QUANTA 2006

X-Ray Structure and Analysis

MAY 2006

Copyright

(1) Copyright

(2) Trademark

(3) Restrictions on Government Use

(4) Acknowledgments, and References

To print photographs or files of computational results (figures and/or data) obtained using Accelrys software, acknowledge the source in an appropriate format. For example:

> “Computational results obtained using software programs from Accelrys Software Inc. Dynamics calculations performed with the Discover program using the CFF forcefield, ab initio calculations performed with the DMol^3 program, and graphical displays generated with the Cerius^2 molecular modeling system.”

To reference an Accelrys publication in another publication, Accelrys Software, Inc., is the author and the publisher. For example:

> Accelrys Software Inc., Cerius^2 Modeling Environment, Release 4.8, San Diego: Accelrys Software Inc., 2006.

(5) Request for Permission to Reprint and Acknowledgment

Accelrys may grant permission to republish or reprint its copyrighted materials. Requests should be submitted to Accelrys Scientific Support, either through electronic mail to support@accelrys.com, or in writing to:

> Accelrys Scientific Support
> 10188 Telesis Court, Suite 100
> San Diego, CA 92121

Please include an acknowledgement “Reprinted with permission from Accelrys Software Inc., Document name, Month Year, Accelrys Software Inc., San Diego.” For example:

> Reprinted with permission from Accelrys Software Inc., Cerius^2 User Guide, June 2006, Accelrys Software Inc.: San Diego.

Contents

Preface: How to Use This Book 1
How to find information 2
Typographical conventions 2

1. Introduction to X-Ray Structure Analysis and Refinement 5
Accelrys' X-Ray crystallography applications 6
X-POWERFIT 6
X-AUTOFIT 6
X-BUILD. 6
3D notebook and other tools 6
CNX 7
X-SOLVATE 7
X-LIGAND. 7
The crystallographic structure determination process . . . 8
Before you begin 8
Collect and process data. 9
Determine phases. 9
Generate initial model. 10
Build and refine model. 10
Refinement 10
Adding ligands. 10
Add solvent and refine final model. 10
Assess model 11

2. Using the X-applications 13
Introduction 13
Application subdivision 13
Electron density Bones 14
Mask generation. 15
Cα tracing 15
Adding Cα trace atoms. 15
Moving Cα trace atoms. 16
Moving Cα-trace segments 16
Cα trace polarity 17
Sequence assignment 17
Loading sequences 17
Sequence assignment 18
Unique sequence 18
Cα-trace -> all atom model 19
Chain polarity 19
Where do the coordinates go?. 19
Model building 20
Refinement 20
Grid refinement. 21
Gradient refinement 21
Monte Carlo refinement 22
Geometry refinement 22
Model building protein/NA 22
Model build ligands and waters 23

General comments 23
Validation 23

3. Managing Maps 25
Interactive Map generation 25
Typical Use. 26
Maps Management table 27
Menu bar 28
Maps menu. 28
Import 28
Add a Map. 29
Delete a Map 29
Replace a Map 29
List Maps 29
Hide Table. 29
Contour menu 30
Contour Maps. 30
Contour Mode. 30
Characteristics of Display Lines 30
Define Contour Levels. 31
Utils menu 32
Purge Contours. 32
Reset Drawing 32
Edit line 32
Column headings 33
Rows 33
Managing maps in various QUANTA applications. 34
X-LIGAND, X-SOLVATE: 34
X-POWERFIT, X-AUTOFIT: X-BUILD: 35

4. Introduction to X-AUTOFIT/X-BUILD/X-POWERFIT (X-FIT) 37
Map generation 37
Map contouring 37
Density skeletonization 38
Solvent masks 38
3D text editor 38
Automated Cα tracing 38
Cα tracing 39
Automated sequence assignment and fuzzy logic assignment 39
Cα trace --> all-atom model 39
Model building 39
Refinement techniques 40
Validation techniques 40
Data logging 41
Data integrity 41
X-AUTOFIT in QUANTA. 41
Memory requirements for X-AUTOFIT 41
Accessing X-AUTOFIT 43
General X-AUTOFIT and X-BUILD behavior 44
Using multiple structures and maps in X-AUTOFIT . . . 44
Palettes and tool activity in X-AUTOFIT 45
Dial box 47
Graph windows 47
Tables. 48

5. Using X-AUTOFIT 49
Solvent boundaries or map masks 49
Parameterization 50
Editing bones 50
Generating a mask 51
Editing masks 51
Summary 51
Bones 52
Displaying and skeletonizing an electron density map. 52
Determining map quality 53
To get information about map quality 53
Improving map quality 53
Modifying the skeletonization initial cut-off parameter 54
To adjust the Start parameter 54
Adjusting branch trimming parameters 55
To adjust branch parameters 55
Using bones with masks 55
Deleting bones points 56
To delete bones points 56
To change the strand type 56
Bones and symmetry 56
Using bones with Cα-tracing 57
Cα-tracing 57
Generating Cα segments using assisted carbon building 58
To generate a new segment 58
Editing segment and Cα atoms 59
Using the pseudo-Ramachandran plot 60
Positioning the next Cα atom 62
Evaluating and changing segment polarity 63
To evaluate segment polarity 63
To change segment polarity 63
Cut/paste Cα segments 64
To join two segments 64
Cutting segments 64
Templates and rigid body editing of Cα traces 64
Sequence assignment 65
Reading in sequence information 65
Fitting a sequence to a segment 65
Generating sequence alignment information . . . 66
Redefining the weights 68
Creating new definitions 69
Finding unique sequences 69
Predicting secondary structure 70
Building all-atom representation from Cα trace 70
Which builder to use 70
How to build 71
Building main chain coordinates by RSR 71
Building mainchain coordinates by database fragment fitting 72
Building the mainchain coordinates by Cα direct correlation 72
Building the sidechain coordinates by RSR 72
Building sidechains by modeling 74
Automated rebuilding in X-BUILD 74

References 75
6. Using X-BUILD 77
Protein or nucleic acids. 77
Proteins 77
DNA/RNA. 77
Nomenclature 78
Hydrogen modes 78
Controlling the display 78
Extent of display 78
The current residue value 79
Modal and amodal/active residue mode. 79
Placement of the view 79
By coordinates 80
By bones. 80
By atom 80
Along sequence. 80
Graph plots 81
Placement by Validation graph. 81
Placement by Validation Table. 81
Placement by Last Command table.. 81
By text markers 82
Graph plots 82
Ramachandran plots. 82
DNA/RNA circle plot 83
Picking screen information 83
Symmetry. 84
Interface to run external programs 85
Advanced validation techniques. 85
Table data 86
Picking data 87
Table contents 87
Table 1 (atom table) 87
Table 2 (residue table) 88
General table 88
Scratch table 89
Fixed size of atom and residue tables 89
Inconsistent data 89
Graphs and plotting 90
Table names 90
X data 90
Legends/annotation 90
Add/deleting annotations. 91
Hardcopy plotting 91
Last commands 91
The table contents. 92
Picking title entries in the table 92
Picking cell entries. 93
X-BUILD features. 94
RSR vs. gradient body refinement 94
RSR vs. Move atom and RSR 94
Defining torsion angles for unknown residues 95
Missing or incorrect atoms. 95
Regularization. 95
Regularization compared to RSR 96

Regularization and disulfides 96
Hydrogen representations 97
Disorder 97
Clamping of alternate conformations 97
C-terminal oxygen atoms in proteins. 98
5´ Terminal residues in nucleic acids 98
Summary of using X-BUILD 99
Editing each residue 99
Waters 99
Color 99
Manual editing 100
Loops 100
Structure refinement. 101

7. Using X-POWERFIT 103
Introduction 103
Data 104
Who can use this application? 104
Map quality. 104
Reduced representation (low resolution tracing only) . . 105
Before starting. 105
Maps 105
Masks 106
Bones 106
The bones parameterization 106
Start parameter. 107
Applying a map mask 107
Tracing high resolution structures. 107
Pathway analysis 108
Fragment analysis. 108
Generation of Cα trace 109
Tracing low resolution structures 109
Determining the secondary structure in the molecule . . 109
Introduction 109
Pattern recognition. 110
Cluster analysis 110
Sheet structure 111
Refinement 111
Overlap analysis. 111
Placing the initial Cα atoms 112
Automatic generation of the Cα trace 112
Start point for tracing 113
Pathway analysis 113
Consensus tracing 115
Adding secondary structure elements 116
Adding more secondary structure 116
Searching the PDB for similar motif patterns 117
Building general structure 117
Checking the trace for knots. 118
Cα refinement 118
Generation of all atom models 118

8. X-AUTOFIT:X-BUILD Tools 121
Introduction 121
To start X-AUTOFIT 121
Map options 123

Calculate map 123
Symmetry palette 123
Define symmetry 124
Define NCS symmetry 124
Unit cell 125
CA Packing Diagram 125
Packing Diagram 125
...Packing radius 126
CA Filled Cell 126
Filled Cell 126
Delete Sym. Ob. 126
Write Sym. to PDB 126
Re-arrange atoms. 126
Hide this menu 127
Pointer palette. 127
Show pointer 127
Show Ruler. 127
Map table 127
Interactive contour. 127
Active residue on and Active residue off 128
Pointer dials. 128
Go to pointer 128
Place at next residue. 128
Place at previous residue 129
Place by atom. 129
Place using bones. 129
Place using coordinates 129
Place by 3D position 129
Suggestions 130
Hide this menu 130
Interactive Contour palette 130
Next Map. 130
Previous Map 131
Next Contour 131
Previous Contour 131
Accept 131
Quit. 131
Notes. 131
Bones palette 132
Bones on/off. 132
Mask bones by mask 132
Set bones/RSR map. 132
Find nice area of map 133
Map quality from bones 133
Smooth bones 133
Side chains on/off 133
Main <-> side 133
Delete 1 section. 134
Delete fragments 134
Delete all fragments 134
Reset delete all 135
Undo last delete 135
Create bones atoms 135
Change start value 135

Change trim/analysis. 136
Calc bones symmetry. 136
Symmetry off 136
Hide this menu 136
Map mask palette 137
Calc. mask from bones. 137
Calc. mask from coord 137
Mask off 137
Mask dials 138
Mask delete radius 138
Solvent content 138
Add mask at pointer 138
Del mask at pointer 138
Check for voids 139
Increase resolution/ Decrease resolution 139
Save mask to file 139
Read mask from file 139
Hide this menu 140
Text palette 140
New text at pointer 140
Delete text 140
Goto next text 140
Goto previous text 140
Goto defined text. 141
...Fix Validate error 141
Load Property 141
Hide this menu 141
CA Build palette 142
Load CA coordinates 142
Next bones box 142
Next CA 142
Delete current CA. 143
New segment 143
Delete current segment 143
Reverse chain 143
Current res/seg 143
Join 2 segments. 144
Unjoin 2 CA 144
Add helix/strand 144
Move current seg 145
Refine current seg 145
Check CA direction 145
Show points ~ 3.8Å 146
Guess next CA. 146
Fit seg by RSR 146
Fit seg by database. 147
Fit seg by CA correlation 147
Fit seg by D.E.E. 148
Delete fitted seg 148
Undo last build 148
Save changes 149
X-POWERFIT 149
Find sec. struct. 149
Delete vectors. 149

Place-edit strand 150
Place-edit helix 150
Save vectors 151
Read vectors. 151
CA Search/Browse 151
Vector Search/Browse 151
Generating a new database for searching. 152
...Next solution 152
...Previous solution. 153
...Solution to CA 153
Current res. and seg.. 153
Vector to CA-trace 153
Next CA as helix/strand 154
Auto extend CA 155
CA refinement. 156
Auto-trace Low 156
Auto-trace High 157
Consensus tracing. 157
Knot finder. 157
Hide this menu 158
Sequence palette 158
Load sequence 158
Delete sequence 159
Show/Hide sequence 160
Current res/seg 160
Unique sequence 160
Clear sequence 160
Return fuzzy sequence 160
Guess Sequence 161
Guess sec. struct.. 161
Show/Hide Amino acids 161
Show/Hide fuzzy 161
Undo last change 161
Hide this menu 162
Build atoms palette 162
Refine 1 residue. 162
Restrictions 162
Geometric conformation. 162
Restrictions 163
Fit side chain by RSR 163
Restrictions 164
Fit main chain by RSR 164
Restrictions 164
Move atom + RSR 164
Restrictions 165
Edit backbone tor. 165
For polypeptides... 165
For DNA/RNA... 165
Restrictions 166
Edit chi angles. 166
Restrictions 166
Move atom 167
Restrictions 167
Move zone 167

Restrictions . . . 167
Model terminal res.. . . . 168
For polypeptides... . . . 168
For DNA/RNA... . . . 168
Flip torsion 180 degrees . . . 168
Mutate residue . . . 169
For polypeptides... . . . 169
For DNA/RNA... . . . 169
Restrictions . . . 169
Add/delete... . . . 169
Hydrogen bonds . . . 169
Move atom + reg. res. . . . 170
Restrictions... . . . 170
Move atom + reg. zone . . . 171
Restrictions... . . . 171
Regularize . . . 172
For polypeptides... . . . 172
For DNA/RNA (for all regularization tools)... . . . 172
Restrictions . . . 172
...fix atoms . . . 173
Fix all CA atoms . . . 173
Fix all CA, N, C, O atoms . . . 173
Fix atoms by picking . . . 173
Clear fix atoms . . . 173
Color atoms... . . . 173
Edit atom info . . . 173
Show residue bumps . . . 174
Save changes . . . 174
Undo last fit . . . 174
Hide this menu . . . 174
Add-delete palette . . . 174
Delete bond . . . 174
Create bond . . . 175
Add torsion by bond . . . 175
Add alternative conformation . . . 176
Delete residue . . . 176
Delete range . . . 176
Add res. at termini . . . 176
For polypeptides... . . . 176
For DNA/RNA... . . . 177
Add water at pointer . . . 177
Add atom at pointer . . . 177
Re-patch terminal . . . 178
For polypeptides... . . . 178
For DNA/RNA... . . . 178
Show B or U . . . 178
Renumber sequence ID . . . 179
Hide this menu . . . 179
Color atoms... palette . . . 179
Color by atom . . . 179
Color by progress . . . 179
Color by (ρ) fit . . . 180
Color by B-value . . . 180
Color by occupancy . . . 180

Color B-alt different . . . 180
Hide this menu . . . 181
Structure palette . . . 181
Fragment fitting . . . 181
Rigid body fit . . . 181
Refine zone . . . 182
Loop fit & Terminal fit . . . 183
Fitting/Continue search . . . 183
Do all... . . . 184
Refine all water . . . 186
...Delete bad water . . . 187
Fit by D.E.E. . . . 187
Undo last fit . . . 188
Regularise range . . . 188
Regularise volume . . . 188
Regularise param . . . 188
Refine volume . . . 188
Auto build . . . 189
Change all B/O . . . 189
Mutate all . . . 189
Hide this menu . . . 190
Tables and Graphs . . . 190
Read file . . . 190
Read molecule . . . 191
Del. Current col. . . . 191
Del. Current row . . . 191
Del. Table . . . 191
Atom Selection . . . 192
Column function . . . 192
Calculations . . . 193
Protein property . . . 193
Differences . . . 195
Plot data . . . 196
Postscript settings... . . . 197
Label graph... . . . 198
Hide this menu . . . 198
User Defined tools . . . 198
Define new tools . . . 199
Define-Edit group . . . 199
...Add menu space . . . 200
...Clear all tools . . . 200
X-AUTOFIT:X-BUILD main palette tools . . . 200
Validate . . . 200
Find Negative density . . . 202
Set bones/RSR map . . . 202
Options... . . . 202
Color table . . . 203
Save built CA to MSF . . . 203
Save built atoms to MSF . . . 203
Save built atoms to PDB . . . 203
Run external program . . . 204
Last commands... . . . 204
Finish . . . 204
Choose the MSF Saving Option . . . 204

X-AUTOFIT Options dialog box 204
Rotamer library 204
Protein build mode 205
DNA build mode 205
Return main dials after edit 205
B-conf clamped to A backbone 205
Center at Ramachandran point for residues... 206
Ramachandran range 206
Show bumps less than 3.00 206
Show bumps to inactive molecules 206
Symmetry picking mode 207
Next residue step 207
Number of DP 207
Hydrogen representation 207
Search time limit 208
Number of steps for regularize 208
Phi/Psi restraint 208
Omega restraint 209
Nonbonding in regularization 209
Regularize across disulfide bonds 210
Coordinate radius 210
Map radius 210
Symmetry radius 210
Resolution factor 211
Color table 211
Map (n) 211
Bones main 212
Bones side fields 212
Symmetry 212
NCS 212
CA trace 212
Text 213
Mask 213
Vectors 213
Moving atoms 213
Pointer 213
Last commands 214
Menu bar at the top of the table 214
FILE: Write log file 214
FILE: Command saving off 214
FILE: Browse old command files 215
FILE: Delete old files 215
VIEW: Info (+) lines: Show/Hide 215
VIEW: User comment: Show/Hide 215
ANALYSIS: Check command file 215
ANALYSIS: Make suggestion 215
The last command table 216
Cell picking 216
External program palette 217
Definitions 218
General form of a script 218
Variables 218
General formatting 218
Program definition line 219

Assignment of variables . . . 219
Un-initialized variables . . . 219
Control statement . . . 220
QUANTA data export statements . . . 221
QUANTA data import statements . . . 221
Script definitions key words . . . 221
Command file control statement . . . 222
Dialog only control statements . . . 222
Dialog and command file statements . . . 224
Format additions . . . 225
Continuation marker . . . 225
References . . . 227

9. Using X-LIGAND . . . 229
Overview . . . 229
Requirements . . . 230
Ligands . . . 230
Rotatable bonds . . . 230
Edited rotatable bonds . . . 231
MSF file requirements . . . 231
Symmetry . . . 232
General use . . . 232
Active site searching . . . 232
Conformation searching . . . 232
Precision . . . 232
Refinement . . . 233
Masked tools . . . 233
X-LIGAND palette . . . 234
Search for ligands . . . 234
Adjust site volume . . . 234
Change search setting... . . . 234
Search threshold . . . 234
Conf search mode . . . 235
Electron density truncation . . . 235
Non-bonding . . . 235
Partial bond torsion search . . . 235
Use 3rd moments . . . 235
Z-weights . . . 236
Regularize . . . 236
Grid search parameters / Max conf to search . . . 236
Max time for search . . . 236
Intra-ligand Non-bonding . . . 236
Grid search parameters / Max tors to search . . . 237
Radii / Map display radius . . . 237
Radii / Calculation radius . . . 237
Position search centre . . . 237
Map on/off . . . 237
Reset Ligand . . . 238
Next site . . . 238
Previous site . . . 238
Goto site . . . 238
Search conformations . . . 238
...Refine all . . . 239
...next conformation . . . 239
...previous conformation . . . 239

...all 20 conformations . . . 239
...Fix search origin . . . 240
...Fix ligand atoms . . . 240
...Clear fixed atoms . . . 240
...fit in stages . . . 240
Flip tor 180 . . . 240
Set active torsions . . . 241
Add/del 1 torsion . . . 241
Edit torsions . . . 241
Edit position . . . 242
Refine . . . 242
Score fit . . . 242
Save ligand to MSF . . . 242
Select new ligand . . . 242
Database/list . . . 242
Exit . . . 243

10. Using X-SOLVATE . . . 245
Molecular coordinates used by X-SOLVATE . . . 245
Saving water molecules on exit from X-SOLVATE . . . 245
Accessing X-SOLVATE . . . 246
Before you begin . . . 246
To begin X-SOLVATE . . . 246
Using X-SOLVATE . . . 247
Searching for water molecules . . . 247
To run a search . . . 247
Adjusting water molecule position . . . 248
Saving the water to MSF . . . 248
Search for peaks palette . . . 248
Search for waters . . . 248
Change search setting... . . . 249
Position search center . . . 249
Map on/off . . . 249
Previous peak . . . 249
Next peak . . . 249
Goto peak . . . 249
Save as water . . . 250
Delete water . . . 250
Save all peaks . . . 250
Exit . . . 250
Quit . . . 250
Peak search parameters dialog box . . . 250
Min density level . . . 250
Min distance to protein . . . 251
Max distance to protein . . . 251
Min sep between peaks . . . 251
Calculation radius . . . 251
Display radius of map . . . 251
Temperature factor . . . 251

11. Using the CNX Interface . . . 253
The CNX Interface palette . . . 253
Palette tools . . . 254
Set Symmetry . . . 254
Set-up Structure Factors . . . 254
Parameter Set . . . 255

Map Generation . . . 256
Rigid Body Refinement . . . 257
Refine Structure . . . 257
Other CNX Interface tools . . . 258
Handling disorder in QUANTA and the CNX interface 259

12. Setting Up Molecular Systems for CNX . . . 261
This chapter describes . . . 261
Before you begin . . . 261
Setting up a CNX system for a simple protein . . . 261
Setting up the structure for CNX . . . 261
Running a CNX job to minimize the system . . . 262
Running standalone CNX . . . 264
Dealing with possible problems . . . 264
If CNX does not run or cannot be found. . . . 264
Seeing the attached-atoms error message . . . 264
Unable to generate a PSF because of unknown atom types . . . 265
Setting up a system for a nucleic acid, solvent, ion, and ligand . . . 265
Doing initial setup . . . 265
Preparing the other molecules . . . 265
Setting up the CNX calculation . . . 266

13. Tutorial . . . 269
Setup for the tutorial . . . 269
Lesson 1: Getting Started with QUANTA . . . 269
Lesson 2: Getting Started with the X-Ray Modules . . . 276
Lesson 3: Manual CA-tracing . . . 284
References . . . 294
Lesson 4: Automated CA-tracing . . . 294
References . . . 302
Lesson 5: Sequence assignment . . . 303
References . . . 307
Lesson 6: Model Building and Real Space Refinement using X-BUILD . . . 307
References . . . 317
Lesson 7: Ligand Fitting using X-LIGAND . . . 318
References . . . 323
Lesson 8: Water Fitting Using X-SOLVATE . . . 323

A. References . . . 329

B. Creating a Fragment Database . . . 331

C. The cnx.bat File . . . 333
The cnx.bat file . . . 333

D. Searching for Fragments . . . 335
Searching for fragments . . . 335
Fragment database . . . 335
Fragment selection criteria . . . 336
Running a search . . . 336
Search Fragment Database palette . . . 337
List Proteins . . . 337
Pick Alpha Carbon Range . . . 337
Pick Alpha Carbon . . . 338
Undo Last . . . 338

Undo All . . . 338
Search Database . . . 338
With Bumps . . . 338
Display All Fragments . . . 338
Display Next . . . 338
Display Previous . . . 339
Select Display... . . . 339
List Residues . . . 339
Accept Fragment . . . 339
Anneal Fragment . . . 339
Reject Fragment . . . 339
Options... . . . 340
Finish . . . 340
Example: Using fragment searching to complete a model of renin . . . 340
Before you begin . . . 340

E. Using RTF and PSF Modes . . . 345
RTF and PSF files and modes . . . 345
RTF mode . . . 345
Setting up a protein in PSF mode . . . 347
Organizing and separating segments . . . 347
Adding polar hydrogens . . . 347
Using CHARMm functions in RTF mode . . . 348
Using the Protein Editor . . . 348
Calculating a CHARMm energy . . . 348
Saving changes . . . 348
Problems . . . 348

F. X-Ray - X-BUILD Command Conversions . . . 351
Map management . . . 351
Specific command conversions . . . 352

G. Extend . . . 355
Where to find the program . . . 355
Introduction . . . 355
-extend . . . 355
-group (space group) . . . 356
-symops (file) [$HYD_LIB/symlib.sym] . . . 356
-limits (xmin xmax ymin ymax zmin zmax) . . . 356
-radial (x y z r) . . . 356
-pdb (pdb-file) . . . 356
-msf (msf-file) . . . 356
-input . . . 356
-boundary (value [0.0]) . . . 356
-oformat (filename) . . . 357
-cutoff (cutoff [0.5]) . . . 357
Examples . . . 357
General comments on the program . . . 357

How to Use This Book

This book provides an introduction to the use of QUANTA's X-Ray crystallography applications. The first chapter contains an overview of the crystallographic process and the components of Accelrys' crystallographic software applications that facilitate each activity in the process. The remaining chapters describe the applications in detail.

How to find information

If you want to know about…	Read…
Managing maps and using the Maps Management table	Managing Maps (page 25)
General information about fitting coordinates to maps	Introduction to X-AUTOFIT/X-BUILD/X-POWERFIT (X-FIT) (page 37)
Background and tips for skeletonizing and fitting coordinates to electron density maps	Using X-AUTOFIT (page 49)
Model building	Using X-BUILD (page 77)
Generate Cα traces from electron density maps	Using X-POWERFIT (page 103)
Using the commands on the X-AUTOFIT or X-BUILD palettes	X-AUTOFIT:X-BUILD Tools (page 121)
Fitting ligand coordinates automatically into electron density maps	Using X-LIGAND (page 229)
Searching electron density maps for water peaks	Using X-SOLVATE (page 245)
Controlling the 3D pointer	Using the 3D Pointer (page 265)
Creating databases to be used by Search Fragment Database.	Creating a Fragment Database (page 331)
Using the cnx.bat file	The cnx.bat File (page 333)
Searching a specialized database to locate a suitable set of fragments for modeling undefined regions of a protein	Searching for Fragments (page 335)
Using RTF and PSF modes	Using RTF and PSF Modes (page 345)
An extension of the mbkall program	Extend (page 355)

Typographical conventions

Unless otherwise noted in the text, this manual uses the typographical conventions described below:

- Words in *italic* represent variables. For example:

  The application saves the file to *filename*.msf

  In this example, the name of the MSF that you want to use replaces the value *filename*.

- Sample syntax and the examples illustrating the contents of files are presented in a `fixed-width` font. For example, the following illustrates a line in an input file:

```
ACAM_HAMMH.CRD
```

- Names of items in the interface or presented on the screen are presented in **bold**. For example:

  Select **X-AUTOFIT** on the **Applications** menu.

1. Introduction to X-Ray Structure Analysis and Refinement

The crystallographic family of products in QUANTA allows the automated addition of solvent (X-SOLVATE), the automatic fitting of ligands (X-LIGAND), de novo density fitting (X-AUTOFIT and X-POWERFIT), general model building (X-BUILD) and Refinement (CNX interface).

These applications provide a complete set of tools, from the tracing of the first map to final refinement and analysis.

The implementation of X-Ray applications within QUANTA provides an integrated environment in which many additional tools from other QUANTA applications, including Protein Design, Protein Health, and Conformational Search, can be used to enhance the model building and refinement processes for proteins and other structures. This chapter describes

- QUANTA's X-Ray crystallography applications (Accelrys' X-Ray crystallography applications (page 6)).

  This section describes the various X-Ray applications that are integrated within QUANTA. Together, these components provide tools for you to process X-Ray crystallography information from raw data to a refined molecular model of your structure.

  Cross references to other chapters in this book and to material in other documentation direct you to detailed information on software applications and functions used to complete the crystallographic process described in the next section.

- The crystallographic process (page 8).

  This section describes the basic activities of the crystallographic process and the software that facilitates each activity.

  Use this chapter to determine which components of the Crystallography Workbench Applications you need in your work, as well as when and how to apply them. The remaining chapters of this book describe the crystallographic applications and the CNX interface.

Accelrys' X-Ray crystallography applications

Accelrys' crystallographic software library consists of the following applications that together facilitate the complete structure determination process from crystallographic data:

X-POWERFIT

X-POWERFIT provides a tracing methods for data at resolutions from 1Å to 4 Å. With higher resolution data it is possible to calculate a single Cα trace from a simultaneous analysis of all the bones. At lower resolutions automated interpretation tools streamline the process using secondary structure analysis of the map followed by sequential interpretation using Cα batons. The algorithms available can provide a 500-fold speed up over conventional tracing methods and result in superior quality models.

X-AUTOFIT

X-AUTOFIT speeds and enhances the process of fitting coordinates to a SIR, MIR, or MAD map. It skeletonizes electron density maps of proteins, intelligently places alpha carbons to create carbon traces, aligns segments to known molecular sequences, and automatically builds atomic coordinates to fit an electron density map. It can also be powerfully applied in automatically fitting a molecular replacement map.

X-BUILD

The X-BUILD application contains many automated model building tools. The application includes interactive regularization, Mote Carlo/grid/gradient real space torsion angle refinement algorithms, and automated water refinement. These and many other tools streamline the process of model building, resulting in a several-fold improvement in productivity.

3D notebook and other tools

The integrated X-AUTOFIT and X-BUILD tools include various support applications: a 3D notebook that allows annotation of a molecule, a pointer palette that allows rapid movement around a protein molecule, and validation tools. A Ramachandran plot is always visible while X-BUILD is active, and a Cα plot is always active for the X-AUTOFIT application.

Automated and advanced validation is provided specific for crystallographic structure determination as well as full data logging in X-BUILD for data recovery.

CNX

CNX is the standard for 3D structure determination of macromolecules using crystallographic or NMR data. The QUANTA interface to CNX export an entire molecular system from QUANTA, including a coordinates file and a script, a principal structure file (PSF), and a parameter file. The interface also launches CNX calculations to perform map generation, simulated annealing, and positional refinement.

You can submit data to CNX interactively or in standalone mode. Results are returned to QUANTA for further manipulation and analysis.

For detailed information about CNX, see Brunger (1992) and the *CNX 2006 User Guide* (published separately by Accelrys). For more information about the CNX interface, see Setting Up Molecular Systems for CNX (page 261).

X-SOLVATE

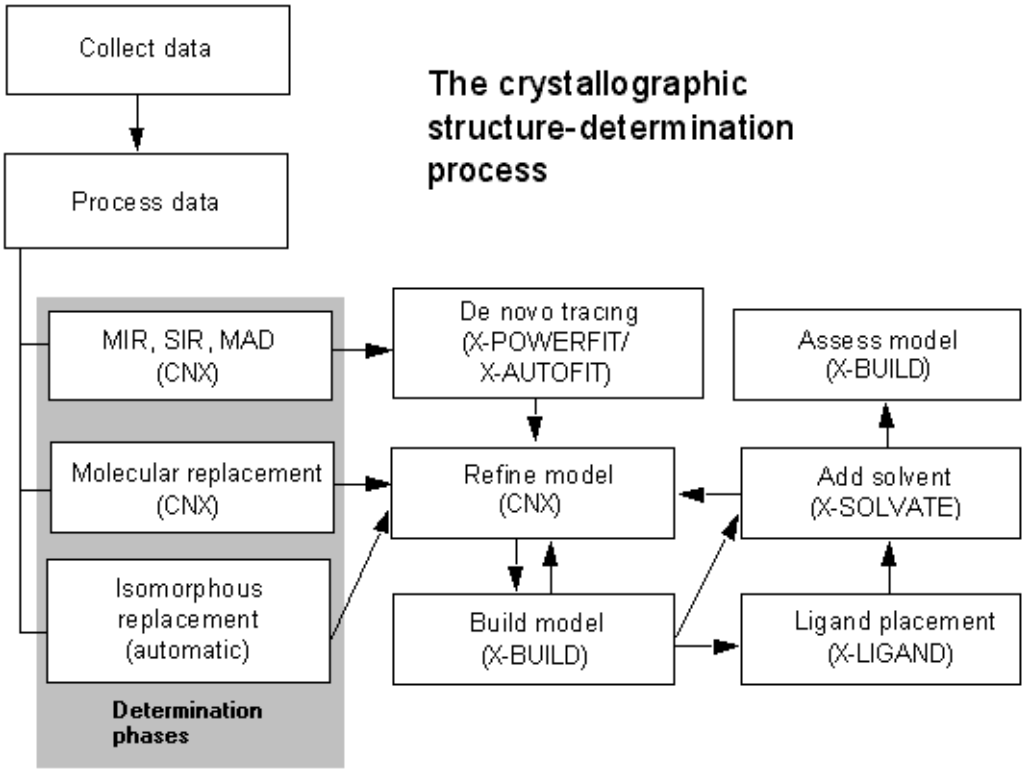
X-SOLVATE is a rapid method for searching an electron density map for water molecules. The application searches the map for electron density not already filled by atoms and assesses the contacts made with the protein, at all times dealing with symmetry equivalents. Peaks can be assessed interactively and saved as water coordinates. For more information on X-SOLVATE, see Using X-SOLVATE (page 245).

X-LIGAND

The X-LIGAND application allows the rapid searching of regions of significant connected density that may be due to a ligand molecule and not already filled by protein atoms. Different ligands can be fitted to the sorted list of sites, internal degrees of freedom of the ligand can be searched at the rate of thousands per second, and the solution refined using real space torsion angle refinement. For more information on X-LIGAND, see Using X-LIGAND (page 229).

The crystallographic structure determination process

The following figure provides an overview of the tasks that the crystallographer can perform to determine and refine a macromolecular structure:



The following paragraphs describe the process outlined in the figure above:

Before you begin

Regardless of where you start in the crystallographic process, you must have a set of structure factors available from. Store these data in a structure factor file (.fob). If no experimental data are available, you can generate a dummy structure factor file. However, in this case, the electron density map that is generated is not physically relevant. For information on how to generate a dummy file.

Collect and process data

Experiments performed on an in-house area detector or at a synchrotron facility result in X-Ray data frames that must be processed to extract the intensity of each diffraction peak. X-GEN provides facilities to process these data to the point where a set of merged, corrected intensity measurements are computed and output.

Determine phases

For a new protein structure, phases must be calculated for each diffraction peak, using one or a combination of techniques. Three basic strategies are available for phase determination:

- Multiple isomorphous replacement (MIR)

  Multiple isomorphous replacement is used when diffraction data are collected for several crystals with various bound heavy metal atoms. You can use the combination of resulting heavy-atom positions to compute phases. Alternatively, you can collect data with multiple wavelength or anomalous dispersion techniques on a single isomorphous derivative to obtain the phase information.

- Molecular replacement

  For a protein that is similar to one with a known structure or where a reliable model structure can be generated, you can use the technique of molecular replacement to obtain phases for the new structure. With this technique, a search model of the protein is rotated and translated through the diffraction data to locate orientations that maximize agreement between calculated and experimental data.

- Isomorphous replacement

  Use isomorphous replacement when your protein structure is essentially identical to a known structure. Existing phase data can be used to determine phases from the data for crystals of a mutant or ligand complexed protein.

CNX provides tools for all types of phase determination, including heavy-atom positioning, density modification, and all the methods described above. Molecular replacement is also available within CNX.

When you have obtained phases, you can compute an electron density map and begin the process of building and refining a model for the protein structure. The procedures to be followed for generating a model are somewhat different for each phase determination strategy. However, inspection of the

model in the electron density map, manual model building, and refinement are common for all approaches.

Generate initial model

If a structure has been solved by MIR, SIR, or MAD, then an initial model for the protein must be constructed from an initial map. X-POWERFIT can automatically trace the Cα atoms form high resolution phased data. For low resolution data, X-POWERFIT allows the identification of secondary structure elements (vectors) and automatic conversion of the vectors to Cα traces. For low resolution structures, as well as the more disordered regions of high resolution structures, X-AUTOFIT can be used to continue the Cα trace construction from skeletonized electron density maps, with built-in intelligence about features of protein structure. This initial Cα trace is converted to an all-atom model using automated tools within the X-AUTOFIT functionality.

Build and refine model

The all-atom models generated from X-AUTOFIT or through the process of molecular replacement can be re-built using the X-BUILD application. This allows manual and automated editing of the atom positions and real-space refinement into electron density maps calculated with CNX.

Refinement

The refinement of model coordinates can be carried out with simulated annealing or traditional least-squares methods within CNX. The QUANTA interface to this program allows automated setup of the required protocols.

Adding ligands

It is possible to add ligand molecules automatically to electron densities using the X-LIGAND application. This application includes conformation searching of ligand flexibility and refinement of the ligand to the electron density.

Add solvent and refine final model

You can use an iterative process for adding solvent and doing small manual rebuilding and careful refinement to produce a final model of the molecule.

Assess model

You can assess the quality of the final model and also analyze it using the extensive validation facilities designed for detection of crystallographic building error in the X-BUILD application. More general structure analysis tools are found in the Protein Health application in QUANTA and in Insight II.

2. Using the X-applications

Introduction

This section is designed to guide you through the large number of tools available in X-AUTOFIT, X-POWERFIT, X-BUILD, X-LIGAND and X-SOLVATE.

The aim of this section is to provide a starting point for understanding the X-RAY applications, where the functionality is found, and how the functionality should be used to give the best results.

Application subdivision

The X-applications are provided so that X-AUTOFIT, X-BUILD, and X-POWERFIT form an integrated system through a single main palette accessed from the **Applications** menu on the main menu bar of the QUANTA window. X-LIGAND and X-SOLVATE are separate applications run from the **Applications** menu.

Table 1. Application subdivision

Application	Function	Access
X-AUTOFIT	Cα-tracing (**CA-build**). Automated sequence assignment (**Sequence**). All atom auto-generation (**CA-build**).	X-AUTOFIT:X-BUILD palette **Applications	XAUTOFIT**
X-BUILD	Model building in a semi-automated fashion.	X-AUTOFIT:X-BUILD palette **Applications	XBUILD**
X-POWERFIT	Automated map interpretation, used in conjunction with X-AUTOFIT.	X-POWERFIT palette **X-POWERFIT** button on the X-AUTOFIT:X-BUILD palette
X-LIGAND	Automatically adds ligands.	**Applications	X-LIGAND**
X-SOLVATE	Automatically adds solvents.	**Applications	X-SOLVATE**

Table 2. Subpalettes launched from the X-AUTOFIT:X-BUILD palette

Button	Palette	Description
Symmetry...	Symmetry	Symmetry annotations and control
Pointer...	Pointer	View placement control
Bones...	Bones	Map skeleton calculation and manipulation
Map Mask...	Map Mask	Map mask creation and editing
CA Build...	CA Build	Semi-automated map tracing and manipulation
X-POWERFIT	X-POWERFIT	Automated map tracing
Text...	3D Text	Notebook annotation for structure
Sequence...	Sequence	Sequence assignment to Cα trace
Build atoms...	Build atoms	Model rebuilding tools
Structure...	Structure	Advanced model rebuilding tools
Tables and Graphs...	Tables and Graphs	Advanced validation and Analysis

Electron density Bones

Electron density bones are integral to the X-Ray applications and provide near automated map analysis and analysis. They are generated from the map only in regions of interest and so are controlled by the map radius. They can be used to generate map masks as well as in the process of Cα-tracing and can be automatically interpreted in X-POWERFIT or semi-automatically in X-AUTOFIT.

Bones are not converted into atoms but represent a background annotation view on which Cα trace atoms are added. Therefore it is not necessary to edit the bones in detail, since the only represent a backdrop guide; let the program auto-edit these with the parameters provided (**Bones | Change start value, Bones | Change trim value**).

Any editing of the bones is lost by recalculating the bones.

If a map mask is to be calculated from the bones, use a large map radius, and use the bones editing tools to trim the bones to a molecular shaped volume. Use the **Map mask | Solvent content** tool and the **Bones | Bones symmetry** tool to check the extent of the mask and bones.

Once a map mask has been generated, it is not necessary to manually trim the bones for any further protocols.

If bones are to be used in X-POWERFIT for automated tracing, use a large map radius, a map mask to delimit the molecular volume (**bones/mask bones by mask**) and the **Auto-trace High** (high resolution data) or **Find**

sec. struct. | Auto-trace Low (low resolution data) tools under the X-POWERFIT palette.

If bones are to be used in X-AUTOFIT (and X-POWERFIT) for semi-automated map tracing then use a small map radius (about 9Å is suitable), and work along the trace using the **Bones | Next bones box** tool to update the view when the trace nears the edge of the calculated volume of bones.

Mask generation

Masks can be generated from coordinate and map information as well as from O format files (old, new and compressed). This functionality is available on the Map Mask palette.

Mask editing is possible by either editing electron density bones before a mask is generated or by directly editing the mask with a spherical pointer.

Masks are used in X-AUTOFIT and X-POWERFIT as bounding masks for all tracing and calculations (**Bones | Mask bones by mask**) and provide a mask for external solvent flattening functionality in CNX.

Cα tracing

Atoms are placed into electron density by the process of Cα tracing using the bones as a background object.

Adding Cα trace atoms.

Cα tracing is only possible when the bones are active because bones analysis is integral to creating a Cα trace:

- As a new segment of two Cα trace atoms (**CA-build | New segment**).
- As a single Cα trace atom (**CA-build | Next CA**).
- As multiple Cα trace atoms automatically (**X-POWERFIT | Auto extend CA**).
- A a single Cα trace atom of a helix/strand (**X-POWERFIT | Next CA as helix-strand**).
- As a helix/strand manually (**CA-build | Add helix or strand**).
- Automatically, as a helix/strand from a vector (**X-POWERFIT | Vector to CA trace**).
- From an MSF coordinate set (**CA-build | Load CA coordinates**).

Cα-tracing may be done automatically using X-POWERFIT or in X-AUTOFIT in a semi-automated fashion.

There are three types of Cα-trace atoms: (1) an active Cα atom in (2) an active Cα trace segment, and (3) other Cα-trace segments placed previously.

The active Cα trace atom (and active segment) is set using the **Current res/seg** tool on the X-POWERFIT, Cα build and Sequence palettes.

Moving Cα trace atoms

If the active Cα trace atom is at a segment terminus, then it can only be moved on the surface of a sphere of radius 3.8Å from the previous Cα trace atom by changing an opening angle and torsion.

If the active Cα trace atom is not at a terminus, then it can be moved in x/y/z screen ordinate. The Cα-Cα distances affected by this are shown on the QUANTA message line.

Only the active Cα trace atom can be moved in one of the following ways:

- By picking a bones point (active Cα will “move” towards this).
- By picking the Cα-plot graphs (Cα will move to this conformation).
- By moving the dials to change torsion/angle or x/y/z (note that a virtual track ball motion is by far the most efficient method).
- By making an alternative auto-placement (**CA-Build** | **Guess next CA**).

Moving Cα-trace segments

Only the active segment can be moved or changed by:

- Rigid body refinement (**CA build | Refine current seg**).
- Manual rigid body movement (**CA build | Move current seg**).
- Flexible refinement (**X-POWERFIT | CA refinement**).

The Cα trace is automatically saved between sessions and can be explicitly saved using the **CA build | Save changes** tool. The last edit of the Cα trace can be undone/redone using the **CA-build | Undo last change** tool.

Cα trace polarity

A Cα trace is *always* built from the C-terminal end. Therefore setting the current Cα trace atom as a terminal atom sets the Cα segment polarity so that this atom is the C-terminal atom of a segment.

Selecting a *non*-terminus Cα trace atom does not affect the polarity of the Cα-trace. The polarity of the Cα trace is shown as an arrow at the current Cα trace atom pointing to the C-terminal (provided that the current Cα trace atom is not a terminus).

Be aware that you must define the polarity of the Cα trace before converting a Cα trace to an all atom model.

If the current Cα atom is at a terminal, the polarity of the Cα trace may be reversed using the **CA-build | Reverse chain** tool; then the current Cα trace atom is set at the other end of the chain. If the current Cα-trace atom was not at a terminus, the current Cα trace atom is not changed.

The polarity of the Cα trace can be checked with the **CA-build | Check CA direction** tool, which essentially tests for the C=O position in electron density. This makes no change to the segment direction, only notes whether the current direction is right or wrong.

Sequence assignment

Sequence assignment is carried out on Cα trace atoms. The residue type is assigned to the Cα trace atom, and, when a sequence is loaded and displayed, the current residue type is shown at each traced Cα atom. The aim is to assign to the Cα trace atom residue types, so the Cα trace is assigned a unique sequence which is part of or all of the loaded template sequence. The loaded sequence represents a guide to this process for the tools provided; once sequence assignment to the Cα trace is complete the loaded sequence template has no further function.

Loading sequences

A sequence is loaded (from a number of different formats) using the **Sequence | Load sequence** tool. If the protein is to be interpreted as a dimer (or higher repeat), it is recommended you load only the sequence for one monomer unit. This way, the application can identify a unique sequence as only one occurrence of the repeat in the sequence.

Sequence assignment

A Cα trace atom can be assigned one of the 20 amino acids or a number of *fuzzy residue types*. There are 10 pre-defined fuzzy residue types, and you can add a further 10 fuzzy residue descriptions of your own.

Multiple propensity values can be assigned to a single residue — for example a lysine residue can be classed both as big and small since density for this residue could be either. You are directed to the documentation in Using X-AUTOFIT (page 49) on specifying new or changing existing fuzzy propensity values.

Picking a residue type sets the residue type of the current active Cα trace atom. Once any Cα trace atom in a Cα-trace segment is set, a sequence alignment is performed.

Sequence alignment is done in a forward and backward direction, and the results are shown against the sequence trace: blue arrows are forward fits, and red arrows are backward fits.

Unique sequence

The Cα trace atoms are marked with a unique assignment if a unique sequence is found. The sequence table is updated so that the unique section is shown in upper-case. Any subsequence addition of Cα trace atoms will be added as the correct residue type defined by the sequence.

It is not possible to extend an assigned Cα-trace segment if this results in an overlap of the Sequence assignment or in the sequence alignment falling off the end of the sequence table.

Dimer (or higher order) structures can never be uniquely sequence assigned if the sequence is for the whole protein. Please read in only one monomer sequence unit.

Sequence assignment of a second (or higher order) monomer unit is carried out by first making the sequence assignment in the first monomer “not unique” (**Sequence | Unique sequence**). This releases the assigned sequence for further assignment.

Sequence assignment is stored with the Cα-trace and does not require the presence of the loaded sequence or a unique setting. On completion of sequence assignment, the loaded sequence can be removed.

Cα-trace -> all atom model

Cα-trace atoms are converted to all atom models automatically by one of four tools on the CA build palette. The quality of the results depends on the quality of the placement of the Cα trace atoms as well as the quality of the electron density. It is highly recommended that the actual position of each Cα-trace atom be checked and adjusted in a final refinement (**CA build | Refine Current Seg**) before trying to build an all atom model.

Chain polarity

The all atom model is built with a polarity defined by the current polarity of the Cα trace. The Cα trace polarity should be defined by making the C-terminus atom the current atom (**CA build | Current res and seg**). This can be checked against the sequence by examining the arrows and by using the **CA build | Check CA direction** tool to check against the map.

The electron density map *must* cover all the Cα-trace atoms if fitting is to be carried out to electron density. Change the map resolution so that all Cα trace atoms are covered. Turn off the map during the building process, since the repeated re-display of this during the building slows the calculation down.

Knowing which of the four tools to use depends on the quality of the electron density. You are strongly advised *not* to use the **CA build | Fit seg by database** tool, since it produces poor results.

- **CA build | Fit seg by RSR** — For good density and a map resolution of < 1.5Å.
- **CA build | Fit seg by CA corr** — For poor density or a map resolution of > 1.5Å.
- **CA build | Fit seg by D.E.E** — For little or no data.
- **CA build | Fit seg by database**

Where do the coordinates go?

The new coordinates produced by the building process will either be put in a new MSF file or placed in a current MSF file:

If a MSF file is *open/active and visible*, then the new coordinates are added to the *first* active and visible MSF file.

If there is *no open/active and visible*, MSF molecule then a new MSF is generated.

Therefore, if you do *not* want to add new coordinates to a currently open molecule, then ensure that the molecule is *not* active.

Model building

Model building is carried out using the following palettes from the X-BUILD application: Build atoms, Structure, Pointer, and 3D text. Additionally, validation tools are used: data logging with the **Last commands** tool and advanced validation using the Table/graphs palette.

Atom selection for building is carried out using:

- The tools on the Pointer palette.
- Picking the Ramachandran plot.
- Picking a validation graph plot.
- Using a 3D text object.

To edit a single residue, use the tools on the Build atoms palette, and use the Structure palette to refine of a region/volume of residues.

The Pointer palette (generally visible all the time) and the 3D text palette should be used to center the view.

Refinement

X-BUILD has three types of real space torsion angle refinement protocols: gradient, grid and Monte Carlo. In all cases you should be aware that if the electron density does not cover the atoms of interest, then the functionality does not produce correct results.

All the refinement tools improve fit to density by modifying torsion angles. The grid refinement tools also vary angles slightly to improve the radius of convergence, and this can be corrected subsequently using geometry refinement tools. The gradient refinement tools allow small variation in bonds and angles, although these are generally small and do not need to be corrected by a model building tool.

Grid refinement

Grid refinement tests all possible conformations by changing a small number of torsion angles to find the best fit to electron density. It is therefore limited to side chain and main chain protein and nucleic acid fitting.

Grid refinement is used to place side chain atoms or main chain peptide planes into an electron density during the residue by residue fitting process. The tools **Build-atoms | Fit side chain by RSR**, **Build-atoms | Move atoms + RSR** and **Build-atoms | Fit main chain by RSR** carry out Grid refinement.

Grid refinement can only be used on proteins and nucleic acids.

Grid refinement places the sidechain atoms into an electron density at some detriment to angles and improper angles. Hence, it will produce a slightly distorted sidechain, especially where the electron density is poor. It is therefore necessary to regularize residues after grid refinement.

Gradient refinement

Gradient refinement is used to refine any residue or group of residues into an electron density by following the electron density gradient. A single residue can be fitted to an electron density using the **Build atoms | Refine 1 residue** tool, and multiple residues are fitted using the **Structure | Refine zone** tool. These tools use torsion angle refinement, so any defined torsion is allowed to change.

Gradient refinement can be used on any residue or group of residues (protein, nucleic acid, water, ligands).

The **refine 1 residue** tool does not improve the geometry but only refines toward a better electron density. If the bonds/angles and improper angles were initially bad before the use of this tool, they will be bad after using the tool as well.

The **Refine zone** tool improves the fit to electron density and improves geometry terms using mixed parameter refinement. Hence it is possible that the fit to electron density may be worse after using this tool if the geometry was initially poor.

The **Structure | Rigid body fit** tool allows rigid body refinement of regions of residues. This is generally useful to improve the fit of an entire domain after MR.

Monte Carlo refinement

Monte Carlo refinement is used in X-BUILD to place main chain conformations of loops and termini where there are too many degrees of freedom and the atoms are too distant or the density too poor to use gradient refinement.

Structure | Loop fit and **Structure | Terminal fit** are used to fit loops and termini respectively. In both cases they should be considered a “try it and see” protocol where the results take a number of minutes to calculate but require no user input.

Geometry refinement

Geometry refinement is used to improve the bonds/angles and improper geometric terms of atoms. It takes no account of the electron density. This is known as regularization and should be used after side chain fitting and single residue refinement, or after a manual modeling session.

The following tools all carry out geometry refinement:

Build atoms | Regularize

Build atoms | Move atom + reg. res.

Build atoms | Move atom + reg. zone

Structure | Regularize volume

Build atoms | Regularize

Structure | Regularize range

Structure | Refine Zone also carries out geometry refinement as part of the electron density refinement protocol.

Model building protein/NA

We suggest that proteins and nucleic acids be fitted by traversing the polypeptide chain one residue at a time using the **Pointer | Next residue** tool. If the residue does not fit the experimental data, then use the **Build-atoms | Fit side chain by RSR** or **Build atoms | Move atom + RSR** tools. The residue should be tidied up with the **Build-atoms | Regularize** tool.

If a number of consecutive residues have problems, use the **refine zone** tool on the Structure palette or the **Build-atoms | Move atom + reg. res**. tool.

Manually editing residues is, in most cases, unnecessary.

Model build ligands and waters

Ligands can be added in the application X-LIGAND, and waters are added en masse in X-SOLVATE. A single or a small number of waters can be added using the **Build-atoms | Add-delete | Add water at pointer** tool followed by gradient refinement with **Build-atoms | Refine 1 residue**.

Ligands are automatically parameterized for both real space torsion angle refinement and geometry refinement. Hence it is possible to edit these without generating complex molecular descriptions. Generally, ligands should be fitted to electron density (if not already placed in X-LIGAND) with the **Build-atoms | Refine 1 residue** tool.

All waters can be refined in a single step with the **Structure | Refine all water** tool; it is also possible to inspect individual water refinements when there is a problem in the **Structure | Do all...** water refinement option.

General comments

Use the Ramachandran plot to identify regions of backbone chain that need rebuilding. The Ramachandran plot can be picked to place the view at a residue in an unlikely conformation.

Alternate conformations can be automatically searched and added from the **Structure | Do all...** tool, but this does require that the structure is near the end point of refinement.

Only the first active and visible molecule can be edited. Any molecule edited will be automatically saved on exit.

Use the 3D text editor to mark problems in model building. These can then be checked in subsequent model building using further refined coordinates.

Validation

Use the **Validate** tool on the main X-BUILD palette to identify problems remaining in the structure at the end of a model building session. Any problem (except a Ramachandran error) can be automatically fixed with the **3D text | Fix validate** error tool.

When the structure is near the end point of refinement use the advanced validation tools on the Table/graphs palette to identify more subtle problems with the structure.

3. Managing Maps

You can generate maps in QUANTA using two new tools on the X-AUTOFIT XBUILD main palette:

Map Options sets up the type of map to be calculated and the reflection file to be used. This is identical to the map calculation dialog from the CNX interface; note that the bottom three options should be ignored: always ask for the calculation to be immediate.

Calculate Map launches a series of calculations that result in a map being generated and displayed on screen.

Map management tools for defining the display of electron density maps are provided through the Maps Management table, making them available throughout all the functionality within QUANTA. This means that the maps functionality can be used at any time while using QUANTA, not just when carrying out crystallographic map fitting in X-AUTOFIT and X-BUILD.

The map management tools convert map formats to QUANTA-compatible brick maps and define active maps, contour levels and display styles. Additional tools save and retrieve the contents of brick map display lists and purge the brick map database of undisplayed bricks to reduce memory requirements.

This chapter describes

- Interactive Map generation (page 25)
- Maps Management table (page 27)
- Set Contour Levels (page 30)
- Cautions about Managing maps in various QUANTA applications (page 34)

Interactive Map generation

The map calculation takes into account any changes in topology or conformation of the molecular system under study. The following checks are made:

If a map calculation is requested before any reflection data has been specified, you are prompted with the map options dialogue.

If only a conformation change has been made and a PSF of the same name exists, then no new PSF is generated.

If a configuration change (add / remove atoms) has occurred, or the molecular system has been re-selected from disk, a new PSF is generated.

To determine whether a PSF can be generated from the molecular system, it is a good idea to click **Generate CNX PSF** in the CNX interface first.

For example, the default atom typing in QUANTA is CHARMm atom types - when working with non-hydrogen proteins, the Engh and Huber are preferable. This can be changed within the CNX interface, or you can use the command: CHAR SET CHAR/ENGH.

The sequence of calculations is as follows:

1. Check whether a reflection file loaded; if not, ask for map options
2. Generate a PSF and parameter file, if needed
3. Create CNX input files (PDB, command file)
4. Run of CNX
5. Convert the map to QUANTA bricked map format
6. If the map generated is not already loaded into QUANTA, a prompt for the contouring details appears.
7. The map is re-contoured according to the current Xfit settings
8. All files generated use the rootname of the first msf selected.

Typical Use

1. Open an MSF of a protein with no hydrogens.
2. Check that it is acceptable for CNX by testing whether a PSF can be generated. You can also change the host on which CNX runs.
3. Select **Applications | X-AUTOFIT** and click **Calculate map**.
4. Select the type of map and define the map extent.
5. Enter the grid size (0.25).
6. Click **OK**.
7. In the Identify reflection file and parameters dialog box, select the reflection file and specify the resolution range.
8. You are then prompted for the contouring options for the map before the map is generated.

9. Do some modelling. Then click **Calculate map** again to recalculate the map.

Maps Management table

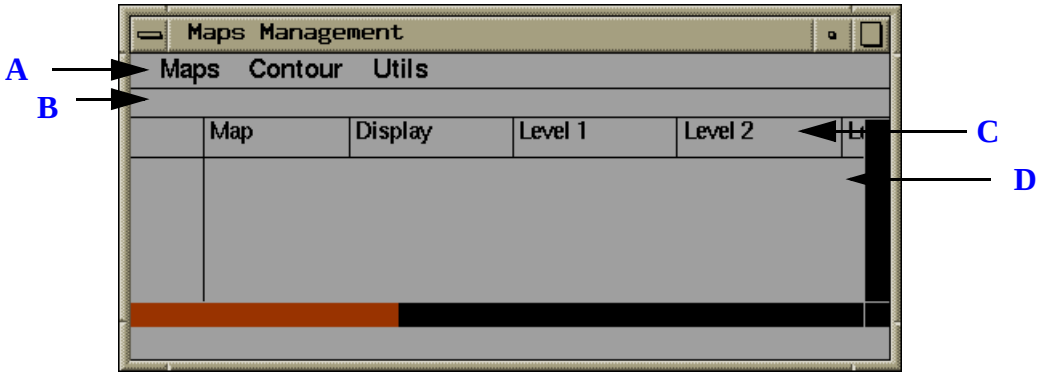
The Maps Management table provides an interface to the general-purpose volume visualization capabilities of the QUANTA program. 3D volume information is stored in brick maps in QUANTA. These brick maps are used for storing electron density, molecular probe maps, volumes, electrostatic potential, and molecular surface information generated within QUANTA. In addition, the maps can be generated by external programs for display within QUANTA. The molecular or volume properties that these maps contain are displayed within QUANTA as either chicken-wire contours or solid molecular surfaces.

The Maps Management table is available from within any application in QUANTA and provides a spreadsheet-like interface to the tools that manage and display the 3D information held in QUANTA brick map files. The display of the table is controlled by the function **Map Table** on the **Draw** menu.

These tools convert map formats to QUANTA-compatible brick maps and define active maps, contour levels and display styles. Additional tools save and retrieve the contents of brick map display lists and purge the brick map database of undisplayed bricks to reduce memory requirements.

The following figure shows the Maps Management table. The table has the following components:

- Its own menu bar of commands (marked A on the figure).
- An edit line in which the contents of the currently selected cell is displayed (marked B on the figure).
- A column heading line (marked C on the figure).
- Rows containing the information for each map in use in the program (marked D on the figure).



Menu bar

The menu bar functions analogous to the menu bar of the main QUANTA molecule window. The menu items perform the following functions:

Maps menu

Import

Imports external map formats and creates a QUANTA-compatible brick map file. Formats available from this interface are:

- **CNX ASCII** — The ASCII format output by the CNX program.
- **Xsight** —A binary map format created and used by programs in Accelrys' Xsight package and the XtalView package.
- **CCP4** — Standard format of the CCP4 suite of programs distributed by the CCP4 group at Daresbury Laboratory in England.
- **GRID** — Map format of the GRID program developed by Dr. Peter Goodford.
- **Ten-Eyck** map format.
- **DSN6** — A dsn6 standard Frodo file generated on a UNIX system.
- **VSN6** — A dsn6 standard Frodo file generated on a VAX/VMS system and copied to the UNIX system. The difference between DSN6 and VSN6 is the swapping of bytes.

After you select a map type, a File Librarian is displayed from which you select the file you want to import.

Selecting **Import Map** spawns a process to run the program $HYD_MAP/mbkall in the QUANTA software. The source for this program is available for user modification in the $HYD_MAP directory.

Add a Map

Displays a File Librarian listing all available files with an.mbk extension. A QUANTA brick map file can be selected and displayed even if it was previously selected. This option is useful for displaying the map with more than one display style.

When a file is selected, the file header is displayed in the textport and the Define Contour Levels and Characteristics dialog box appears. In this box, you can specify one to six contour levels, line style and thickness, and the value and color of each contour level. For information on this dialog box see the Set Contour Levels section.

Delete a Map

Deletes a map from the active list and deletes all bricks generated from this map both from the display list and the brick map database. If only one map is in use, that map is automatically deleted. If more than one map is in use, a scrolling list appears permitting you to select the map to be deleted.

Replace a Map

If more than one map is in use, a File Librarian is displayed from which you select a map to replace the deleted one.

List Maps

In the textport, lists the active brick map files. The listing includes the following information:

- Name of the brick map file.
- Header information in the file.
- Assigned line style and width.
- Specified contour levels and colors in which they are displayed.

Hide Table

Hides the table. The table can be redisplayed by selecting the **Show Map Table** from the top-level **View** menu.

Contour menu

Contour Maps

On Displayed Atoms

Contours the map at the currently specified contour levels for the currently displayed atoms in the viewing area.

All the Map

Contours the entire brick map at the currently specified contour levels. Contouring can be interrupted at any time by clicking the mouse button. This is useful if you select this option by mistake for a large map.

In a Volume

Displays the following dialog box for entering the dimension of a cube as illustrated. After you specify this value and click **OK**, the brick map is displayed as a cube with the specified dimensions.

Contour Mode

Controls preservation of currently displayed bricks when a new set of bricks is calculated. You can choose whether the current selection of bricks is cleared and replaced when a new selection is specified or whether subsequent bricks are added to the current selection.

Set Contour Levels

The Define Contour Levels and Characteristics dialog box is used to change map specifications for currently defined maps. If more than one map is in use, a scrolling list of current maps is displayed from which to select one.

The number of currently defined contour levels can be changed in this dialog box. If the contour levels are changed for currently displayed bricks, the displayed bricks are deleted from memory and the brick map database.

Characteristics of Display Lines

Line Width (0 to 4)

Sets the thickness of the contour lines in pixels. The default is zero. System limitations prohibit setting line thickness greater than 1 for anti-aliased lines for some graphics boards.

Drawing Style for Lines

Sets the line type to solid, dotted, dashed, or dot-dash. The default is solid. When anti-aliasing is on in QUANTA, all non-solid line styles are displayed as solid lines.

Display Map after Contouring

Toggles the display of the entire map. If this option is off, contours for newly selected bricks are not displayed.

Characteristics of Density in Map

Displays the current minimum and maximum densities and the sigma value for each map. These values are not editable. The sigma values aids in assigning contour levels.

Define Contour Levels

Include

Determines whether a contour is calculated. As many as six contours can be calculated per map regardless of its display status. By default, two levels are calculated.

Display

Specifies whether the individual contour level is to be displayed initially. Each contour can be displayed in a different color.

Level

Displays the value for a calculated contour.

Color

Defines the color to be used for the contour level. Any of display colors 1 through 14 can be entered.

OK

Accepts changes and exits the dialog box.

Range of Levels

Automatically sets the level of included contours to values distributed between the minimum and maximum values of the map. This selection only affects contours that are included. Several contours should be marked as included before you choose this selection.

Levels from Sigma

Automatically sets the level of included contours to values starting at zero and incrementing in steps of sigma. If the sigma value of the map has not been calculated or is set to zero, this option does not work. Several contours should be marked as included before you select this option.

Cancel

Exits the dialog box with no changes.

Options

Displays a dialog box in which the parameters used to select the bricks of density to be displayed are set.

Extra Map Radius

Is an additional radius added to the coordinates of atoms when selecting bricks on the basis of the currently displayed display.

Mask map to only cover atoms

Toggle can be used to control the option to delete portions of the data within a brick of density that are beyond a certain distance from any atom. This can be used to remove extraneous density that is more than the **Cover Radius** from an atom. This mask map option should be used with great caution, since it may remove the display of pieces of density that indicate that the model may need further refinement or rebuilding.

Suppress contouring message

Controls whether progress-monitoring messages are displayed in the textport when maps are being contoured for display.

Utils menu

Purge Contours

Deletes all undisplayed contoured bricks in system memory. This tool is useful if many large maps are contoured, and the system's resources for graphical object management are depleted. Other objects that may compete for these system resources are dynamics trajectories and dot surfaces. Tools to remove these objects from memory are found in the associated applications.

Reset Drawing

Resets the map contour line width and style for all maps to 0.

Edit line

Shows the coordinates and contents of the currently selected cell in the Maps Management table. The contour level values can be edited here—this is a shortcut that removes the need to enter the Set Contour Levels dialog box. Once changed, the contours corresponding to the previous contour

level are deleted. The map can then be recontoured using other QUANTA commands, such as Contour Maps (page 30).

Column headings

This row of the table contains the headings for each column of the table. The following table explains what each column contains and what action follows selection of the heading:

Column	Contents	Column header pick
Map	The name of the map truncated to the final 11 characters.	Lists all details of all maps in the tables in the textport.
Display	A yes/no toggle controlling the display of the individual maps.	The first pick switches all maps to their consensus state (the display state of the majority of the maps). Subsequent picks toggle between displaying or hiding all maps.
Level 1 … Level 6	The map value at which the 6 contour levels are made.	No effect.
Width	The line width used to draw the vectors representing the contours of the map.	Selecting this cell switches the line width for all maps to that of the first map + 1. Subsequent picks increment the line width of all maps up to the limiting value (set with the SET LINE MAX command).
Style	The line style used to draw the vectors representing the contours of the map.	Selecting this cell switches the line style for all maps to that of the first map + 1. Subsequent picks increment the style in the sequence 0,1,2….
Map	A repeat of the name of the map (same as first column). It is included in the table so that when the Maps Management Table is quite narrow, the line width and style can be changed with the name of the map still visible.	(Same as first-column Map.) Lists all details of all maps in the tables in the textport.

Rows

Each map selected for display in QUANTA is represented with a different line. A maximum of six maps can be selected.

Each cell in a row contains information about the map or contour level displayed. Picking each cell performs a particular task, as described in the table below:

Column	Cell pick
Row Number	Picking this lists all available information about that map to the textport.
Map Name	Displays the Define Contour Levels and Characteristics dialog box (see Set Contour Levels)
Display	Toggles the display of the contours for that map on (**yes**) or off (**no**).
Level 1–6	Picking a level toggles its display on and off and selects that cell for possible editing in the edit line of the table.
Width	Increments this map's contour line width by 1, up to the maximum (set with the SET LINE MAX command), and then back to 1.
Style	Increments the line style for the contours of this map in the sequence 0, 1, 2, 0. etc.

Managing maps in various QUANTA applications

The control of maps in QUANTA is available in all parts of QUANTA.

Some tools available in the Map palette and the Maps Management table are *not* relevant within X-LIGAND, X-SOLVATE, X-AUTOFIT, X-POWERFIT, and X-BUILD. These applications automatically control the map display; so changes to some general map options can interfere directly with the tools' use of the maps, with serious detriment to speed and results produced. These include:

X-LIGAND, X-SOLVATE:

Do not use Map palette or Maps Management table commands.

No commands from the Map palette or Maps Management table should be used while within these applications. These applications are entirely automated and control all map functionality directly. Open all maps *before* entering these applications.

X-POWERFIT, X-AUTOFIT: X-BUILD:

Do not use the commands found under **Map | Contouring options** or **Maps Management table | Contour | Options** while in X-AUTOFIT: X-BUILD.

These applications check the status of the open maps before every command, so all the Map palette or Maps Management table commands can be used with these applications. In particular, opening, closing, and changing contour levels is important for use of X-AUTOFIT and X-BUILD.

The map options that should *not* be used while in X-AUTOFIT and X-BUILD are those that affect how the map is actually contoured. This is because X-AUTOFIT and X-BUILD use a streamlined process of map contouring where a box of density is automatically generated around the working position, and the only change allowed is to the radial size of this box. This is controlled under **X-AUTOFIT: X-BUILD | Options | Map radius**.

The use of the following commands is strongly discouraged. In particular the identical commands found under **Map | Contouring options** or **Maps Management table | Contour | Options** *must not be used* while in X-AUTOFIT:X-BUILD because:

1. Small values of **Cover radius** produce false contours (especially for difference density) and result in incorrect bones and incorrect solutions to real-space refinements.
2. These commands significantly slow down the contouring process, because contouring by algorithm is more complex than simple contouring.
3. These commands do not work if no coordinates are present.
4. These commands may produce other unwanted side effects.

The following table shows which contouring options can or cannot safely be used in X-AUTOFIT:X-BUILD:

Map…		
Contouring mode	Ignored by X-AUTOFIT:X-BUILD	
Contouring options	MUST NOT BE USED (see reasons above)	
Maps Management table | Contour		
Contour maps	On displayed atoms	Can be used, but any subsequent repositioning of the display center (i.e., **Pointer | Go-to-pointer**) will reset the map display to the defined sphere.
	All the map	Can be used, but any subsequent repositioning of the display center (i.e., **Pointer | Go-to-pointer**) will reset the map display to the defined sphere.

managing maps

	In a volume	Can be used, but any subsequent repositioning of the display center (i.e., **Pointer | Go-to-pointer**) will reset the map display to the defined sphere.
Contour mode	Replace density	Ignored
	Add density	Ignored
Options	Extra map radius	This is the same value as **X-BUILD | Options | Map radius**. Changing this value is the same as changing the map radius in the **Options** menu.
	Map mask to only cover atoms/cover radius	MUST NOT BE USED (see reasons above this table).

4. Introduction to X-AUTOFIT/X-BUILD/X-POWERFIT (X-FIT)

Fitting coordinates to an SIR, MIR, or MAD map can be difficult. X-POWERFIT, X- AUTOFIT and X-BUILD (together called X-FIT) are QUANTA applications that speed up and enhance the process of de novo map interpretation, as well as the general model building in later stages of macromolecular refinement.

X-POWERFIT, X-AUTOFIT and X- BUILD capabilities include:

- Fully automated CA tracing.
- Determine secondary structure from map.
- Intelligently place alpha carbons into the density.
- Carry out sequence assignment.
- Automatically build residue coordinates to fit the electron density map.
- Powerful molecular coordinate modeling and editing.
- Refinement using grid, gradient, and Monte Carlo algorithms.
- Validation and protein structure assessment.
- Calculate bones.
- Generate solvent masks.

Map generation

X-AUTOFIT permits generation of maps on the fly from within the X-AUTOFIT X-BUILD main palette. After selecting the reflection data file and generating the first map, you can add atoms/residues or rebuild the existing molecule (psf file). You can then generate a new map using CNX with a single click.

Map contouring

X-AUTOFIT permits up to six maps and, for each map, seven contour levels to be active at one time. It uses one of the currently open maps as the

basis for the bones and real-space calculations. Bones are calculated from electron density maps.

Density skeletonization

The skeletonization algorithm used in X-AUTOFIT is based on the original rules described by Greer (1974). The four rules were modified and the algorithm reimplemented to incorporate improvements in the speed of calculation and the quality of the resulting bones. Mainchain and sidechain bones are determined by analysis, and a spline function smooths the bones segments to improve interpretability.

Solvent masks

X-AUTOFIT generates solvent masks from coordinate and bones data. The solvent mask facility uses fast algorithms to determine solvent boundaries from coordinate information or bones (and, hence, map) information. The solvent mask can be interactively edited with a spherical pointer, and voids within the mask can be automatically deleted with a single tool.

3D text editor

The 3D text editor can be used to place annotations throughout the macromolecular structure. After notations are created, you can select a notation from a list and the display re-centers on the associated point in the macromolecular structure. You can also load information about the macromolecular structure into the note text utility, thereby allowing you to rapidly find problem areas during crystallographic model building.

Automated Cα tracing

X-POWERFIT provides an algorithm to determine the Cα trace directly from the electron density for high resolution data. For lower resolution (4.0 Å - 2.0 Å) data, there is an algorithm to determine the secondary structure from the electron density, to automatically place Cα atoms into these parts of the density, and then to extend the Cα trace until the model is close to complete. There are also algorithms for Cα refinement.

Cα tracing

X-AUTOFIT also provides semi-automated Cα tracing and manipulation. The alpha-carbon (Cα) building facility in X-AUTOFIT intelligently places Cα coordinates into the electron density map using a rule-based process. It also allows cut-and-paste of fragments and manual editing of Cα atoms. X-POWERFIT can take the current Cα trace and search a set of pdb files for matching geometry.

Automated sequence assignment and fuzzy logic assignment

Once the Cα tracing is complete, you can automatically assign the sequence to the Cα trace in cases where there is high resolution data and there are aromatic residues in the structure. In cases where the automated assignment does not work, or there are no aromatic residues in the structure, you can also carry out a fuzzy sequence assignment (such as: big, aromatic) for each residue of the Cα fragment(s). There is also an option to explicitly assign one of the 20 amino acids to a Cα atom. QUANTA uses this to show a weighted forward and backward alignment to the sequence. A secondary structure prediction from sequence tool returns the secondary structure prediction of a sequence and colors the sequence view on screen using red for helices and blue for strands.

Cα trace --> all-atom model

X-AUTOFIT can create an all-atom representation using refinement techniques, database fragment fitting, and direct correlation of the Cα conformations to all-atom models. These three techniques can be used within X-AUTOFIT to fit the atoms of a residue to the map from just the Cα positions of the Cα trace fragment. The quality of fit is reported by color coding atoms in the fitted segment.

Model building

Model building is carried out with the aid of real-space refinement, regularization, and rigid-body refinement algorithms, as well as by traditional

manual editing. An automated tool (**Structure | Auto build**) that runs through all build residues and fits them using a new mixed grid and gradient protocol. The automated tools of X-BUILD generally give a 200-to-500-fold decrease in time for a model building session and often result in improvement in the subsequent refinement of coordinates compared to traditional manual model building.

Refinement techniques

X-AUTOFIT supports three refinement techniques. Single residues can be fitted by grid searching about torsions, torsion angle real-space gradient refinement, and Monte Carlo fitting. X-AUTOFIT is designed to support the following special cases associated with electron density proteins fitting:

- Alternative conformations (disorder).
- B-value and occupancy editing.
- Polypeptide chain editing, allowing changes in connectivity.
- C-terminal oxygen atoms.
- Rebuilding incomplete amino acids.
- Non-hydrogen, polar hydrogen, and all-hydrogen representations.

Validation techniques

X-BUILD supports two kinds of validation tools designed for the crystallographic process.

The first is an entirely automated system where common errors associated with model building are detected and can be automatically fixed.

The second method of validation provides a set of functions that can derive data from molecular coordinates, apply further functions to them, and then plot them in a graph window. The molecule display, graph display, and tables are integrated, so that the table and graph can be picked to update the molecular display.

Data logging

All X-BUILD tools are logged automatically in a table of previous commands, which can be used to create a log book, undo/redo any edit, assess the use of the application, and even recover and analyze changes from previous model-building sessions.

Data integrity

An automatic save of the current state is made every 5 minutes when in the X-POWERFIT, X-AUTOFIT or X-BUILD modules. This is in addition to the facility for user-directed saves.

X-AUTOFIT in QUANTA

Implementation of the X-AUTOFIT tools in QUANTA provides an integrated environment in which many associated tools in Protein Design, Protein Health, and Conformational Search can be combined with features of the X-AUTOFIT application to enhance the model-building and refinement process.

The rest of this chapter and the four that follow describe X-AUTOFIT, X-BUILD, and X-POWERFIT and their operation within QUANTA. These sections explain:

- Accessing X-AUTOFIT (page 43).
- General X-AUTOFIT and X-BUILD behavior (page 44)

Memory requirements for X-AUTOFIT

QUANTA has been designed to handle maps of up to 500 grid points in the x, y, or z dimensions. This size represents a maximum brick map size of 125 MB or, when a map is stored as real numbers (as in a CCP4 map), 500 MB, which is a very large map. The size of map that can be handled in X-AUTOFIT is defined by the virtual swap space of the system running QUANTA, since all map information and derivatives of these are dynamically allocated as required. A virtual swap space of at least 300 MB (preferably 400 MB) is best for the use of X-AUTOFIT/X-BUILD. This amount

of virtual swap space allows the use of maps of up to 20 million grid points and any bones and mask information derived from this map. Of this space, QUANTA will use about 55 MB and the operating system approximately 30 MB. The remaining space can be allocated by the X-AUTOFIT application or used by other programs. If this happens, then you should leave QUANTA after saving any changes and restart the program. This returns all the swap space allocated for map manipulation to the system. It is *not* enough to just restart QUANTA from the **File** menu, you must exit the program.

You can display a pie chart of memory usage. There is no interface option for this, so you need to edit the xfit.pack file. The line containing "MemS 0" should be changed to "MemS 1" while the X-AUTOFIT:X-BUILD *is closed*. Upon entering the X-RAY application, a pie chart shows current memory usage on the computer. When available memory drops below 50Mbytes the pie chart turns from green to orange, and when available memory drops below 20Mbytes the pie chart turns red.

A *real* memory size of at least 48 MB is recommended for using X-AUTOFIT. If large structures are to be studied, 64 MB provides a more responsive program.

> **Note:** All requests for allocated memory for maps are checked to see if they can be satisfied, and if not, the requested functionality or calculation is aborted and a message printed to the textport.

Example message:

```
MEM_Alloc: Allocation error [-27316]
==============================
A request for memory was made:-
memory =   27 Mbytes
But the system refused this request
Either : Increase system swap space
         Use a small map
         Finish other programs running
==============================
```

Graphical objects generated under SGI GL and Open GL cannot have their memory returned to the system after use. This means that continued re-display of different areas of the map gradually uses the available swap space when very large maps are used. X-AUTOFIT therefore continually monitors the available swap space and prints warning messages to the textport if the available space drops below 10 MB.

Example message:

```
=========================
You are strongly advised
to save and exit QUANTA as
swap memory is low
The current OS does not allow
this to be fixed
=========================
```

Since it is impossible to prevent a program from being killed by the system when all the available swap space is used, you should take one of the following actions if this message appears:

a. Close other programs on the system to release memory.

b. Delete bones and/or delete a map mask in X-AUTOFIT.

c. Close the current map and use a smaller map.

d. Finish editing, save all changes in X-AUTOFIT, and exit QUANTA.

The memory tests in X-AUTOFIT help prevent crashing due to memory depletion.

Accessing X-AUTOFIT

This section describes how to access and get started with X-AUTOFIT.

Before you begin

Prior to starting X-AUTOFIT, do the following:

- In the QUANTA molecule window, if you can, display any known coordinates or previously built coordinates of the protein molecule you are studying.
- Have an electron density map of your molecule available in brick map format.
- Have a standard QUANTA sequence file available. This file contains a one-letter code for each protein residue.

For learning purposes, start with a 2fo-fc electron density map with a corresponding set of coordinates. You may have to convert an existing map using the map conversion facility on the Maps Management table and palette. For more information, see Managing Maps (page 25).

To start X-AUTOFIT

X-AUTOFIT is accessed from the **Applications** menu in the QUANTA main menu, as follows:

1. Create a new directory in which to run X-AUTOFIT. Move to that directory and start QUANTA.
2. Select **X-AUTOFIT** from the **Applications** menu. The main X-AUTOFIT:X-BUILD palette and the Pointer palette appear, as well as a graph window containing the allowed regions for a Ramachandran map and any points representing the phi-psi angles for any known coordinates. If the map table is not open, then open it using the **Map show table** (select **Map table | Show map table** from the **Draw** menu) tool.

   An Object Management table also opens. As objects are generated, they are added to the table: bones, mask, symmetry atoms, Cα trace, and 3D text.

   The Object Management table can be used to toggle the relevant information on and off. Objects can be deleted from this table, but X-AUTOFIT generates them again if required.

Note: If you exit X-AUTOFIT without using **X-AUTOFIT | Finish**, the next time you start X-AUTOFIT, a prompt appears asking if you want to recover from the last building session. If you select **Yes**, this recovers the changes made to the coordinates from the last building session.

General X-AUTOFIT and X-BUILD behavior

The following description indicates the aims behind X-AUTOFIT and generalizations about its use.

Using multiple structures and maps in X-AUTOFIT

X-AUTOFIT allows only one structure to be edited at any time, to prevent possible confusion and ambiguity. The current editable molecule is the first active and displayed molecule in the object management table. To change the molecule that you are editing, use the Molecule Management table to change the activity of the molecule.

QUANTA can use six active and contoured maps at the same time. Because X-AUTOFIT uses crystallographic information for real-space refinement and bones, you must identify which map to use if more than one map is open. To change the current map, use the tool **X-AUTOFIT | Change RSR-bones map**.

By default, if any maps are open when you first enter X-AUTOFIT, the first map that you open is the map used for RSR and bones calculations.

Palettes and tool activity in X-AUTOFIT

X-AUTOFIT differs slightly in the use and behavior of palettes from the rest of QUANTA. Twelve palettes can be open in X-AUTOFIT, and all can be open and used simultaneously. This allows almost any type of manipulation and calculation necessary for macromolecular crystallography, while simplifying the number of commands visible at one time.

All tools in the X-AUTOFIT application are active and can be used from any of the palettes at any time, with these restrictions:

1. When an editing process is active, (when the Accept or Quit dialog box is present), **Accept** or **Quit** are the only options you can choose.
2. X-SOLVATE and X-LIGAND are separate applications and do not allow other X-AUTOFIT tools to be used while they are active. Because they are automated procedures, they take control of screen placement and map display.

The main X-AUTOFIT:X-BUILD palette has ten sub-palettes, the X-AUTOFIT | Sequence palette has two sub-palettes, and the X-AUTOFIT | Build palette has two sub-palettes. The Accept or Quit dialog box appears when the application requires you to choose whether to accept or abort the current editing or picking process.

The general use of X-AUTOFIT can be divided by activity. The following table indicates the palettes required for a given operation. You can have more palettes open at a time.

Activity	Palettes
Defining solvent boundaries	X-AUTOFIT | Map mask
…	X-AUTOFIT | Bones

Activity	Palettes
Cα tracing to density …	X-AUTOFIT | CA Build
	X-AUTOFIT | Bones
	X-POWERFIT
Sequence assignment of Cα trace…	X-AUTOFIT | Sequence
	X-AUTOFIT | Sequence | Show Hide amino-acids
	X-AUTOFIT | Sequence | Show Hide fuzzy
Building all-atom representation from Cα trace…	X-AUTOFIT | CA Build
Model building	X-AUTOFIT | Build atoms
	X-AUTOFIT | Build atoms | Color atoms
	X-AUTOFIT | Build atoms | Add/delete
	X-AUTOFIT | Structure
Validation	X-AUTOFIT
	X-AUTOFIT | 3D-text
	X-AUTOFIT | Tables and Graphs

A user-defined palette is also provided that allows various tools from different palettes to be put in one place. This also allows you to use macros.

The Text and Pointer palettes can be used at any time, and normally the Pointer palette should be open at all times. Each palette is made active by selecting the tool on the parent palette and made inactive by selecting **Hide this menu** at the bottom of each palette. If a palette is active then the parent tool is highlighted. If the parent tool is highlighted but the associated palette is covered by another window, selecting the highlighted tool for that palette brings the palette to the front.

The X-AUTOFIT:X-BUILD main palette contains three other tools, for saving the results of the Cα-tracing, saving the results of the model building, and restoring the old MSF from disk. When you exit X-AUTOFIT, it writes a Cα trace session file to disk. It also writes this file when the **X-AUTOFIT | CA Build | Save changes** tool is used. This file also contains the sequence for the alignment and any sequence assignment made to Cα atoms. The data is read back into X-AUTOFIT the next time the program is used. X-AUTOFIT also checks to see if any changes have been made to any molecular coordinates, and a dialog box asks you to save each changed molecule before leaving X-AUTOFIT.

When you exit X-AUTOFIT, it stores the current status of parameters and settings and the current open palettes, so that a new session starts from where you left off.

Dial box

There are four default sets of dials in X-AUTOFIT. While using the edit functions, the dial set appropriate to the requested function is displayed. Because some functions can use more than one dial set, there are options on the Pointer, Mask, and CA Build palettes to set the type of dial in use.

Two of the default dial sets are used to adjust the position of Cα atoms while fitting Cα atoms to MIR/SIR or MAD density. These control the position of the current active Cα atom.

The third dial set controls the placement of a cursor on the screen. This cursor is normally used for all other building procedures not involving Cα-atom tracing into density. The selection of the dial set also controls the type of plot displayed in the graphs window, as described in the next section.

The fourth dial set controls the mask pointer. This set allows you to position a spherical pointer and the adjust its radius. It becomes active after the calculation or reading of a mask. The **X-AUTOFIT | Mask | Mask dials** tool changes the dials to the mask dial set, if a mask is currently active.

The **X-AUTOFIT | Pointer | Pointer-dials** tool changes the dial set to the pointer dials, and the **X-AUTOFIT | CA Build | CA dials** tool changes the dials to one of the Cα-building dial sets relevant to the current active Cα atom.

The default X-AUTOFIT dial set fills the U dials and by default is always displayed while using X-AUTOFIT, X-POWERFIT, or X-BUILD. A toggle on the X-AUTOFIT:X-BUILD | Options... dialog box can be used to reset the default dials to that of the main “1” dial set of QUANTA if preferred.

Graph windows

Two graph windows are generated in X-AUTOFIT. The first is always present in the lower right corner of the screen. This graph window displays three sets of information.

- If the CA Build dials are active, a contoured plot of Cα torsion and angles observed in the protein databank is shown. A pointer indicates the torsion and angle of the currently active Cα atom.
- If the pointer dials are active and nucleotide structures are being edited, then a DNA concentric circle plot is displayed. This shows all the backbone torsion angles and the base torsion for all the nucleotides. If any nucleotide is edited, then just the information for the residues edited is shown.

- If the pointer dials are active and protein structures are being edited, then a Ramachandran allowed region plot is displayed with all the phi-psi angles of the current active and displayed molecule. During any editing procedure that changes specific phi and psi angles, only these angles are displayed, and they move as the residues are edited.

The second graph window is generated as part of the advanced validation tools and can display many different types of data from the validation tables. Details on the use of this graph, generation, annotation, and plotting can be found in Advanced validation techniques (page 85). The tools are described in more detail in Tables and Graphs (page 190).

Tables

Four tables can be created from X-AUTOFIT | X-BUILD. The first is the last-command table. The last-command table is opened from the main X-AUTOFIT | X-BUILD palette and provides a catalogue of tools used, time and date, undo/redo option, and user-defined remarks. The menu items at the top of this table allow various actions to be carried out based on the last-command table

The remaining three tables are generated while using the advanced validation tools. One table is generated for any property or function based on atom data, the second is generated for residue data, and the third when data is produced that has no direct relation with coordinate information. The atom and residue tables can be picked to place the molecular view and selected (by row or column) for the various manipulations and plotting. Use of the tables can be found in the section on Advanced validation techniques (page 85). The tools are described in more detail in the section on Tables and Graphs (page 190).

5. Using X-AUTOFIT

This chapter describes the use of X-AUTOFIT to generate map masks and bones and their use in de novo Cα tracing. This is followed by sections on automated or semi-automated sequence assignment and building all-atom representations from a Cα trace.

Solvent boundaries or map masks

X-AUTOFIT allows you to define solvent boundaries using either coordinate information or bones (that is, map), information. The mask can also be imported from an external file, and it is possible to read all forms of O format masks into X-AUTOFIT and X-POWERFIT. A program **$HYD_MAP | mbkall** allows the conversion from other formats (i.e., Xsight, CCP4) to an O compressed format file for reading into the X-applications. The masks generated and edited within X-AUTOFIT can be saved as an O compressed format file.

Not only does X-AUTOFIT allow the calculation of map masks, it can use the map mask as an integral part of the de-novo Cα tracing process. The map mask can be used both as a visual boundary, and as a calculation boundary in X-AUTOFIT and X-POWERFIT, removing the problem of building structure in symmetry related molecules.

The mask size is defined by the current volume of map present as defined by **X-AUTOFIT | Options | Map-radius**. This allows a smaller mask to be defined from a map that may cover a whole unit cell. If the mask must fill the same volume as the map, set the map radius to some very large value. The mask will be generated with the same unit cell parameters as those of the map.

If the mask is to be generated from bones, these must be calculated first (**X-AUTOFIT | Bones | Calculate-bones**), but bones are not required for the calculation from atoms. (See the following section for an overview on the use of bones.) In both cases, the mask covers all selected bones/atoms, so use the atom selection tools to delete atoms not required for the mask. If the mask is to be generated from the bones, then use the **X-AUTOFIT | Bones | Symmetry** tool to indicate where overlap of bones will produce clash by symmetry, and edit the bones accordingly.

Parameterization

The initial bones parameterization is optimized for Cα-tracing. These parameters are sensible for the global editing of the bones for masks, but it may be helpful to increase the bones trim level to a larger value as this will remove the smaller unwanted fragments automatically. This is set with the **Bones | Change trim value** tool which will open a dialog box. Set the bones trim value to 5-10 depending on the map. Turn on the bones with **Bones | Bones on-off**.

Editing bones

The bones will now need to be edited to generate a single volume of bones that represents a single connected asymmetry unit. This is usually easier with the map turned off using the map management table. Initially, the quickest method of editing the bones is to remove the largest fragments of bones using **Bones | Delete fragments**. This tool allows you to pick a bones point resulting in the deletion of the entire fragment connected to the bones point picked. At any stage during the editing, deletion of the required volume can be undone using **Bones | Undo last delete**. Only the large fragments should be deleted initially as the smaller fragments can be removed with another tool more efficiently.

If any large fragment of unwanted bones is connected to the main volume required for the mask, then it is possible to cut this off using **Bones | delete 1 section**. The unwanted fragment can then be removed.

The display should now consist of the main volume of bones, plus many smaller fragments of bones. **Bones | delete all fragments** allows you to delete smaller fragments in a single go. This will remove all the fragments of bones that contain less than 2% of the total number of bones points remaining on the display. If a significant number of fragments remain, then a second use of this tool will move fragments less than 4% of the total number of bones point. Each use of this tool will double the deletion threshold and thus remove progressively larger regions of bones. **Bones | ...Reset delete all** resets the threshold. **Bones | delete all fragments** will miss small fragments of bones containing no main chain points, so these will need to be delete with the **Bones | delete fragments** tool.

Bones | Calc bones symmetry allows you to calculate the bones symmetry and study the bones overlap. Where symmetry bones overlaps with real bones (and later with the mask), the bones will need further editing.

Generating a mask

After editing the bones, use the tool **Mask | Calc mask from bones** to generate a map mask. The progress of the calculation is indicated on the message line of the molecular display. The algorithm uses a radial value, which, by default, has a value of 4 Å, that extends the mask beyond the bones. The radial value can be changed with the tool **Mask | Mask delete radius**. On completion of the calculation the mask will appear as a white dot surface.

It is also possible to generate a mask from coordinates. First, open the MSF file (or multiple files) for the coordinates, and then use the mask tool **Mask | Calc. mask from coord** to generate the mask around these coordinates.

Editing masks

The voids in the mask can be removed with the **Mask | check for voids** option and the mask extent adjusted using **Mask | Add mask at pointer** and **Mask | Del mask at pointer**. Save the mask, when complete, with **Mask | Save mask to file**.

Summary

At this point the mask represents a bound region that can be used in subsequent calculations. The tool **Bones | Mask bones by mask**, when active, will delete all bones pointers outside the mask, and hence any calculation based on these bones will be bounded to the molecular mask. The display will reflect this, and any subsequent changing of the view, map radius and parameterization will result in a new set of bones that lie only inside the mask. Adjustment of the mask using the mask editing tools will change the bounding mask in subsequence calculations.

The solvent mask is displayed as a dot surface rather than a net, to reduce the graphics processing. This also has the advantage that generating this dot surface is 10-100 times faster than net contouring, it allows almost interactive recalculation of the surface as you make changes. The number of dots in the surface can be changed; by default, all points on the surface are shown. If the graphics on your system are not fast, or if the surface is particularly large, then the surface dot density can be reduced (**X-AUTOFIT | Mask | reduce-resolution**) in ten steps. The reduction in number of points follows the numerical progression 1/2, 1/3, 1/4, … 1/10. Reducing the dot density of the surfaces increases the refresh rate of the surface during manipulation of the image. The number of dots can be increased again using the **X-AUTOFIT | Mask | increase-resolution** tool.

Bones

The Bones palette is used to set up the skeletonization process. The skeleton can be used with the map mask calculations and Cα-tracing. There are very few tools accessible under the CA Build and X-POWERFIT palettes when the bones are inactive, as the building process requires the presence of the bones. The bones are necessary for the map masking if the mask is to be generated from the bones, and hence from the map information.

The default values for the bones parameterization are reasonable for 2fo-fc maps (start = 1.2 σ), but you may need to change the start value to increase or decrease the connectivity of the bones. The following values should be used as a guide for different maps.

- For 3fo-2fc maps: start = 1.6 to 1.8 σ.
- For 2fo-fc maps: start = 1.1 to 1.2 σ.
- For Sigma A weighted maps: start = 0.8 to 1.2 σ.

The connectivity of the bones and the quality of the map can be judged quantitatively using the **X-AUTOFIT | Bones | Map quality from bones** option.

Displaying and skeletonizing an electron density map

The skeletonization process used in X-AUTOFIT is a three-dimensional data reduction algorithm that removes points from the electron density map. This process follows four basic steps, repeating steps 2 through 4 until no more points can be removed from the map:

1. All points below a threshold are removed.
2. Only edge points are removed from the density.
3. Electron density connectivity is not broken.
4. Electron density chain length is maintained.

Mathematically this process is simple to implement, but it is difficult to make fast enough for an interactive display. The original four rules of J. Greer[1] were modified slightly, and the algorithm implemented from scratch to incorporate improvements in the calculation and memory usage.

Determining map quality

The quality of any electron density map changes with location. Generally, good density can be found in the core, and marginal density in external loops. The perfect map of one molecule contains only one bones tree with only one pathway from the C-terminus to the N-terminus (except for disulfides).

Information about the quality of a map can be determined by the level of over-connectivity measured as false links and the extent of under-connectivity measured as broken density.

To get information about map quality

Select **Map quality** on the X-AUTOFIT:X-BUILD/bones palette. When you make this selection, X-AUTOFIT calculates the number of connected bones trees, the size of each tree as a percent of total bones points, and the number of false links. A tree is a set of branching bones segments.

The information generated when you select **Map quality** is listed in the QUANTA Textport. The list of trees is sorted by size. Generally, a good map has a few large trees and the remaining trees are small (containing less than one percent of the total bones points). A poor quality map has many moderately sized trees.

A good map also has fewer than 50 false links (a 2fo-fc map has fewer than 100 false links because aromatic rings appear as false links). A poor map may have hundreds or thousands of false links.

Use the information about the quality of your map to make decisions about how to optimize it.

Generally, if the bones appear disconnected, decrease the bones start value, and if the bones appear as “spaghetti”, increase the start values. The bones start value of 1.8 σ is recommended for 3fo-2fc maps. It is recommended that the bones display be set to smooth bones as this makes the bones easier to interpret.

It is possible to use the tool **X-POWERFIT | Find sec struct** to determine the quality of several maps. This acts as an interpretive methods of map quality assessment. Refer to the Determining the secondary structure in the molecule.

Improving map quality

To improve map quality, optimize the skeleton for your current map section by adjusting skeletonization parameters, by adjusting branch trimming parameters, by manually removing bones points from the skeleton, or by

changing the type of bones point from main to side change or the reverse. You can make changes interactively, using map quality data to fine-tune the algorithm for a particular section of your map.

Modifying the skeletonization initial cut-off parameter

The **Start** parameter is a cutoff parameter that you can adjust. All density below the defined **Start** value is removed from an electron density map and ignored when a bones skeleton is generated.

Examine your map to determine how to adjust the **Start** parameter. If no connected bones are visible, or if they appear as single dots, the **Start** value needs to be decreased. If the bones appear as spaghetti, the **Start** value needs to be increased until individual segments become evident.

To adjust the Start parameter

This parameter can be modified in two ways: by using either the X-AUTOFIT dials or the tool **X-AUTOFIT | Bones | Bones start value.**

- To adjust the **Start** parameter using the X-AUTOFIT dials:

> **Note:** The X-AUTOFIT dials can be variously used for the pointer, masks, or Cα building. To activate the dials that allow the start value to be changed, use the tool **X-AUTOFIT | CA Build | CA dials**. You can then set up the required values as described here.

1. Click once on either side of the center line on the **Start Value** dial. The start value is automatically modified.

   Each click fully recalculates the bones. Multiple clicks result in multiple recalculations. The **Start Value** is increased by clicking to the right and decreased by clicking to the left. The further to the left or right you click, the more the value changes.

- To adjust the **Start** parameter using the **X-AUTOFIT | Bones | Bones start value** tool:

2. Select **X-AUTOFIT | Bones | Bones start value**.
3. Select **Change start value** on the Bones palette. The Set up bones parameters dialog box is displayed.
4. Enter a value in the starting value data entry box. The minimum value is -10,000 and the maximum is 10,000.
5. Click **OK**. The dialog box is removed and the bones are recalculated. A new skeleton is displayed when the calculation is complete.

Adjusting branch trimming parameters

Branch trimming provides a way to clear short branches from the bones skeleton. Any branch of density less than or equal to the length of the trim parameter is deleted from the bones skeleton and takes no part in any calculation or display. Any single piece of bones longer than twice the length of the trim value is also deleted. The delete value can be varied from a minimum of 0 to a maximum of 50. The default value is 3. The higher the value, the more deletion will occur.

The sidechain detect level affects the number of bones that are assigned to sidechain status. Any branch of the bones skeleton with a value less than the sidechain detect value is assigned a sidechain type. Any trunk of bones that is part of a sidechain subtree structure is also assigned to the sidechain type if its length is less than the sidechain detect value and if the subtree depth is less than one half the sidechain detect value. This parameter can vary between a minimum of three and a maximum of fifty. The default value is eighteen. The higher the value, the more bones points assigned to the sidechain type.

To adjust branch parameters

1. Select **Change trim parameters** on the Bones setup palette. The Trim parameters dialog box is displayed.
2. Enter the values you want to use for the delete level and sidechain detect parameters.
3. Click **OK**. The dialog box is removed and bones are recalculated using the new values for the trim parameters.

Using bones with masks

If the bones are to be used with the mask calculations then it is usual to set a map radius (**X-AUTOFIT | Options | Map-radius**) that gives you a clear view of the entire molecular packing. If the map radius is set to a very large number (for example, 1000) then the entire map will be used to calculate the bones.

If a mask has been calculated, it can be used to automatically delimit any part of the map within all the map interpretation calculations. In general, it would be used to delimit a molecule that represents a single asymmetric unit.

Deleting bones points

Bones points are deleted as multiple points. The tool **X-AUTOFIT | Bones | Delete-1-section** deletes points that lie in a single chain extending either from a branch point or from a terminus to another branch point or terminus. The **X-AUTOFIT | Bones | Delete-fragment** tool deletes all points in a tree fragment that are connected to the point picked. The tool remains active until clicked again. When you delete a branch, the smoothing function changes so some branch points will move.

The tool **X-AUTOFIT | Bones | Delete-all-fragments** allows the deletion of all small fragments, where the threshold is incremented on every subsequent use of the tool.

Although you may delete a section from the bones skeleton, the electron density map is not altered. Any recalculation of bones (for example, through addition of a new contour level or moving to a new bones box) will override the deletion modifications you have made.

To delete bones points

1. Select one of the delete tools on the X-AUTOFIT:X-BUILD palette. When you make this selection, the message prompt at the bottom of the molecule window instructs you to select a bones point using the mouse.
2. Click a point in the section of skeleton that you want to delete. The section, fragment or multiple fragments are deleted. You can undo your last selection by selecting **Undo last** on the X-AUTOFIT:X-BUILD palette.

To change the strand type

You can change the type of a bones strand from sidechain to main chain or the reverse. A status change in a bones section is indicated by a color change for the strand. If you have sidechains hidden, when you change a main chain to a sidechain, the section seems to disappear.

Select the **Main ↔ side** selection on the X-AUTOFIT:X-BUILD palette. The current strand of bones changes type and color. To undo this change, reapply the **Main ↔ side** selection or use **Undo last** on the Bones palette.

Bones and symmetry

The bones should be edited with the **X-AUTOFIT | Bones | Delete-fragment** option to leave only a single molecule/structure. If the bones symmetry is turned on (**X-AUTOFIT | Bones | Symmetry-on**), a reduced representation of the bones is generated by symmetry. Where symmetry-related bones overlap with the bones required for the mask, you must deter-

mine which real bones fragment that, when deleted, will remove the symmetry-related section. Note the symmetry-related bones are not updated during the bones editing, but can be refreshed by clicking on the **X-AUTOFIT | Bones | Symmetry-on** tool again. The symmetry-related bones can be removed with the **X-AUTOFIT | Bones | Symmetry-off** tool. If the fragment of bone to delete is joined to some bones that must be saved, use the tool **X-AUTOFIT | Bones | Delete-1- section** to cut a link between the two parts of the bones. Once the bones have been edited so as to give no symmetry overlaps, a mask can be calculated.

Using bones with Cα-tracing

The recommended map radius (**X-AUTOFIT | Options | Map-radius**) is approximately 9–12 Å for use with Cα-building. This radius allows quick changes in the bones start value and allows the display to be manipulated on less powerful graphical workstations. Also, you can change the start values for the bones calculation and get an almost immediate change in the bones when using this radius of data. The bones display is often updated when carrying out Cα-tracing (**X-AUTOFIT | CA Build | Next bones** box), recalculating the bones again from the map. Therefore, any editing of the bones is lost when the bones are recalculated. It is therefore recommended that rather than edit the bones when using them from Cα-tracing, the parameters for the auto generation and analysis of the bones be adjusted to give the most interpretable results. Use **X-AUTOFIT | Bones | ChangeStart value** to change the bones generation parameter; increasing this value will reduce the connectivity of the bones and vice versa.

Cα-tracing

To carry out Cα-tracing of density, bones must be active. Please read Using bones with Cα-tracing, in the preceding section. This overview assumes that the bones have been turned on with **X-AUTOFIT | Bones | Calculate-bones** and that required adjustments to the bones and trimming parameters have already been made. If there is no obvious starting place in the map for Cα-tracing, it is possible to use the tool **X-AUTOFIT | Bones | Find nice area of map** to search the entire electron density map for a region of electron density that may have meaningful electron density. This is performed by an algorithm that scans cubes of electron density to determine which volume of the map contains the most density above 1 sigma. The map then is centered at this location, and X-AUTOFIT calculates the bones for this region.

Generating Cα segments using assisted carbon building

Alpha-carbon coordinates for a segment of a protein are built from a skeletonized electron density map using assisted (or smart) alpha-carbon building. X-AUTOFIT evaluates the skeleton and projects the most likely placement of each alpha carbon, determining the open (beta) angle and the torsion (gamma) angle of the alpha carbon with respect to the previous four alpha carbons. Central to this evaluation is the use of a pseudo-Ramachandran plot that defines the probabilities of specific alpha-carbon geometries.

To generate a new segment

1. Start an alpha-carbon segment by selecting **New segment** on the X-AUTOFIT:X-BUILD palette. The Pick Density palette is displayed.
2. Find a branch point on the bones skeleton that looks like a piece of main chain with a long sidechain. Remember that C=O looks like a short sidechain, but you want a Cα-R point. Click on this branch point.

   A red cross appears on the picked location. This is the starting alpha carbon. If you are not satisfied with this point, simply click on another location on the bones skeleton.
3. When you are satisfied with the starting alpha carbon placement, choose **Accept Point** on the Pick Density palette. The palette is removed.

   X-AUTOFIT recalculates a new box around this coordinate. The first alpha carbon position is fixed. The moving end of the segment (marked by a yellow line) represents the next alpha carbon, 3.8 Å from the first alpha carbon.
4. Position the next atom by one of several mechanisms:

   Pick a point on the skeleton by clicking the bones (see “Positioning the next Cα atom” on page 62)

   or

   Ask the program to auto-fit the atom by selecting **Next CA** on the X-AUTOFIT | CA Build palette

   or

   Pick a point on the pseudo-Ramachandran plot (see “Using the pseudo-Ramachandran plot” on page 60)

   or

   Use the dials to set the angle (beta) and torsion (gamma)

   or

Click **X-AUTOFIT | CA Build | Guess next CA** to see the different fitted Cα positions from the auto-fit routine.

5. Continue adding alpha carbons until you are in an un-interpretable part of the map or until you reach the edge of the bones box.

   When you have more than four alpha carbons, the utility of the pseudo-Ramachandran plot becomes obvious. The current conformation of the last four alpha carbons is reported. The torsion angle (γ) is along the y-axis, and the open angle (β) is along the x-axis. Double clicking on the plot sets γ + β to the mouse position. These changes are reflected in the alpha-carbon chain displayed in the molecule window.

   Guess next CA on the X-AUTOFIT:X-BUILD palette uses the plot to score different positions on the bones skeleton.

6. When the trace reaches the edge of the currently drawn map, select **Next bones box** on the X-AUTOFIT:X-BUILD palette to load a new bones box. The box centers on the current alpha carbon and draws a new skeleton and map.

7. If the current alpha carbon is misplaced, select **Delete current CA** on the X-AUTOFIT:X-BUILD palette. This deletes the current alpha carbon and makes the previous one active. This selection should also be used if the current Cα cannot be placed on the map.

Watch the plot and use it to try different solutions, especially when you reach patchy density with poor connectivity. If you have a good map, the program does most of the work.

With difficult maps, fitting the first few atoms is the most difficult. It requires trying Cα placements, then deleting them. You can also use the interactive properties of X-AUTOFIT to adjust and modify trim parameters as you work.

At any stage, you can reverse the growing chain and build in the opposite direction until both ends reach regions that cannot be interpreted.

Also, multiple segments of chain can be fitted to each piece of interpretable map and then joined into a single trace.

To add a further segment of Cα-trace use the tool **X-AUTOFIT | CA Build | New segment** again and a new segment of two Cα atoms will appear. The previous segment will be displayed in color 1 (pale green), and the new segment will be color 3 (red), and the current Cα atom will be color 4 (yellow).

Editing segment and Cα atoms

The tool **X-AUTOFIT | CA Build | Current res seg** can be used to select the current Cα atom as the current atom. After you select this tool,

X-AUTOFIT prompts you to pick a Cα atom. Picking a Cα atom causes the following color changes:

- The segment containing this Cα atom will become color 3 (red).
- All other segments turn to color 1 (pale green).
- The picked Cα atom turns to color 4 (yellow).

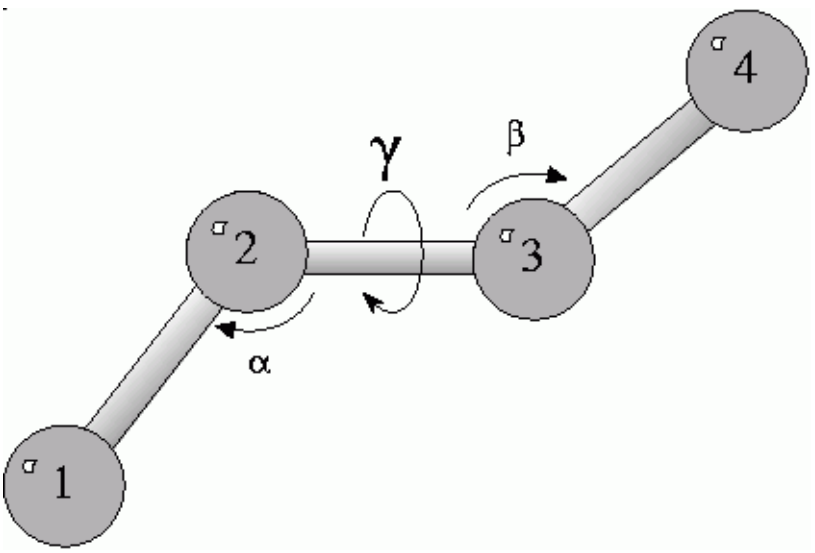
If the new current Cα atom is at a terminus of a Cα-trace segment, the dial box will contain dials to allow the adjustment of the opening angle and torsion relative to the previous Cα atom, and this end of the chain will now become the C-terminus. Any sequence alignment is adjusted appropriately.

If the Cα atom selected is not a terminal atom, the dial changes to allow the movement of this Cα atom in the xyz screen coordinates. The Cα atom has an arrow next to it to indicate the direction of the C-terminus. The chain direction is not changed.

Using the pseudo-Ramachandran plot

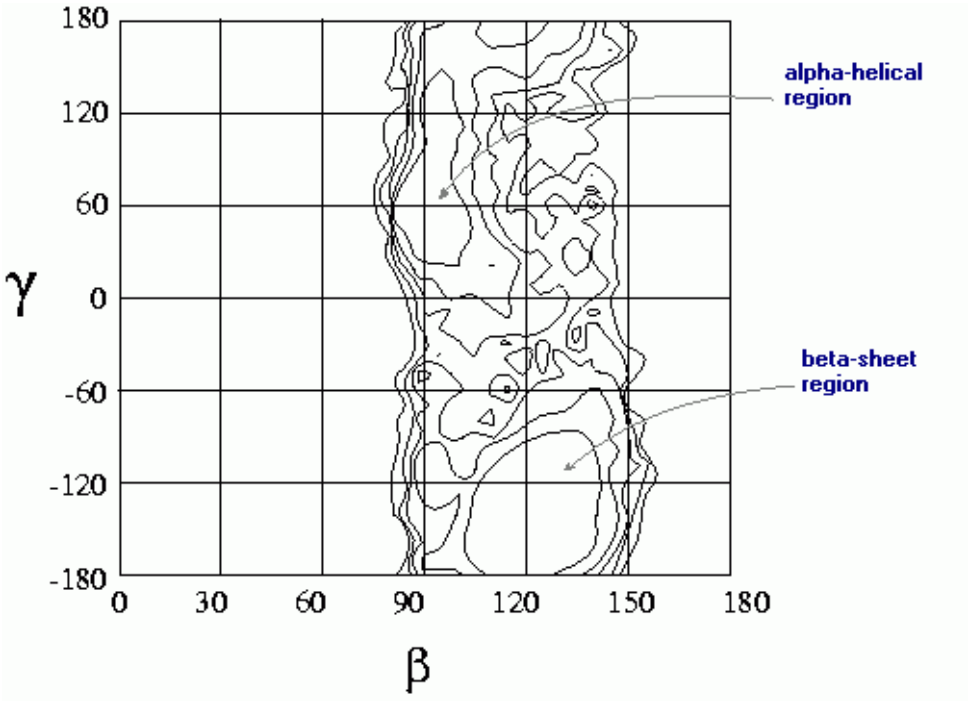
A pseudo-Ramachandran plot is generated for alpha-carbon geometry using well-resolved protein structures. This plot defines the probabilities of specific alpha-carbon geometries and is central to the evaluation process for generating an alpha-carbon trace.

The torsion angle of four consecutive alpha carbons is plotted against the open angle that is defined by three consecutive alpha carbon atoms as illustrated in the following figure:



A probability map of alpha-carbon geometry can then be generated. The resulting plot shows the probability of the alpha-carbon geometry being restricted to certain regions. Different areas on the plot correspond to specific conformations of the protein backbone, including alpha helices, beta sheets, and turn structures. This empirically derived probability surface is used to direct fitting of alpha carbons to the displayed electron density pattern.

The pseudo-Ramachandran plot is displayed in an independent window labeled **CA angle | torsion**. A typical plot is illustrated in the following figure. This plot is generated using QUANTA graph facilities. For more information on QUANTA graph facilities, see Chapter 9 of *QUANTA Simulation, Search, and Analysis*.



A colored pointer (by default, a light-blue oval) indicates the current value of the open angle (β) and torsion angle (γ). As the contoured surface represents observed structure in the protein databank, the position of the pointer on the plot indicates the probability of this geometry occurring in a protein. Its position also designates whether the atoms are being fitted to a helix or β-strand conformation.

The plot provides an interactive tool for monitoring and manipulating the conformation of the alpha-carbon trace. The plot reports the current values of gamma and beta for the current alpha carbon. Alternatively, when a specific β and γ are chosen from the plot by clicking on the selected location, the current alpha carbon is positioned in the appropriate conformation. For example, an alpha helix can be generated by clicking in the alpha-helix zone of the pseudo-Ramachandran plot after adding each new alpha carbon.

Positioning the next Cα atom

Markers are provided to indicate all points on the skeleton that are 3.8 ± 0.3 Å from the current alpha carbon, regardless of the connectivity.

The program has simple logic that positions the next alpha carbon in the best location with regard to the skeleton and alpha-carbon geometry. This positioning is based on the following rules:

- Points 3.8 Å from the previous Cα are linked by continuous skeleton.
- Alpha-carbon geometry is weighted as a function of the alpha-carbon conformation map with respect to the previous Cα atoms.
- A 3.8-Å point has a higher weight if it lies at a branch point on the skeleton. It has a lower weight if it is near a branch point, and a yet lower weight if there is no branch point nearby in the skeleton.
- A main-chain skeleton has a higher weight than a side- chain skeleton.
- The best density path, defined as having the largest mean map value and the highest minimum map value for the path.

The proportion of correct solutions for alpha-carbon placement that the program finds is dependent on the quality of the map. The algorithm places the next atom correctly about forty percent of the time using a map of average quality. The function can significantly accelerate Cα trace building.

Automatic fitting of the next alpha carbon is always carried out when you select **Next CA** on the X-AUTOFIT:X-BUILD palette. The new atom placed by this mechanism is in the best possible position as evaluated by X-AUTOFIT. However, you can override this positioning by moving the current alpha carbon using one of three mechanisms:

- Clicking on any point on the skeleton.
- Using the dials to set the angle (beta) and torsion (gamma).
- Clicking on any position within the pseudo-Ramachandran plot.

You can return to the original best-guess position again by cycling through the auto-fit positions by repeated use of **X-AUTOFIT | CA Build | Guess**

next CA. You also can request automatic fitting whenever you modify skeletonization parameters.

Evaluating and changing segment polarity

When you save a bones skeleton as a set of alpha-carbon traces or when you build a polypeptide, polarity of the segment is determined by the direction of the construction of the trace from the origin to the current alpha carbon (N-terminus to C-terminus).

X-AUTOFIT assesses the probability of the polarity being correct and reports that information in the text port. If you have defined any residue types for any segments in the molecule, sequence alignments are marked in the molecular sequence table at the top of the molecule window. Blue arrows indicate forward alignment and red arrows indicate reverse alignment.

To evaluate segment polarity

To get polarity information for the current segment, select **Check CA direction** on the X-AUTOFIT:X-BUILD | CA Build palette. X-AUTOFIT attempts to fit polyglycine to the trace in both directions, checking geometries. The following information is then reported in the textport:

- The percent fit.
- A statement that the chain is the correct/wrong way around.
- A fit ratio that gives a probability on the correctness of orientation.

> **Note:** The percentage likelihood values for the forward and reverse directions are independent values. The sum will not necessarily equal 100%.

The polarity of the active alpha-carbon trace can be reversed so that building can be carried out at either end of the chain. Evaluate the polarity of the alpha-carbon trace before you generate a peptide backbone.

To change segment polarity

Select **Reverse chain** on the **X-AUTOFIT | CA Build** palette. Any further additions to the segment will occur at the opposite end of the chain. The colors of any arrows indicating sequence alignments for the segment are reversed (that is, red becomes blue and blue becomes red).

Cut/paste Cα segments

You can join two or more alpha carbon segments to build bigger segments and, eventually, to generate a single Cα chain. To use this X-AUTOFIT capability, you must have at least two segments that are within 5 Å of one another but at least 2 Å apart. The newly joined segment becomes the current segment and the current atom is the new last atom in the chain.

The alpha carbons that you pick to define the segments to join must be in two different segments because X-AUTOFIT does not allow cyclic peptides. Also, the alpha carbons must be terminal atoms since branched peptides are not allowed.

If both segments have sequence data assigned to them, the sequences must be in the same direction as defined by the alignment. Also, sequences must be consecutive: no deletions or insertions allowed.

To join two segments

1. Select **Join 2 segments** on the X-AUTOFIT:X-BUILD palette. The message line prompts you to select two alpha carbons from two built traces.
2. Select two close segments by clicking the terminal alpha carbon in each segment. If X-AUTOFIT considers your selection to be reasonable, the two segments are joined.
3. If any sequence information is present in either joined segment, a sequence alignment occurs.

Cutting segments

A segment of Cα-trace can also be cut using the tool **X-AUTOFIT | CA Build | Unjoin 2 CA**. Use this tool to insert a Cα atom and then rejoin the trace with the **X-AUTOFIT | CA Build | Join 2 segments** tool or just to edit some incorrect connectivity. The two new sections of trace are both checked for sequence alignment if some alignment has taken place, (as one may now be too short to be unique), and the C-terminal section becomes the current segment. The sequence alignment table is updated accordingly.

Templates and rigid body editing of Cα traces

The tool **X-AUTOFIT | CA Build | Add helix strand** allows the addition of idealized helix or strands or Cα trace. When this tool is selected, a dialog box appears that allows you to select a secondary structural element of a user-defined length. This secondary structural element can be moved so

that it fits the density. Once placed, this new template of secondary structure becomes the current segment. It can therefore be used just as a template to aid in building a Cα trace accurately to the density, or can extend, by the normal building methods, any user-defined element as part of the structure. The tool **X-AUTOFIT | CA Build | Move current segment** allows the positioning of an already-built Cα trace, regardless of whether it is a secondary structure element or a Cα trace built with the auto build commands.

Sequence assignment

The sequence alignment palette, X-AUTOFIT/Sequence, contains the tool to move around the Cα-trace segment (**X-AUTOFIT | Sequence | Current-res-seg**) and the tools to assign segment information to the Cα atoms.

Reading in sequence information

The tool **X-AUTOFIT | Sequence | Load-sequence** allows you to load sequence information from various format sequence files, including from an MSF or PDB file. If a sequence is successfully loaded into X-AUTOFIT, it will be displayed at the top of the main molecule window in lowercase. Once the sequence table has been loaded, you can assign sequence information to the Cα trace and observe the alignment in relation to this sequence. The Cα trace will be marked with the current sequence assignment of each residue. If the Cα trace is built de novo, using the tracing tools, then all the residues are labeled unknown. If the sequence is read in from an MSF, the sequence information from the MSF is retained and displayed on the Cα trace.

Fitting a sequence to a segment

X-AUTOFIT has an algorithm that automatically searches the electron density map for patterns of aromatic residues and matches these patterns to the aromatic patterns in the sequence that has been read in. The method does not work for sequences that do not contain aromatic residues as it is based on the analysis of these residue positions. If a unique sequence is found, X-AUTOFIT labels each Cα with the name of the amino acid. When no solution is found, QUANTA returns the message: **no unique solution found**. In this case, the semi-automated protocol described below can be used.

X-AUTOFIT has a sequence alignment algorithm that allows you to generate alignment information that matches molecular sequence information of

the structure you are studying with alpha-carbon segments you have generated. The algorithm can be applied after you have labeled at least one residue either specifically or using a fuzzy residue type.

Generating sequence alignment information

With the map turned on, select **X-AUTOFIT | Sequence | Current res. and seq**. You are prompted to pick a Cα atom. When you make the selection, the atom you have selected is colored yellow to indicate that it is the current Cα.

To use the assignment, open the two palettes that allow assignment to Cα atoms. These are accessed as **X-AUTOFIT | Sequence | Show-Hide Amino acids** and **X-AUTOFIT | Sequence | Show-Hide fuzzy**. These palettes provide selections for choosing the residue type of the current alpha carbon. The Fuzzy Residues palette is used to assign fuzzy residues and the Specific Residues palette is used to assign one of the twenty standard amino acid residues. You can make a selection from these palettes at any time.

After selecting an amino acid or a fuzzy descriptor, X-AUTOFIT shows all forward and backward sequence alignments from that residue for the Cα trace. The alignments are displayed as arrows (blue for forward alignment and red for reverse alignment) underneath the aligned residues in the residue sequence table at the top of the molecule window.

The current residue is marked by blue boxes for forward fitting and red boxes for reversed fitting. The current residue box is shown for all alignment arrows when more than five residues have been fitted. The boxes may not be clear at first where there are multiple possible solutions which often overlap.

The sequence alignment algorithm assigns sequences using a weighting system where fuzzy residues are weighted as follows:

Fuzzy Residue	Specific residue																			
	G	A	V	L	I	M	P	F	W	N	Q	T	S	C	D	E	K	R	H	Y
Big	1	2	4	5	5	5	4	8	10	4	6	4	2	3	4	6	6	8	7	9
Medium	1	1	7	9	9	9	7	3	1	7	7	7	3	5	7	7	7	3	5	1
Small	10	9	6	5	5	5	6	2	1	6	4	6	8	7	6	4	4	2	3	1
Aromatic	1	1	2	3	3	3	1	10	10	2	3	2	1	2	2	3	3	5	8	10
Aliphatic	1	1	5	6	6	6	2	10	10	5	7	5	3	4	5	7	7	10	10	10

Fuzzy Residue	Specific residue																			
	G	A	V	L	I	M	P	F	W	N	Q	T	S	C	D	E	K	R	H	Y
Polar	1	1	1	1	1	4	1	1	3	7	8	6	6	4	10	10	10	10	8	8
Nonpolar	9	9	9	9	9	6	9	9	7	3	2	4	4	6	1	1	1	1	2	2
Charged	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	10	10	10	5	0
Acid	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	10	0	0	0	0
Basic	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	10	5	0

Where:

- 0 indicates that a residue is not a member of the set.
- 1 is a bad alignment.
- 10 is a perfect alignment.
- 5 is neutral.

All other numbers fall between these in a continuum.

The algorithm uses weighting for the fuzzy residue types where:

- Big, medium, and small form three ranges and (*big* + *small* = 10).
- Aromatic and aliphatic types form a mutually exclusive list. The aromatic group is judged true or false and the aliphatic group is weighted by the length of sidechain.
- Polar plus non-polar types equal unity.
- Charged, acid, and basic types are judged true or false.
- Unknown residues take no part in the weighting.

These conditions apply to the alignment as a whole:

- If sequence alignment returns a weight of zero, then the alignment is rejected regardless of any other any other residue fit.
- If

$$\frac{1}{N} \sum_{i = 1, N} F_i < 5 ,$$

- where *F* is the fit and *N* is the number of residues, then the alignment also is rejected.
- If

$$\frac{1}{N} \sum_{i = 1, N} F_i \geq 5 ,$$

- then X-AUTOFIT displays an arrow under the alignment sequence.
- The thickness of the arrow is calculated as

  $\frac{1}{N}\sum F_i - 5$.

  Thickness, therefore, varies between 0 and 5 units.

It should be noted that “0” mean complete exclusion

Redefining the weights

You can change the weights table by providing a file in the local directory sequence.weights. This file can contain any number of lines starting with the labels: BIG, MEDIUM, SMALL, POLAR, NONPOLAR, ALIPHATIC, AROMATIC, CHARGED, ACID, BASIC. It is also possible to create new definitions as shown in the next section if a new label is used.

On the same line as the labels, supply twenty values to replace those in the table on page 66. List the values in the same order as the amino acids are listed in the table. For example, to change definition for small residues so that arginine and lysine are also classed as small residues with a high weight (because there is often no density for these amino acids) but still retains the high weight:

Create the file sequence.weights, containing this single line:

```
SMALL    10 9 6 5 5 5 6 2 1 6 4 6 8 7 6 4 9 9 3 1
```

Cha

If you place this file in the current working directory, it will be read on entry to X-AUTOFIT. X-AUTOFIT will then write:

```
Reading user sequence weights
Sequence weights table : 1 re-definitions found
```

to indicate the successful reading of the sequence.weights file.

If you provide a valid keyword, but not enough values, or values less than 0 or greater than 10, the following message will be printed to the Textport:

```
Reading user sequence weights
Sequence weights table :   1 new definitions found
Number of values invalid or incomplete =   1
Set to default values
```

Creating new definitions

You can create a your own tables with a new label if required. For example, a definition for a CB branched residues would allow the specification of valine and threonine, and to a less extent, isoleucine as the same type of residue. To add a new specification, in the file “sequence.weights” create a new line with a different keyword from one of the pre-defined specifications, For example:

```
CB-BRANCH 1 1 10 3 8 1 1 3 3 3 1 10 1 1 3 1 1 1 3 3
```

This specification indicates that all residues branched at CB have weight 10 (for valine and threonine), or 8 (in the case of isoleucine). All residues branched at CG are give a weight of 3, and all other residues are given a weight of 1. You may want to completely exclude glycine from any alignment, in which the first number on the line should be a zero.

To use this new specification, close the **X-AUTOFIT | Sequence | Fuzzy residues** palette if open, then open this palette. The text port will indicate that the new specification has been read:-

```
Number of new definitions =   1
```

A new tool will appear on the fuzzy residue palette called **CB-branched**. This tool can now be used in the same way as any other definition of residue type, and the sequence weights will reflect the definitions provided. The Cα atom will be labelled with the name CB-b (i.e. the first 4 characters of the new definition), and saved in the normal way with the Cα trace.

- You can create up to 10 new definitions.
- The new definitions are invoked on opening the fuzzy residue palette.
- Removal of the new definition from the sequence.weights file when a Cα trace contains these new definitions will results in the Cα trace being labelled with USRn (n=1..9,0).

Finding unique sequences

When you have assigned several residues, X-AUTOFIT may identify a unique sequence for a segment. If a unique sequence is found, X-AUTOFIT shows that sequence in uppercase letters in the sequence table. If you select or start another segment, the unique sequence remains in uppercase letters and the residues are not used in any subsequent alignment.

Three selections on the X-AUTOFIT:X-BUILD palette become active when a unique sequence is identified: **Unique sequence**, **Clear sequence**, and **Return fuzzy alignment**. The Unique sequence selection is highlighted. If you click **Unique sequence** when you move to another segment,

X-AUTOFIT changes the unique sequence code to lowercase letters and makes these residues available for additional sequence matching.

Any action that changes the alpha carbons in the segment with a unique sequence forces a new sequence alignment. For example, if you have an identified unique sequence, delete an alpha carbon--X-AUTOFIT checks to see if there are any other sequence alignment solutions.

Clear sequence removes any sequence assignment and **Return fuzzy sequence** reverses the last unique alignment so as to return the last non-unique fuzzy alignment.

Predicting secondary structure

X-AUTOFIT has a tool that allows you to predict the secondary structure from sequence (**X-AUTOFIT | Sequence | Guess sec. struct**). This tool can be used to check the correctness of the assigned sequence, which will return the secondary structure prediction of a sequence and color the sequence table on screen using red for helices and blue for strands.

Building all-atom representation from Cα trace

Once a Cα trace has been generated, or a Cα trace has been loaded from an MSF file (**X-AUTOFIT | CA Build | Load CA coordinates**) then the production of an all-atom model is a trivial process. There are four methods of generating all atom models from Cα-traces

1. Just using real space refinement (RSR)
2. Fitting the main chain with database fragment fitting (and the sidechains with RSR)
3. Fitting the main chain atoms by direct correlation of the Cα conformation with main chain Ramachandran geometry and fitting the side chain atoms by RSR, and
4. Fitting main chain atoms by direct correlation of Cα conformation with main chain Ramachandran geometry and the side chain atoms using the modeling technique of dead-end elimination.

Which builder to use

Which builder you use to build the all atom model depends on the quality and resolution of the electron density map. If the map quality is good and the resolution is 2.0 Å or better, then fitting entirely with RSR is recommended, otherwise use the Cα-Ramachandran correlation fitting. Building

with database fragments is also available but fitting by RSR or Cα-Ramachandran correlation should yield better results.

How to build

1. Select the segment that you want to use for the procedure, and be sure the chain direction is as required.
2. Select **Fit seg by RSR**, **Fit seg by database**, **Fit seg by CA corr.** or **Fit seg by D.E.E** on the X-AUTOFIT | CA Build palette. The fitting process begins. When the process is complete, the polypeptide structure is displayed over the Cα trace.
3. If you want to eliminate the fitted segment, select Delete fitted segment on the X-AUTOFIT | CA Build palette.

On completion, the coordinates are colored by fit to the electron density. The goodness of the fit is color-coded as follows

- Green atoms fit density well
- Yellow atoms fit density moderately well
- Red atoms fit badly
- Blue atoms have negative density or no density

Building main chain coordinates by RSR

X-AUTOFIT has a real space refinement procedure (**X-AUTOFIT | Fit seg by RSR**) that builds coordinates using the alpha-carbon positions of the current segment as a starting point. Built atoms of other segments are retained. The process involves the following steps:

1. X-AUTOFIT fits polyglycine to the alpha-carbon traces using the real space alignment algorithm.
2. The resulting geometry is checked. Good and bad sections of the structure are flagged. Where the geometry is poor, the program writes a warning to the text port, but continues the fitting process.
3. Good sections of the chain are used as seed points to adjust residues in poor regions for a better fit.
4. Beta carbons are added.

Building mainchain coordinates by database fragment fitting

The tool **X-AUTOFIT | fit seg by database** builds mainchain coordinates based on fragment fitting five residue segments to the Cα trace. The process involves the following steps:

1. For each fragment of five Cα atoms in the Cα trace built, the tool searches the Cα distance matrix for equivalent Cα conformations observed in the protein databank.
2. The best three conformations are then fitted by least squares refinement to check for mirrored solutions, and the best solution selected.
3. The overlapping five residue segments are merged to improve the main chain connectivity, and the polyalanine model coordinates are built directly from the merged fragments of database coordinates.

Refer to Creating a Fragment Database (page 331) to set up the Cα distance matrix for the first use, since this is not supplied by Accelrys.

Building the mainchain coordinates by Cα direct correlation

The tools **X-AUTOFIT | Fit seg by corr.** and **X-AUTOFIT | Fit seg by DEE** build the mainchain atoms by direct correlation of the Cα conformations found in the entire protein databank and the equivalent Ramachandran values.

1. The program reads in the Cα-Ramachandran correlation matrix if not already read into memory.
2. For each four-residue fragment of the Cα trace, the tool determines three parameters that describes the Cα trace of four Cα atoms.
3. The three parameters are used to look up the equivalent values of four Ramachandran angles from the correlation matrix.
4. The four-residue fragment is built with the Ramachandran angles from the matrix
5. The four-residue fragments are merged.

Building the sidechain coordinates by RSR

The polyalanine coordinates are used as a basis of adding the sidechain atoms by adding the atoms by RSR to the map.

1. If the resulting chain has a reasonable geometry, sidechain atoms are added progressively, changing χ angles and adjusting open angles.
2. If the main chain geometry is still poor, sidechain atoms are added but flagged as unfit.
3. At the completion of the process, atoms are colored by fit using a green-yellow-red scale (green is good, yellow is intermediate, red is poor). Any residue sidechains which are in areas of zero density (no map) or not refined are colored blue.

The refinement process refines alpha-carbon coordinates to full atom representation. It refines atoms, but does not affect the electron density map.

The building algorithm builds one segment at a time (**X-AUTOFIT | CA-Build | Fit seg by RSR**) where a segment is defined as the currently active Cα trace segment. This will be colored as color 2 (usually red). For a successful build, the map must cover the entire current Cα segment, since the atoms are built into electron density. Once the "first" segment is selected (**X-AUTOFIT | CA Build | Current res seg**), and the map is extended to cover the entire volume of the Cα atoms, you should turn the map off. During the building process, the progress is displayed. So turning off the map will eliminate the need for the continual redisplay of the map.

To begin building, select the tool **X-AUTOFIT | CA Build | Fit seg by RSR**. The algorithm first generates a backbone trace using electron density fitting, and then correlated geometry analysis of areas where the density is too poor to give correct conformation from the map fitting. This is almost instantaneous. Next, each residue is fitted to density by progressive torsion angle searching to the map, and the progress of this calculation is displayed as each residue is fitted. On completion the coordinates are colored by fit to density. The goodness of the fit is color-coded:

- Green atoms fit density well
- Yellow atoms fit density OK
- Red atoms fit badly
- Blue atoms had negative density or no density

Since each new set of coordinates is added to the QUANTA data structure in the order generated, build the segments of Cα trace in order of connectivity. If a molecule is displayed and active, the coordinates are added to the end of the current molecule's coordinates in the data structure. If no coordinates are displayed and active, the coordinates are placed into a new molecular structure at the end of the QUANTA data structure.

Building sidechains by modeling.

It is possible to fit the side chain coordinates using modeling techniques only where the experimental information is very poor or non-existent. The tool **CA-build | Fit seg by D.E.E.** places main chain atoms using the Cα-Ramachandran correlation matrix and then places the sidechain atoms using rotamer searching. Each sidechain is adjusted to match the sidechain angles from a rotamer library and the energy of interactions determined for the build molecule. The process carries out several cycles of single residue analysis followed my multiple residue simultaneous analysis. The energy is computed for each possible conformation and the lowest conformation is retained at each residue. It is recommended that the rotamer library **X-AUTOFIT X-BUILD | Options…Rotamer: Old.** (Oldfield, unpublished results) should be used since these represent a multidimensional analysis of possible observed conformations of sidechain atoms and contain more variants than other libraries.

Automated rebuilding in X-BUILD

X-BUILD contains two new automated rebuilding tools.

1. **X-AUTOFIT X-BUILD | Structure…| Auto build**

   This tool runs through the entire fitted residues and fits them using a mixed grid and gradient protocol. It take three residues, fits the side chain of the center residue by Grid refinement. **Refine Zone** is then used on the main chain atoms only of all three residues, followed by side chain fitting of the two edge residues by Grid refinement. This is repeated for five cycles. The final stage is a round of regularization.

2. **X-AUTOFIT X-BUILD | Structure…| Refine Volume**

   This tool carries out a real space torsion angle refinement for all residues (including ligands, but not water) within a volume about a selected atom. The tool requests a single atom pick, or uses the last picked atom when using “active residue mode.” The radius of the refined volume is set using **Regularise param.** on the same palette, and the default is 6 Å.

   All sections of the molecular structure that are covalently bound to non-refine regions are automatically restrained with fixed atoms.

3. **X-AUTOFIT X-BUILD | Structure…| Do all…**

   This tool allows a refinement protocol to be used on a range of residues. For example, the side chains of the protein could be real space refined using GRID refinement which searches all possible combinations of torsion angles to a precision of 1°. If any changes occur that are outside

user-defined limits, **Do all** stops and centers on this residue so that this residue can be edited.

One of the following protocols can be run with **Do all**: refinement, fitting side chains, fitting main chain, finding alternate conformations, and regularization. The protocol can be applied to waters, the entire protein, or a zone selection.

References

1. Greer, J., *J. Mol. Biol.*, **82,** 279-301 (1974).
2. Oldfield, T. J.; Hubbard, R. H., *Protein: Structure, Function and Genetics*, **18,** 324-337 (1994).
3. Ramachandran, G. N.; Sasisekharan, V., Conformation of polypeptides and proteins. *Adv. Prot. Chem.*, **23,** 283–437 (1968).
4. Dickerson, R. E.; Guis, I., *The Structure and Action of Proteins*, Benjamin/Cummings, ISBN 0-8053-2391-0.

6. Using X-BUILD

The model building functionality in X-AUTOFIT contains many features that accelerate the modeling of crystallographic coordinates. Model building allows real-space refinement in torsion angle space, rigid-body refinement, regularization, and general manual editing of macromolecular coordinates. Auto Build allows you to walk through the structure and rebuild the main chain and side chain of each residue in an automated manner. The Add/Delete palette allows the addition of amino acids, waters, and ions, deletion of residues and ranges of residues, and addition of alternative conformations to amino acids.

To avoid confusion about the target of issued commands, only one structure can be edited at a time. To define the structure to be edited in X-AUTOFIT, use the “active” property on the molecule management palette. The first active and displayed molecule in the molecule management table is taken as the molecule to edit.

Protein or nucleic acids

X-BUILD supports the editing of polypeptide and polynucleotide structures. Since the basic structure of these macromolecules is different, the tools described in this section behave differently depending on the residue selected. The program determines from the residue selected which functionality to use and which information to show in the plot window.

Proteins

Protein atom naming is very strictly observed, and the conventions used are those for all the major crystallographic programs. The 20 basic amino acids and some modified amino acids (that is, HISH) are recognized by the program.

DNA/RNA

X-BUILD automatically supports both ribose and deoxyribose units when editing nucleic acids, and any addition of residues is carried out using a sugar of the same type as the proceeding residue in the chain. Currently, only the five standard bases, adenine, cytosine, guanine, thymine, and

uracil, are supported for full editing, and no restriction is made on the uracil/thymine pair for RNA and DNA. The 5’-phosphate group is optional and is not required for editing the terminal residue in a chain, but dummy atoms are added during regularization and refinement and then removed on completion of these functions. Also supported are the 5’-terminal patches of no-phosphate, phosphate + no terminating oxygen, and phosphate + terminating oxygen.

Nomenclature

The atom- and residue-naming convention for DNA and RNA tends to be less strict than for proteins. The PDB standardization has been used for X-BUILD: 3-letter residue names, and the use of the quote and not * for sugar atom naming. To make the program more general, entry into X-BUILD automatically changes the residue names and atoms names if these conventions are not used in the original file. The phosphate atoms are called “P”, “O1P”, “O2P”, and the terminating 5’-oxygen “O5T”. The terminating hydrogen atoms are “H3T” and “H5PT” where applicable. Some of the editing facilities in X-BUILD are not available if these naming conventions are not used. (The same conventions are used within CNX, CHARMm, and CCP4).

Hydrogen modes

Both protein and nucleic acids can be edited in nonhydrogen, polar-hydrogen, and all-hydrogen modes. As with proteins, the program automatically detects the hydrogen mode on entry to X-BUILD, although you may override the choice at any time.

Disorder, nonbonds, energies, and other general functionality with X-BUILD is supported for both polymer types.

Controlling the display

Extent of display

The number of atoms displayed around the specified center is defined by the tool **X-AUTOFIT | Options... | Coordinate radius.** The volume of map displayed is defined by **X-AUTOFIT | Options... | Map-radius**, while the number of symmetry atoms is defined by the tool **X-AUTOFIT**

| Options... | Symmetry-radius. The size of the calculated bones region is defined by the map radius.

The current residue value

X-BUILD has the notion of a *current residue*. The initial value of the current residue is “1”, but this is changed by several commands. Its value is restricted to between 1 and the total number of residues in the QUANTA data structure. It can be set but not used if it lies outside this range. The actual value therefore refers to the sequential position of the residue in the entire QUANTA data structure. The value of the current residue is saved between sessions. The value is set by the placement of view commands and used to color the current residue point within the Ramachandran and DNA circle plots red, which provides feedback information between these plots and the molecule display.

Modal and amodal/active residue mode

You can switch the program to amodal/active residue mode, in which the current residue is highlighted and any tool acts immediately on this residue. (A residue range cannot be handled in this way.) This mode of action can be used to make single residue editing quicker, since less mouse-clicking is required.

- **Modal**: Each tool requests that you pick a residue to edit.
- **Amodal**: Each tool assumes a current residue and acts on it.

When building all-atom models, X-BUILD can act in both modal and amodal fashion. This is controlled by the **Active residue on** tool in the Pointer palette.

When in amodal mode, the current residue is labeled with pink rhomboids. The current residue can be changed by picking another visible atom. Tools are provided to regularize a range of residues from a single residue, and a tool to regularize a volume of residues from a single residue — these permit the use of the modal action on multiple residues.

Placement of the view

The center of the display can be updated using several methods. The application moves the screen center, updates the map contouring around the region required, recalculates the symmetry atoms, and, if bones are active, recalculates them for the current map region.

By coordinates

Placement by atom coordinate: The tool **X-AUTOFIT | Pointer... | Place using coordinates** prompts you to select a single atom position from the main screen display. Selecting an atom (including a symmetry atom) results in the display center moving to this coordinate. Selecting a non-atom position on the screen or selecting any palette tool aborts this operation. This tool sets the current residue pointer to the residue number of the picked atom, and the Ramachandran plot or DNA plot is updated.

By bones

Placement by bones: If the bones are active, the tool **X-AUTOFIT | Pointer | Place-by-bones** prompts you to select a bones point from the main display. Selecting a non-bones position on the screen or selecting any palette tool aborts this operation.

By atom

Placement by atom name: The tool **X-AUTOFIT | Pointer | Place-by-atom** opens a dialog box that allows specification of an atom name, a sequence ID, and a segment name. If this atom is present in the currently active and displayed molecule, it becomes the new display origin. If it is not found, no change occurs. This tool sets the current residue to that of the residue number of the named atom, and the Ramachandran plot or DNA plot is updated.

Along sequence

Place at next residue: The tool **X-AUTOFIT | Pointer | Place-at-next-residue** moves the display center to the next residue. If the next residue is an amino acid, the display is centered on the Cα atom. If the next residue is a water or an ion, then that residue becomes the display center. If the next residue is neither an amino acid nor a single-atom residue, the display center is set to the first atom in the residue. This increments the current residue number by one. (You can change the next-residue step to values greater than 1 in the **Options...** dialog box

Place at previous residue: The tool **X-AUTOFIT | Pointer | Place-at-previous-residue** moves the display center to the previous residue. If the previous residue is an amino acid, the display is centered on the Cα atom. If the previous residue is a water or an ion, that residue becomes the display center. If the previous residue is neither an amino acid nor a single-atom residue, the display center is set to the first atom in the residue. This decrements the current residue number by one. (You can change the previous-residue step to values greater than 1 on the **Options...** dialog box

Graph plots

Placement by Ramachandran plot: The graph window is always active in X-AUTOFIT and shows a Ramachandran plot of the current active and displayed molecule while the Cα placement dials are inactive. Green points on this plot indicate non-glycine residues, and blue points indicate glycine residues. The red point indicates the current residue. To place the screen origin, click any point in this Ramachandran plot; the residue that has the phi/psi values for this point becomes the screen origin, and the information for this residue is written to the textport. The point picked turns red to show that it is the current residue. There is an **X-AUTOFIT Options | Center at Ramachandran point** toggle that allows a Ramachandran point to be selected without the display center being updated, to allow the Ramachandran plot to be interrogated without waiting for the screen center to be updated. This **X-AUTOFIT | Options** option must be turned on for the display centering function to be effective.

Placement by DNA circle plot. This plot is available when DNA or RNA coordinates are edited. Centering using this plot is identical to that for Ramachandran plots.

Placement by Validation graph.

The validation graph is generated from the Tables and Graphs functionality in X-BUILD. If any point in the graph is picked, then the residue/atom that was used to generate this point becomes the center of the display. If the graph plot was generated from general data (i.e., does not relate to a atom or residue), then no update is made. All maps and bones and additional properties are updated by this pick.

Placement by Validation Table.

If the atom or residue table is open from the Validation functionality, then this can be used to update the display. If a table row is picked, the atom/residue within the current molecule that relates to the row of data in the table becomes the center of the display. The map and bones are *not* updated by this pick. To update this information: after picking the validation table, use the **Goto pointer** tool from the Pointer palette.

Placement by Last Command table.

The Last Command table can be used to center the display at the residue defined in the residue field for the command card. If a residue range is shown in the residue card of this table entry, the display is centered at the first residue of the range.

By text markers

Placement by text marker: The screen origin can be updated using 3D text marks defined with the Text palette. To go to the next text in the list of text marks, use the tool **X-AUTOFIT | Text | Next-text**; go back to the previous text with **X-AUTOFIT | Text | Previous-text**. The current loaded text marks (that is, user defined or loaded property) can be used with the next and previous tools; the most obvious text mark for this process is "loaded property=water molecules". To go to a particular text mark, use **X-AUTOFIT | Text | Defined-text**. A scrolling list of the current text marks appears, from which you can select a single text. This tool does *not* change the current residue number.

Graph plots

Ramachandran plots

A Ramachandran plot appears after any edit of an amino acid and remains until any subsequent edit of a nucleic acid or Cα trace atom. The Ramachandran plot drawn in X-AUTOFIT is a hard-sphere contact and 10% overlap surface of the angles ϕ (C_{i-1}–N_i–$C\alpha_i$–C_i) and ψ (N_i–$C\alpha_i$–C_i–N_{i+1}) (Ramachandran 1968). The contour lines shown are taken from Dickerson and Guis. The definition of the Ramachandran angles are described in Ramachandran and Sasisetharam (1968).

When you are not editing any part of the structure under **X-AUTOFIT | Build atoms**, the Ramachandran plot shows the current values of ϕ and ψ for all the residues in the active and displayed protein as colored points on the Ramachandran plot. All non-glycine residues are green and all glycine residues are blue. The current residue is drawn with a red circle.

The number of points drawn on the Ramachandran plot can be changed using the X-AUTOFIT:X-BUILD | Options... dialog setting:

[X] Center at Ramachandran point...
Segid : [*] start : [1] End : [9999]

This option allows the specification of a range of residues to be drawn on the Ramachandran plot, enabling you to limit the amount of information presented in this plot window when editing small regions of the molecule. The specification is defined for the first active and visible molecule. The wildcard specification (*****) indicates that any segment name is allowed, and the start/end values are those of the sequence ID of the residues. This tool is also valid for DNA and RNA models.

To identify a point on the plot, click the point. The residue that corresponds to that point picked is shown in the textport. If the option **X-AUTOFIT |**

Options | Goto-Rama residue is active, the display is also centered on this residue.

When editing some property under **X-AUTOFIT | Build atoms**, such as the backbone peptide plane (**X-AUTOFIT | Build atoms | Change backbone**), the current Ramachandran angles of only the affected residues are shown. In this case, when the angle $C\alpha_i$–N_i–C_{i+1}–$C\alpha_{i+1}$ is rotated, the ϕ/ψ angles for residues *i* and *i*+1 are shown. For a detailed description of which point corresponds to which residue while editing the coordinates, refer to X-AUTOFIT:X-BUILD Tools for descriptions of the relevant tools.

DNA/RNA circle plot

The DNA/RNA plot window appears automatically after any edit of a DNA/RNA residue and remains until any subsequent edit of an amino acid or Cα trace atom. The plot window style is retained between editing sessions.

The DNA/RNA plot consists of seven fields, α–β–γ–δ–ε–ζ–ξ, in decreasing concentric circles. The coordinate frame is polar, where the radial values have discrete values for each torsion in DNA residues. The polar angle is the value in degrees for each of the seven allowed torsions. Seven green circles are drawn for each residue in the polynucleotide, one in each of the seven fields of the plot. (Terminal residues may have incomplete fields where atoms are not present in the residue.) The current residue (see The current residue value on page 79) is indicated by red circles. If the plot is picked using the cross-hair cursor, then the nearest point on the plot becomes the current residue (as indicated by a red circle), the graphical display changes to place this residue at the origin (when **Center at Ramachandran point** is on), and information on this residue is shown in the textport.

When you are editing some property that will affect a torsion value, the fields are filled only with those values relevant for the edit.

Picking screen information

Various kinds of information displayed in the molecular view and plot windows can be picked, although there are restrictions.

1. The atom information can be picked at any time to return an atom label for the atom picked.

   If X-BUILD is in active-residue mode (amodal), then picking any atom changes the current residue. This changes all graphical objects associated with the current residue.

2. The Cα trace can be picked only when bones are *not* turned on. This restriction results from the use of bone-picking to define the Cα position. The Cα trace atom is labelled with the current segment number and Cα atom position in the segment. The segment number and Cα position are transient and depend on the current segment and atom (**CA build | current seg-res**) when the label is generated.
3. The bones can be picked, when the Cα dials are active, to define the position of a terminal Cα trace atom. Only the terminal Cα trace atom can be placed by this action. The Cα trace atom is moved so that the current position points at the bones point picked, and the pseudobond to this is retained at 3.8 Å.
4. The Ramachandran and DNA plots can be picked at any time to reset the molecular view origin, map display, and bones display to the residue picked. Information on the Ramachandran/DNA mainchain torsion angles are written to the textport with the residue information.
5. The Cα plot can be picked when active, and the current Cα atom terminates the current segment. The current Cα atom is placed in a conformation defined by the point picked.
6. Pick the pointer at any time to return the real-space coordinate of the pointer.
7. Picking an entry in the Last Command table will position the molecular view, provide details on an edit, undo/redo an edit, or provide a comment card, depending on the table cell picked.
8. Picking the validation tables for atom and residue information updates the molecular view so that the center is defined by the atom/residue row in the table. Map and bones information is *not* redrawn.
9. Picking the general graph updates the molecular view so that the center is defined by the atom/residue information that generated the plot. Map and bones information is also redraw.

Symmetry

Symmetry cards for each molecule used are written to files with the same name as the MSF file, but with a .sym suffix. This allows symmetry to be accessed quickly. The symmetry of the Cα trace and bones is stored in the Xfit.sym file, since there is no saved MSF information for this data. If the symmetry is the same for all molecules, the program uses the single set of symmetry information stored within the program. This allows the symmetry to be drawn very quickly for simple cases but also allows more complex cases of multiple symmetry to be handled.

X-BUILD now allows symmetry atoms to be picked and the information of the picked atoms is displayed.

Non-crystallographic symmetry (NCS) is handled by X-AUTOFIT, X-BUILD, X-LIGAND, and X-SOLVATE. Up to 60 NCS matrices can be entered for each open molecule and the Cα trace. The NCS symmetry is displayed in a different color from the normal symmetry information and can be picked.

Interface to run external programs

A facility is provided in X-AUTOFIT-X-BUILD to run an external program. Up to 20 external programs can be associated and, when set up, appear on the Run External Program palette.

The external program facility is set up by writing scripts that define:

- The layout and contents of a dialog box that provides parameters for the external program.
- The handling of I/O between QUANTA and the external program.
- The content of a command file to run the external program.

The scripting language supports hidden dialog boxes, variables, and various language constructs that provide a rich variety of possibilities to set up support programs.

You should provide a script as a file named script.# where # is a number between 1 and 20. When present, this file is checked and added to the Run External Program palette when this palette is opened. When this new tool is selected, the script is read and the dialog opened as defined by the script. When the dialog is completed, the command file defined by this script is written and run.

A debugger is provided that indicates problems with the script.

Advanced validation techniques

The tables and graphs palette is used for advanced analysis of proteins (in particular) and other macromolecules. The advanced validation techniques are carried out by generating tables of data based on atomistic and residue properties and applying functions to them. The data can then be plotted and the graphs used to identify features of the molecule that are interesting or in error.

This palette's tools are used to generate tables containing information on atoms, residues, and other general data. The data tables can be operated on by several functions and plotted in various styles. The graphs and tables can be picked to center the molecule view, and the graphs can be annotated and plotted to a PostScript file.

Table data

Tables of atom and residue information can be generated:

- From the current visible and active molecule.
- From an external file of data (into a general table of data).
- By calculating property information from the first current active molecule.
- By applying functions to columns of table information.
- By applying difference functions between two molecules or two segments of one molecule.

The four tables consist of:

- Atom data - attached to molecule (*natom* data items).
- Residue data - attached to molecule (*nresidue* data items).
- General data - can be any length.
- Scratch data - (used by the auto validation tool).

The length of the tables is defined by the number of visible active molecules in the molecule table and the atom selection defined by the atom selection tool. If two molecules are both active and visible, then the table contains the basic information for both molecules. Once created, the table length of the atom and residue tables cannot be changed; it is necessary to delete the tables to change the table length. The General table is the length of the longest set of data.

The protein property functions produce data only from the first active and visible molecule in the molecule table, all remaining values not associated with this molecule that are in the table are given a value of NoData. These values are *not* plotted. If you want to plot data of the same property from two molecules, then change the activity of molecules in the molecule table and re-apply the same property for the second molecule. Both data columns can then be plotted.

Difference data is generated with respect to the first current active molecule in the molecule table.

Column functions and calculations apply directly to the data columns selected and are independent of molecule visibility and activity.

Picking data

Data can be picked from the tables using the following procedures:

- A *single column* can be selected by clicking the column header.
- *More than one column* can be selected by <Ctrl>-clicking subsequent columns.
- *Multiple consecutive columns* can be selected from one column to a second column using the following events: Select one column by clicking the header, and select the second column by <Shift>-clicking the header.
- *Deselect a column* by <Ctrl>-clicking any currently selected column
- All the above applies also to *rows* in the table.

Additionally:

- The data values cannot be edited.
- Clicking a cell updates the molecule view so that the viewing origin is set to the atom/residue that generated the data for that cell.

Table contents

Table 1 (atom table)

Table 1 is the atom table and contains rows of data that correspond to the atoms in the current visible and active molecules that are selected based on the selection criteria defined by the tool **X-BUILD | Tables and Graphs | Atom Selection**.

- The current active and visible molecule's information is placed in Table 1with the **Read molecule** tool.
- Table 1 is always overwritten when data is read in from a molecule.
- The table created from **Read molecule** consists of columns for: atom name / residue name / sequence number/ x / y / z /occupancy / B-value.
- Column 1 is a hidden column used to reference the molecular atom data from the table. Column 2 is a hidden column to reference the molecular residue data.

- It is inadvisable to delete the Atom (3), Residue (4), SeqNo (5), or Segment (6) columns of data from the atom table. These columns are used to generate information for the residue properties and difference data. These functions abort if data columns were deleted.
- Data can also be added into the atom table using the **Column function** tools, **Protein property** tools, calculations, and **Difference** tools. If the atom table is not open when atom data is generated by one of these tools, then the atom table is generated from the current active and visible molecules.

Table 2 (residue table)

The residue table contains rows of data for each residue in the current active and visible molecules. The selection of residues is based on the selection criteria defined by the **X-BUILD | Tables and Graphs | Atom Selection** tool.

- Table 2 can only be created by use of a residue function on the atom table 1 or by applying a property or difference on the current molecule that generates residue data
- New columns can be created using functions on columns of the residue table or calculations based on the residue table.
- The table is always created with columns of Residue name / Sequence number / Segment ID.
- Column 1 is a hidden column that is used to reference the table row to the molecule.
- It is inadvisable to delete the Residue, SeqNo, or Segment columns from the table as this prevents further residue functions from working.
- If the residue table is not open when a function is applied, then residue data is generated when the table is opened.

General table

The general table is used for data that has no direct relationship with either the atom or residue data of the current selected and visible molecules.

- The general table is generated by reading an ASCII file. The ASCII file should be a space/tab/comma delimited columnar file. The table can contain data of undefined length, and data can be appended to columns and rows when a new file is read.
- Some functions also place data in the general table, for example, the user bond function and probability function.

- The general table has *no* reference back to the molecule data, and graphs generated from this table cannot be used to update the molecule center of display.

Scratch table

You cannot add data directly to the scratch table, which is used by the program to store information for automated validation.

Fixed size of atom and residue tables

Once data has been created in a table, no other rows of data can be added to the table via the properties, difference, calculated, or function tools. If any of these tools are used, only data that already exists in the original atom table is updated in the residue table; all other data is set to NoData (if in the residue table but not in origin data) or ignored (if in origin data and not in the residue table).

Inconsistent data

Inconsistent data may be generated if:

- The active and displayed molecules are changed. If this happens, then any data in tables are only updated based on the current displayed and active molecule. Thus, most of the data will probably be changed to NoData for molecules that are no longer visible and active. This does not produce incorrect data but does result in NoData data for molecules that have been changed. To generate new data for the new active and visible molecules, delete the table and generate a table.
- Deleting or adding atoms or residues in the displayed molecule. This results in incorrect data being generated in the tables. If residues are added to or deleted from the molecule, you *must* delete the tables and regenerate the data.
- A different molecule is read into the atom table without refilling the residue table with the required residue data. The residue data is now inconsistent with the atom data and the molecule data. The residue table must be deleted before any residue data can be created from the new atom table.
- Changing the atom selection *deletes* all tables, to prevent inconsistencies.

Graphs and plotting

General graph drawing facilities are found on the Tables and Graphs palette. Graphing works in conjunction with the tables facilities described above to produce publication-quality plots and for general analysis of properties.

Three tools on the Tables and Graphs palette affect data plotting. Selected columns of any of the atom/residue/general data tables can be plotted on the same graph in different colors and styles. The number of graphs to be drawn is defined by the number of selected table columns. If no table columns are selected, three blank plotting definitions are provided for you to fill in. The plot data tool opens a dialog that provides options to define the look of a plot. This includes the color, line width, style, legends, and axis labelling.

Table names

The allowed table names requested in this dialog are Atom, Residue, and Data and specify the origin table of the data. If a column is selected, this information along with the column number is filled in by the program.

X data

Only one x column is provided by this functionality, and all data is plotted against this one set of x data. By default, the column header for the x data is zero. A zero column number specifies that the data is incremental data, starting at 1 and incremented by 1 for each data value. Any plotted data from the residue or atom table is automatically linked to the molecule data, regardless of the x ordinate.

Legends/annotation

Legends can be added from the plotting setup palette and can also be added using the annotation facility. The font size used for the legends and the axis labels is defined by the **Axis font** option list, and the **Axis style** changes the border around the graph.

General annotation is added with the Label graph... palette, which allows you to place text, boxes, lines, and circles in various colors and styles. The text style is defined using the **Label** string, **Size** option, **Orientation** value, and **Color** value. The box, circle, and line styles are defined by the **Color** value.

The **Attachment** defines whether the annotation is to be added to the **Graph** data or is just an annotation in the **Figure** window. Most annota-

tions, such as legends, are added as **Figure** annotations. If the annotation relates directly to a data point(s) (such as a data point desciption) then it should be added as a graph annotation. Graph annotations *cannot* be placed outside the bounding box of the graph.

Add/deleting annotations

To delete an annotation, select the **Delete** button. The application waits for you to select an annotation on the graph. Selecting the **Add** button requests a single point pick for a text string (the bottom left end of the text is placed at the point picked), while line, circle, and box annotations are placed by picking two points in the graph window.

Hardcopy plotting

The PostScript settings... dialog affects only a plot produced in a publication-quality plot style and only direct plotting of the graph window contents as a PostScript file is possible from this dialog.

The scaled font toggle is provided so that the whole plot plus any annotation can be scaled to the size of the page. When this is not checked, the font size is set as defined when the plot was generated and labelled.

The **color** options allow you to generate plots in black and white, color with the black and white objects (including background) reversed, and inverted color. These options are provided because the graph plot on the screen has a black background, while a hardcopy plot has a white background. **Color invert** often produces better looking plots than **Color B<->W** because of the inverting of the background.

The middle nine buttons are used to place the plot on the page. The large rectangle defines the page and the smaller rectangle within this is the graph border and does not include the axis labels. These include setting the aspect ratio, translation, and scaling. The **Step** defines the increment of movement of the plot and is in inches.

The **set** button sets the current values of the options for future use, the **plot** button plots the current validation graph in the current position and with the current settings, and **cancel** aborts all actions.

Last commands

The **Last Command** tool opens the Last Command table, where the record of all edits done in X-BUILD is stored. The table can be used to check the progress of the model-building process, undo and redo each command, analyze the work done, and create log files of X-BUILD functionality

To hide the table, click the **Last Command** tool again. The presence of the table does not affect the saving of commands to the file. As each command is issued (and if you accept the changes made by that tool) it is added to the top of the table.

The application records the tools used in X-BUILD, and on normal exit (finish), the current session information file is moved to a new file named command#.stack (# = 1–999). When there is a premature exit, the current session information file remains as command000.stack. Thus, the program is made aware of an incorrect exit and previous tools can be recovered. If a new session is started after the previous session was exited correctly, the Last Command table is empty (except for a save changes always made on entry). If the previous session ended incorrectly, then the Last Command table contains the previous session entries.

The table contents

For each tool used in X-BUILD, a new entry is added to the Last Command table.

- Tool name used, some with marks to indicate the type of command:

  (+) Information only.

  (*) Nonreversible command, normally insertion/deletion of residue data.

- Residues edited.
- Time and date.
- Undo/redo columns to indicate:

  a. N/A - The command cannot be repeated/undone (various reasons — usually due to residue deletion/insertion).

  b. No - The command has not been undone/redone.

  c. Yes - The command has been redone/undone.

- Index is the sorting order.
- A comment card is provided for user comments for a particular edit and can be used as a notebook. (these can be labeled on the molecule and written in the logbook). The current comment is placed here with hot code.

Picking title entries in the table

The headers at the top of the tool entries can be picked to sort the data and analyze the content of the table.

- If the **Command title** is picked, the entries are sorted by command name.
- If the **Residue title** is picked, the entries are sorted by residue.
- If the **Time title** is picked, the entries are sorted by time.
- If the **Undo | Redo title** is picked, the current Last Command list is analyzed to determine the number of tools used, the percent of time each tool was used, the rank in order of use, and any suggestions.
- If the **index title** is picked, then the data is sorted by order i.e., placed in order that the command was issued.
- If the **comment title** is picked, the data is sorted by comment.

Picking cell entries

The cells in the table can be picked to place the molecule view, determine the changes made by the tool, undo changes, and redo changes.

- If the command cell is picked in the table, then the changes made to that residue are written to the textport as the min change/max change, rmsd change. Results may not be meaningful if the number of atoms has changed in the residue edited.
- If the residue range field is picked, the molecule view is placed at this residue (If the edit is for a range of residues then the view is placed at the first residue). Only the view is placed here — the map and bones are not recalculated. (But the pointer position is updated, so you are reminded to use the Goto pointer to update the display).
- If the time is selected, nothing is done.
- Undo undoes the edit to before that edit for that residue
- Redo redoes the edit to just after that tool edit.
- Index picking does nothing
- Picking **Comment** opens a dialog to allow you to add a log entry for the edit made. A hot code is also provided to define the color of annotations. The annotations can be displayed on the molecule at the residue centroid using a pulldown menu on the last command table (view). The annotations are also added to the log file.

X-BUILD features

RSR vs. gradient body refinement

For amino acids, X-AUTOFIT provides two forms of real-space refinement. The first (**X-AUTOFIT | Build atoms | Refine 1 residue**) is a true refinement procedure that carries out a gradient minimization of a residue to the current electron density. This can take a few seconds to complete and can be aborted by clicking the screen. This algorithm can refine the position and orientation of any residue type, and if torsion angle information is present in the .GSD file or a user editable file (lig.rot), then a full torsion angle minimization can be accomplished. Refinement of multiple residues by real-space refinement torsion angle refinement is carried out with the tool **X-AUTOFIT | Structure | Refine zone**.

Specific to amino acids and nucleic acids are **X-AUTOFIT | Build atoms | RSR sidechain** and **X-AUTOFIT | Build atoms | RSR main chain**, which use an optimized grid-search algorithm to refine to completion amino acid coordinates in 0.1 sec. The specific amino acid refinement routine also searches all possible conformations available to the amino acid and so does not get into the nearest false minima. These specific routines are recommended for amino acid fitting to density, since they are far more powerful than the traditional gradient-refinement protocol. The specific amino acid routines are also fast enough to allow editing of atom positions in an amino acid while the refinement is active. This means that the effect of placing an atom at a position can be observed as a function of electron density fit.

RSR vs. Move atom and RSR

Where the mainchain is not yet well fitted or a residue has a poor Cα chiral volume, the simple **RSR sidechain** tool sometimes seems to produce spurious results. Often, the geometry of the backbone interferes with fitting the sidechain. If this happens, you may want to pick up the Cα atom with the **Move atom and RSR** tool and move the atom until the sidechain snaps into the density. It is usually obvious when the sidechain has found the right density. Sometimes you only need to move the Cα atom a very small distance.

Defining torsion angles for unknown residues

A tool **X-BUILD | Build atoms | Add-delete | Add 1 torsion** allows the definition or removal of a torsion angle in a residue that is *not* an amino acid or nucleic acid. This can then be rotated manually (**Build atoms | Edit chi angle**s, **Flip torsion 180 degrees**) and also refined with any of the real-space torsion angle refinement algorithms (**Build atoms | Refine 1 residue**, **Structure | refine zone**).

Missing or incorrect atoms

Missing and incorrect atoms found in residues described in the .GSD file (for example, amino acids) can be built or removed automatically using the **X-AUTOFIT | Build atoms | RSR sidechain** and **X-AUTOFIT | Build atoms | Geometric conformation** tools. Both these routines replace a sidechain of an amino acid with a template structure that has a reasonable conformation, regardless of the starting coordinates. This includes the adding or removing of hydrogen atoms when the current X-AUTOFIT non-hydrogen/polar-hydrogen/all-hydrogen building mode is different from the MSF model. When missing or incorrect atoms are found in these template residues, X-AUTOFIT initializes the temperature factors, retypes all the atoms, and resets the charges to template values. You can use **X-AUTOFIT | Build atoms | Edit residue info** to access the full editable table, which allows you to change B values and occupancies for all residues of a macromolecule.

Regularization

The regularization protocol corrects bond, angle, and improper-torsion parameters for all templated residues in the .GSD file. If the residue to be regularized is not found in the .GSD file, the program generates the required parameters from a PSF that is generated using the type definitions of the atoms of the residue. This allows the regularization of any set of atoms in QUANTA or X-AUTOFIT. Before regularization, you must have assigned the proper atom types to the atoms.

The regularization implementation allows atom position editing while the refinement is active. This allows regions of atoms (for example, a loop in proteins) to be manually edited, with the program retaining the correct geometry during the editing process. This powerful facility for loop editing is an improvement over the usual practice of placing the Cα atoms in specific positions while the program moves the rest of the structure to compensate for the loop changes.

Regularization compared to RSR

Regularization changes the bonds, angles, planes, and chirality of residues. It does not explicitly change the rotatable bonds (that is, chi, phi, and psi). The torsions *are* affected if this is the only way the algorithm can satisfy the restraints.

RSR changes the χ angles and adjusts some bond angles to help find density for sidechains. The algorithm rebuilds the residue first from the template, so after using the RSR sidechain tool, all bonds have the correct value regardless of the starting positions of the atoms. The bond angles and improper torsions may have a deviation of up to 12° from the parameter values to enable the algorithm to search density when the mainchain atoms are not in the correct positions (as judged from the density). Hence, although the regularization and RSR algorithms are complementary and change different parameters, you should use real-space refinement on a residue first and regularize it afterwards.

The tools **Refine 1 residue**, **Fit sidechain by RSR**, and **Move atom +RSR** on the Build atoms palette do *not* carry out geometry minimization and thus do not correct the geometry of a residue. This tool is specifically designed to place atoms into electron density with some detriment to angles and improper torsions, the latter being easy to correct with the **Regularize** tool or with general reciprocal-space refinement.

The **Refine Zone** tool on the structure palette uses mixed parameterization and therefore fits atoms to electron density and optimizes all geometric terms.

Regularization and disulfides

X-AUTOFIT has two ways of treating disulfide links during regularization. It can ignore them or treat them by including the link and the connected residue. You control this behavior by setting the parameter **Regularize across disulfides** on the X-AUTOFIT | Options dialog box.

When the **Regularize across disulfides** parameter is on, if the zone you want to regularize contains a cystine as part of a disulfide, X-AUTOFIT checks whether the other cystine of the pair is already in the regularize zone. If it is, then it just sets the restraints for the bond. If the other cystine of the pair is not in the regularize zone, then the cystine is included, the N and C atoms are fixed for this residue, and disulfide bond restraints are added. Thus, if you want to regularize a disulfide bridge, select just *one* of the cystine residues to give a single residue range; the other residue is added automatically.

Hydrogen representations

X-AUTOFIT supports no-hydrogen, polar-hydrogen, and all-hydrogen representations of residues. The mode is automatically determined on entry to X-AUTOFIT. The X-AUTOFIT Options dialog box has an option to set the hydrogen representation mode. If the loaded MSF structure has a different hydrogen representation from that defined in X-AUTOFIT | Options, then only **RSR sidechain** and **geometric conformation** tools can be used. These tools then build or remove hydrogens from the residues. Once a residue has the correct hydrogen representation, it can be edited by all the **X-AUTOFIT | Build** tools. Any residue not in the .GSD template file can be edited regardless of the hydrogen representation, since no atom editing on these residues can take place in X-AUTOFIT.

Disorder

X-AUTOFIT supports disorder for all residues, up to four alternative positions of atoms. For a disordered residue, most commands first split the residue into multiple complete residues, depending on which conformation was selected.

If the atom selected for any tool is in a disordered residue, then:

- If the atom is disordered, the alternative conformation defined by that atom is used in that tool.
- If the atom itself is not disordered and is a mainchain protein or nucleic acid atom, the tool defaults to the A conformation.
- If the atom itself is not disordered and is not a mainchain protein or nucleic acid atom, the application asks whether to edit an A or B conformation. C and D conformations cannot be selected by this process.

Since most tools act on a residue (and not on single atoms), it would be normal to explicitly select the conformation to edit by selecting the required conformation from sidechain-disordered atoms. After editing a disordered residue, the separation of the atoms from the multiple conformations is checked. If the separation is less than 0.01 Å, then this atom group becomes a single atom. Otherwise, the multiple atom positions are retained as separate conformers.

Clamping of alternate conformations

The clamping option is found on the X-AUTOFIT | Options dialog box. When clamping is turned on, X-AUTOFIT moves the B (C and D if relevant) conformer so that the mainchain atoms are aligned with the A confor-

mation mainchain atoms.This means that when a residue A conformation is edited (for example, regularized), then on completion of the edit, the B (C and D) conformers move to the equivalent before-edit position, relative to the mainchain atoms of the A chain. You should use clamping of the B (C and D) conformer to the A conformer for all single-residue disorder, and only turn off this option when a disordered loop is being added.

It is possible to convert a multiple conformation back to a single conformation by fitting the disordered atoms in the same place. When the edit is complete, a routine removes all the excess atoms from the alternative conformation, and it is possible to remove all of them if they are very similar in conformation. The residue is therefore no longer classed as an alternative conformation. The most likely scenario is when multiple disordered conformers are both fitted with **RSR sidechains**: this causes the same solution to be found. B conformations are fitted when added with the **Add alternate conformation** tool, so a B conformation can be refitted by deleting it and adding it again.

C and D conformations cannot be fitted automatically to density, though the **Refine 1 residue** tool may fit these if close enough to a solution. Be aware that if the mainchain atoms move for a clamped residue, then this edit may affect other disorder of this residue. (You should fix the mainchain atoms for a clamped residue before using this tool.)

C-terminal oxygen atoms in proteins

X-AUTOFIT supports various naming conventions for C-terminal COO groups. X-AUTOFIT does not explicitly use or show the second oxygen position during editing, but adds it back to the residue on completion of the edit.

5´ Terminal residues in nucleic acids

The tool to repatch the 5’-terminal in DNA supports three conventions. The list of options is found in Using X-AUTOFIT. X-AUTOFIT tools handle all forms of the patch, but when there is no 5’-phosphate present, these are added as dummy atoms during regularization and refinement and then removed afterwards.

Summary of using X-BUILD

Editing each residue

For most residues, it is often only necessary to RSR the sidechain coordinates, followed by some regularization. This makes model building a very easy process. When you need to more extensively edit sections of a chain trace, use the tool **Move zone and regularize**. During this process, the Cα atoms of the zone being regularized and edited should be moved to the region of density that represents the Cα density. This should drag the rest of the residue atoms in the zone to a new location. Once this Cα atom movement has been completed, use the tools on the **X-AUTOFIT | Build atoms** palette, **RSR sidechain** and **RSR main chain**, to automatically place the other atoms.

Where the sidechain positions are indeterminate, the tool **X-AUTOFIT | Build atoms | Geometrical conformation** can generate a theoretical conformation based on the conformation observed from the protein databank. Each conformation is shown with the nonbond contacts to other displayed atoms.

Waters

Gradient body refinement can be used on waters to place them into the density after refinement. The process of refinement often shifts waters out of the center of the density. This tool simply returns them to a refined position. Used in conjunction with **Goto next residue**, this allows a simple way to traverse all the waters. This procedure is fully automated under **X-AUTOFIT | Structure | Do all**.

Color

The X-AUTOFIT | Build atoms | Color atoms palette allows easy checking of a poor structure visually. After refinement, **Color by B-value** allows visualization of the badly fitted regions, and **Color by fit** shows the current quality of fit to the map. **Color by progress** allows the visualization of the modeling session, so that it is easy to find how much has been fitted. Also it can be used to see when the last save to disk occurred and whether a failure to fit a region would be worth undoing from disk or editing again.

Manual editing

The manual editing facilities provide all the necessary facilities for fitting, but for most model building it may only be necessary to use the auto-fitting facilities. You may need to fit the B conformation of a sidechain manually, since fitting with real-space refinement produces the same solution as the A conformation. This can be done using **X-AUTOFIT | Build atoms | Refine 1 residue** by gradient refinement, since this will move to the nearest density.

Loops

There are several ways to rebuild loops in X-AUTOFIT. If it is necessary to delete residues in a loop, use either **Delete residue** or **Delete range** on the X-AUTOFIT | Build atoms | Add delete palette to remove one or more residues. The new termini created by this process can be manually adjusted using **X-AUTOFIT | Build atoms | Model first last 4 res.** tool to bring them together or forced together using the **X-AUTOFIT | Build atoms | regularize** option. It will be necessary for the regularize option to work to insert a peptide bond between the two cut ends of the loop. Use the tool **X-AUTOFIT | Build atoms | Add delete | Create peptide link** to add this new bond; and since there is *no* length restriction on this new peptide bond, you can temporarily create a very long peptide bond, which will be regularized to the correct length.

To insert residues into a loop, you must first cut a peptide bond to create two new termini, using **X-AUTOFIT | Build atoms | Add delete**.

Delete peptide link. You can now use the **X-AUTOFIT | Build atoms | Model first 4 res.** tool to manually move the two ends to make room for the new residue(s) and then use the tool **X-AUTOFIT | Build atoms | Add delete | Add res at termini** to add the residues from either of the new cut ends. As well as deleting residues, a new peptide link can be created, and the new loop regularized to remove the strain. The new loop can be fitted as two cut ends before joining these together, refined with **X-AUTOFIT | Structure | Refine zone**, or searched with the tool **X-AUTOFIT | Structure | Loop fit**. Large changes can be searched with loop fitting algorithm, but there is a limit of around six residues on the size of loop that can be fitted in a reasonable period of time.

The **Save changes** and **Undo last** options under **X-AUTOFIT | Build** write and read a session file to and from disk. This is much quicker than restoring the data from an MSF file or writing a new file and allows you to save your editing in process and to undo a mistake very easily.

Structure refinement

Rigid-body refinement (on the Structure palette) of a zone allows the fitting of sections of structure and very quickly finds the nearest minima for all atoms in the zone. The radius of convergence is not as great as in rigid-body refinement in reciprocal space, so it may not be able to improve rotation/translation function solutions for a whole protein. It may be possible to refine each domain or even secondary structure element.

The structure-building section includes rigid-body refinement and full torsion-angle refinement for a specified region of the structure. Thus it is possible to rigid-body refine a domain, segments, or even just local folds to improve the initial fit to a map. The refinement protocol uses torsion-angle refinement and rigid-body refinement on a residue basis while retaining geometry with constraints on bonds, angle, impropers, and nonbonds. This results in a refinement method with a very high radius of convergence (1.5 Å). This refinement protocol should not be used to replace normal xyz refinement, since it is only applied in real space and does not affect the map. It does provide a rapid method of refining regions using a very powerful algorithm. Loop fitting and terminal fitting are accomplished using a Monte Carlo sampling method that can screen thousands of conformations per second against the electron density while retaining the ten best current solutions during the search. The progress of the search is displayed, and you can interrupt it when the ten solutions have converged to a single solution or when an obvious fit can be observed.

7. Using X-POWERFIT

This chapter describes the use of X-POWERFIT to generate Cα traces directly from electron density maps. X-POWERFIT is designed to analyze experimental electron density maps and then provide automated methods of generating and adding Cα-trace atoms into electron density. There are three automated protocols available: one optimized for maps with high resolution (better than 2.0 Å), and the other two for low resolution (2 to 4 Å).

X-POWERFIT has been designed as a tool for crystallographers to help speed up de novo Cα tracing and model building. As in all model building, the crystallographer should carefully review the automated results for accuracy.

Introduction

X-POWERFIT can be accessed from the X-AUTOFIT: X-BUILD palette as the tool X-POWERFIT. This application is designed to analyze an experimental map and trace the Cα chain in an automated or semi-automated manner. It has three different tools for automated de novo Cα tracing:

- **X-POWERFIT | Auto-trace High** for automatic tracing of high resolution maps (better than 2 Å). The algorithm uses a simultaneous multiple path analysis to give a single Cα trace of the structure.
- **X-POWERFIT | Auto-trace Low** designed for lower resolution maps (between 4 and 2.0 Å). This tool can be used after the manual placement of one or more Cα atoms or the automatic placement of secondary structure elements with **X-POWERFIT | Find sec. struct**. The tool attempts to automatically trace the entire Cα chain based on the placed Cα atoms or the secondary structure elements.
- **X-POWERFIT | Consensus tracing** carries out nine rounds of the **Auto-trace Low** protocol with different sigma values of the skeletonized representation of electron density.

Also available are algorithms to check for internal knots in the Cα trace, refine the Cα traces, and edit the positions of the vectors (helices and strands) identified by **X-POWERFIT | Find sec. struct**.

The application requires an extended map that covers the entire region to be traced. The knowledge of the space group of the molecule to be built is highly recommended but not absolutely necessary.

Data

X-POWERFIT is optimized for the use of real data that is better than 4 Å in resolution, and works best on the original MIR/SIR/MAD map. Although it is possible to use a final map to try out the application, the best results are with real, unrefined data. You can also try the X-POWERFIT tools on MIR/SIR/MAD maps that have been through density-modification. Be aware that if no information is present in the electron density map, then no information will be found by this application. The main advantage of X-POWERFIT is with large structures where the large amount of information in the map can make initial tracing difficult for a person to complete. X-POWERFIT has been used successfully to identify most of the Cα atoms in structures with 200-500 residue in 1-5 minutes.

Who can use this application?

X-POWERFIT is simple to use, as it is mostly automated, but it must be noted that it is still an expert system. It is easy to build an incorrect structure very quickly when used by a non-crystallographer. At each stage, the application should be used as a guide to modeling a crystal structure and not as a black box building application.

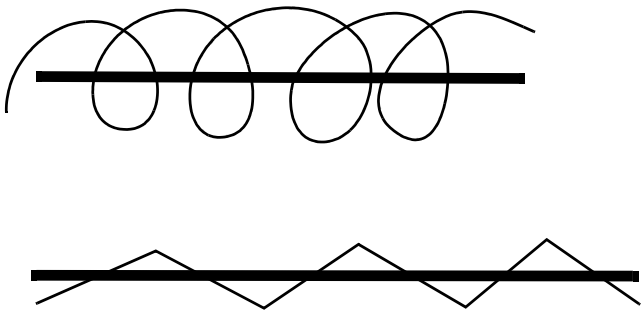
Map quality

As an extension to the model building aim of the application, the tool for determining the secondary structure (**X-POWERFIT | Find sec. struct.**) is useful in determining the quality of lower resolution electron density maps. This represents an interpretative method of classifying the over- and under- connectivity of electron density maps. Hence, the quality of the map can be defined as the quantity of secondary structure that can be determined successfully by the application. Of course this may not indicate the quality of the loop regions, which are always the most difficult sections of the model to build.

Reduced representation (low resolution tracing only)

For the low resolution tracing protocols (**Auto-trace Low** and **Consensus Tracing**), the starting point is not the electron density bones but either one or more manually placed Cα atoms or the secondary structure elements (helices and strands) of the molecule. The latter can be identified using the tool **X-POWERFIT | Find sec. struct.** This tool identifies secondary structure in the map.

Find sec. struct. is based on the calculation and placement of vectors that represent the principle components of the secondary structural elements.



Vector = principle component of helix/strand

Length = magnitude of vector

This palette can be used to generate an initial Cα trace, which can be then extended if needed in a semi-automated manner using the CA Build palette for a Cα -tracing session.

Before starting

Maps

Before you use X-POWERFIT, you must calculate and extend an electron density map so that at least one protein molecule is covered by electron density. This map should be converted into brick map format and then opened using the map handling tools. Because the automated routines

determine structure in all the currently visible maps, it is recommended that the first stage in analysis be to delimit the molecule with a map mask.

Masks

To generate a map mask from an experimental map, the map must be open and the Map Management table displayed (**Draw | Maps table | Show map table**). The following description presents the generation of a map mask from bones, but if you have a homologous set of coordinates, these can also be used to generate the mask. The X-AUTOFIT application should be started (**Applications | X-AUTOFIT**), and the palettes for **Bones...** and **Map masks...** should be opened from the X-AUTOFIT control palette. Using the **Options... | Map radius** dialog option, set the map radius to a large value, such as **100** Å. This results in the entire map being displayed and read for analysis, but be aware that systems with less than 64 MB of real memory will be slow due to memory swapping.

Bones

All of the tracing protocols in X-POWERFIT work with the skeletonized representation (bones) of the electron density map. The bones can be edited, and turned on and off, using **X-AUTOFIT | Bones…**. For more details on how to generate and edit the bones, refer to Bones (page 52).

The bones parameterization

Since the algorithm is based on the analysis of the bones network it is to be expected that there is some correlation of the bones parameterization to the number of elements found in the electron density. Because the algorithm sets the **Trim level** and **Side /main chain detect level** to optimal values, the initial values of these parameters have no effect on the results. The algorithm also modifies the bones slightly, so as to create a particular density of network start points, so the presence of side chain density is not necessary. This is important for lower resolution densities where the bones can end up particularly featureless. The only parameter that can therefore affect the algorithm is the **Start value** of the bones. Generally, if the bones look interpretable then the algorithm will work. There is actually little detriment to the quality of the results for bones levels that are slightly small (that is, over-connected) except that the search takes longer as there are more possible connections to explore. There is major detriment to the results where the bones are significantly fragmented.

Start parameter

The type of map used affects the bones start value. The following numbers for the bones start value are provided as a guide only, and you should optimize the connectivity of the bones using **Bones | Map quality from bones**. The aim is to reduce the number of false connections without increasing the fragmentation of the map within the bounding mask. You should note that the values suggested are slightly lower that would normally be used for manual model building. This is to reduce the fragmentation of the bones which seriously affects the quality of the results produced.

- For 3fo-2fc maps: start = 1.6 to 1.8 σ.
- For 2fo-fc maps: start = 1.1 to 1.2 σ.
- For Sigma A weighted maps: start = 0.8 to 1.2 σ.

Applying a map mask

Once the mask has been generated (Introduction to X-AUTOFIT/ X-BUILD/X-POWERFIT (X-FIT) (page 37)), it should be used as a boundary for the calculation by turning on the toggle option **X-AUTOFIT | Bones| Mask bones by mask**. This increases the speed of the calculation and prevents finding structure outside the volume of interest. The X-POWERFIT palette can now be used to auto trace the Cα atoms.

Tracing high resolution structures

If the resolution of your data is between 2.0 Å and 1.2Å, you can use the new high resolution tracing protocol in X-POWERFIT. This protocol is dependent on the quality of the map. Since the tracing algorithm is extremely fast (normally within 1 second), we recommend that you try **Auto-trace High** with several different **Bones start values** (sigma levels).

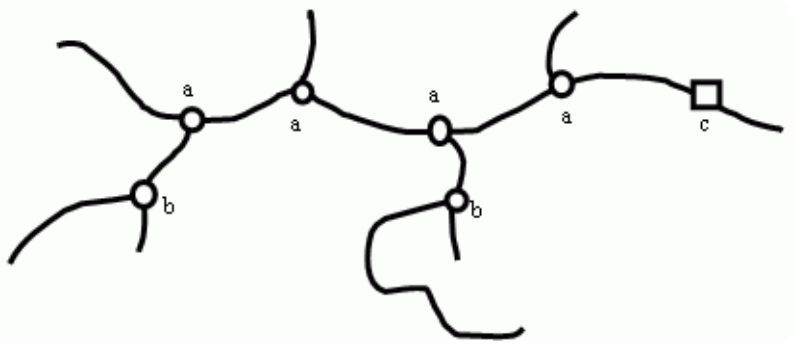
Open the X-POWERFIT palette from the main control palette X-AUTOFIT: X-BUILD and select **Auto-trace High**.

This is a pathway analysis method and is designed to work on connected electron density-this generally occurs in maps with resolution worse than 1.2 Å. It also requires some peaks in reasonable Cα atomic positions. This generally occurs in maps with resolution better than 2.2Å. This method differs from the low resolution density tracing method in that it correlates all the possible paths at one time rather than sequentially follows a path.

The tracing method looks for all the possible 3.8 Å batons within the map, then does a correlated analysis to work out the ideal path from all the possible batons found from the bones.

Pathway analysis

The bones path is generated from the masked out region of electron density. Branched points are identified (In the figure below, points labeled a are probably Cα atoms; points labeled b are possibly Cα atoms).



The presence of Glycines or poor density can result in featureless regions in the bones representation. For such regions, the depth analysis of the bones looks for sections longer than 5 Å, and divides these up into the best integer number of 3.8 lengths (point c in the figure above).

At this stage, multiple Cα–Cα batons are created using a depth search from each branched bones point to find neighboring branches. The variance allowed in the baton length is 3.0 Å - 4.8 Å.

The resulting batons are clustered to remove degeneracy (multiple interconnected branch-points). Adjacent batons are merged to generate fragment chains, with a length restraint of 3.0 Å - 5.0 Å, and an angle restraint (opening angle) of > 75°. Fragments with a single gap of between 3.0 Å and 4.8 Å are closed to form a single continuous fragment.

Fragment analysis

The fragment analysis is based on the assumption that the most likely trace in the electron density is the longest trace. The algorithm starts with the longest trace and deletes all fragments that overlap with the trace. The result is a unique set of non-overlapping fragments.

Generation of Cα trace

The final step is conversion of the best set of fragment chains into a Cα trace. Cα refinement (page 118) should be run on the resultant Cα trace because of the errors that can result within the pseudo bond lengths.

Tracing low resolution structures

Before using the automated low resolution tracing tools (**Auto-trace Low** or **Consensus tracing**), one of the following steps must be performed:

- Determine the secondary structure of in the molecule
- Manually place one or more Cα atoms in the molecule

Determining the secondary structure in the molecule

One way of starting a low resolution tracing experiment is by determining the secondary structure from the map. The calculation uses the bones information so the bones should be turned on. Specify a map radius large enough to completely fill the bounding mask, with bones trimmed beyond the mask. Turn off the mask display, as it is not necessary to view the mask, and it reduces the performance of the graphics.

Introduction

Open the X-POWERFIT palette from the main control palette X-AUTOFIT: X-BUILD and select **X-POWERFIT | Find sec struct.** Depending on the volume of the asymmetric unit this will take some minutes to calculate, normally about two minutes for a protein < 200 residue, and maybe up to 30 minutes for an 800-residue protein. The progress of the calculation can be observed on the message line and the molecular display. Upon completion of the calculation the display should have a number of vectors in color 5 and color 14 (normally white and pink, respectively). The color 5 vectors represent possible helices and the color 14 vectors represent possible strands.

The process of determining the secondary structure is carried out in several stages. On picking the **Find sec struct tool**, the bones display is hidden to improve the speed of redrawing of the screen. Maps and masks are not

automatically hidden by the tool and you are recommended to make them invisible before using this tool. To abort the search during the pattern recognition, click the left mouse button. All the following steps are carried out with no intervention by you, unless aborted by a mouse click.

Pattern recognition

The first part is a 13-stage pattern recognition step where every part of the bones network is analyzed for parts that look like a helix or strand. The program attempts to determine the longest piece of secondary structure that will fit before the pattern of the secondary structure breaks down. The smallest piece of secondary structure that can be found by the algorithm is 2.5 turns of helix, and a four residue strand. The longest section of secondary structure that can be determined is 50 Å long. This phase of the analysis is observed as a series of color 5 and color 14 lines appearing on the screen. The lines adapt and merge as the algorithm continually modifies the results during the analysis.

The progression of this part of the calculation is indicated on the message line at bottom of the molecular view. The program indicates that it is searching the map and shows the progress as the proportion of the total number of bones networks to be searched

```
Pattern recognition Done/To do  256/433
```

In this example there are 433 networks to search for this map, 256 have been searched. As a guide, this example took 70 seconds to carry out the 433-network search.

Cluster analysis

The next stage of the search is to cluster the resulting solutions.

The multiple solutions found are combined by cluster analysis to reduce the time taken in the next steps. For smaller structures (less than 300 residues), the time taken for the clustering is not very long, normally on the order of seconds, but for large structures of more than 500 residues, the clustering can take minutes. The prompt will change during the clustering process:

```
Clustering  244
```

The number reduces in size, initially very quickly, and more slowly as the problem becomes more difficult to solve. For the example here, the clustering took two seconds. For a much larger protein of 750 residues the clustering took about five minutes.

Sheet structure

The next stage of the analysis is the determination of sheet structure from the strands found by the search. This is instantaneous and weights the probability of the strands towards the generation of supersecondary structure within the map.

Refinement

The next stage of the analysis is to carry out directed refinement of fragments of Cα-trace into the electron density at each of the proposed sites. The detail of the directed refinement is described under the tool heading Vector -> Cα trace. The application attempts to refine a secondary structure element into the electron density, and then determine a weighted fit of the element at the final site of refinement. The weighted fit is then used to screen out elements unlikely to be real structure, which are therefore deleted from further analysis. The refinement takes approximately 0.2 seconds/element. The progress of this part of the calculation is indicated on the message line by the comment:

```
Refining Error : 45/70 = 0.210
```

This prompt indicates that the 45th structural element of 70 to refine has a residual from refinement of 0.210. If the residual is below 2.0, the element is accepted as a possible secondary structural element. The units for the residuals are undefined, as they are a weighted fit of modified atoms to electron density as a function of their position in the secondary structural element.

The application also writes out to the textport when it has found a likely helix/strand when the residual is below 2.0 so that you can at least observe if some success is likely during the calculation. For the example with 70 elements to refine, the time taken was 12 seconds.

Overlap analysis

The next stage is an overlap analysis. The secondary structural elements are weighted by the directed refinement algorithm as the likelihood of fitting at a search position. The tool checks to see if the remaining elements overlap since secondary structure cannot physically overlap within a protein. (Some overlap is allowed at the ends of elements so that bent helices can join together.) This analysis is very rapid and is normally less than 1 second as the number of remaining elements is very small. The message line shows:

```
Deleting overlaps  15
```

The number indicates the number of remaining secondary structure elements during the analysis of overlaps. In this example the time taken was less than one second to carry out the overlap analysis, and the final number of secondary structural elements was 10.

```
Number of vectors = 10
```

X-POWERFIT describes the results as vectors, since the secondary structure at this stage is only presented as single lines representing the principle components of the secondary structure. This reduced representation is to prevent too much information being displayed on the screen as the result of the analysis.

On completion of the calculation the vectors that represent the principle components of secondary structure are shown using lines in color 5 (helices) and color 14 (strand) with a line thickness of 5 units. These default values can be changed using the Color table... dialog box.

The bones display reappears if initially visible before the calculation. By default, the bones are not permanently modified by the calculation (the extra points are removed). However, you may want the bones to be modified by the calculation to improve the search analysis. If so, a subsequent use of **Find sec. struct.** gives slightly different results as further modification is made. This feature is included as an option because it allows subsequent searches to give different results, which can be advantageous with difficult problems.

Placing the initial Cα atoms

The second way of starting a low resolution tracing experiment is by placing one or a few Cα atoms manually using the bones atoms as a guide. For this, use the tools **X-AUTOFIT:X-BUILD | CA-build | Next CA** or **X-AUTOFIT:X-BUILD | CA-build | Add helix or strand**.

Automatic generation of the Cα trace

Once the secondary structure elements have been identified, or a few Cα atoms have been placed manually, you can automatically generate a full Cα trace for the molecule using **X-POWERFIT | Auto-trace Low**. This protocol is uniquely suited for low resolution structures (between 4 and 2.0 Å). The number of Cα atoms placed by this tool and the precision of the atom placement depends on the quality of the map, but even with poor maps

(FOM ≈ 0.5), the algorithm can provide a significant saving in time over conventional methods of interpretation. The protocol is also strongly dependent on the bones sigma and trim values (as given by the **Bones Start** values).

The algorithm uses a sequential trace method; that is, it places one Cα atom at a time based on the position of previously placed Cα atoms.

Start point for tracing

The starting point for automated low resolution tracing is either the placed vectors or the Cα atoms. If vectors are used, these are converted to a Cα trace by rigid body refinement.

- A good map traces from the best secondary structure element.
- A bad map traces from each secondary structure element in turn.

The vectors have an associated weight defined by the fit to density for this vector trace, so the multiple secondary structure vectors are ordered by fit to density. The tracing therefore starts at the best secondary structure vector.

On failure of fitting a chain, the **Auto-trace Low** button tries the next secondary structure element (unless it has already been fitted by a previous tracing attempt).

Pathway analysis

The pathways within the electron-density map are defined by analysis of bones ridgelines starting from the bones seed point (Figure 1., blue circle). The bones are analyzed to determine all points that lie 3.8 Å away on the bones path (Figure 1., red circles). The green points in Figure 1. are derived form the red points and are determined by taking the unweighted mean position of the bones path to the red point extended to 3.8Å from the blue circle. This is done because sometimes the Cα atom is on the bones path,

but sometimes it is best found as fitting the average path of the bones but does no lie exactly on the bones.

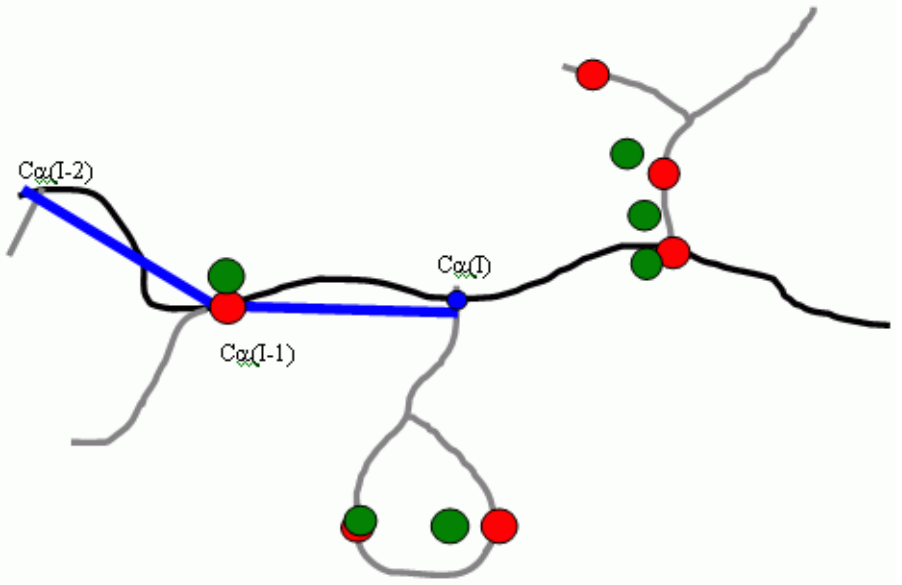


Figure 1.

The Cα trace is extended based on the following rules:

- Trial points are given more weight if they fall on the longest continuous bones main chain.
- The opening angle (angle made by a new Cα atom with respect to the previous two Cα atoms is ideally between 70° and 160°.
- The density minimum and mean are high throughout the trace, and the fit at each Cα atom position is appropriate for a tetrahedral atom.
- If the secondary structure vectors have been calculated, the Cα trace should come from a point within 4 Å of the end point of the element, and is weighted depending on the angle of entry (rule for entering a secondary structure element).
- If the secondary structure vectors have been calculated, the Cα trace should remain within 4 Å of the element and form an angle with the element within the specified limits (rule for continuing along a secondary structure element).
- The Cα trace cannot overlap with already traced sections (this is enforced by a simple weighting function for non-bond overlaps with already traced map, so that close clashes are heavily penalized and there

is a “shift” function between 2.5 and 2.8Å, and no weight at greater distances.

- A mapping between Ramachandran space and Cα geometry is used to define and rank the probability weights for each Cα trace angle and torsion.
- Each peptide plane in independently fitted to electron density (free-Geometry fit: without any restraints on the N-C-C angle)
- Exit conditions: no density, clash, bad free-Geometry fits.
- For a bad exit condition (no density or bad free-Geometry fit), the last Cα atom is removed and the next best position of the last Cα atom is used as the start point (tracing through side chains is avoided by allowing failure back-tracking for two Cα atoms).

Consensus tracing

The quality of the tool **X-POWERFIT | Auto-trace Low** depends on getting the right bones values (**Bones Start** and **Bones Trim**). If the values are too low, there will be too many branch points which can lead to the wrong trace. If the values are too high, there are too many break points and therefore an incomplete or fragmented trace. The method of consensus tracing involves interpreting the electron density map nine times with a different bones start parameter, then attempting to automatically identify the ideal trace. Each of the nine traces is analyzed in turn and the quality indeed (TQ_i) is defined for each trace as the variance along the trace atoms compared with all other traces. The best trace is defined as that with the lowest quality index and represents the trace with the lowest variance from all other traces,

Eq. 1

$$TQ_i = \frac{1}{N_i^{\,2}} \left(\sum_{j=1,9\,(i \neq j)} \sum_{n=1,N_i} y^2_{n,i,j} \right)$$

Where TQ_i is the trace quality for trace *i* and $y_{n,i,j}$ is the separation between atom *n* in the trace *i* and the nearest atom in trace *j* . The index is normalized by $1/N^2$ to decrease the value of TQ_i for long trace lengths, as these are more desirable. The variance of each atom $C\alpha_n$ is defined as the sum of the distances to the nearest atoms within the eight other traces.

Adding secondary structure elements

If the automated tracing protocols described above fails to give a satisfactory trace, you can extend the trace in a semi-automated manner using the tool **X-POWERFIT | Vector -> CA trace**. This tool uses directed refinement to place Cα atoms into the electron density using the vector as a starting point. You should note:

- Helices are added with great precision.
- Strands are fitted with less precision due to the large variation in the number of possible conformational possibilities, so you should be aware that some editing may be necessary for strands added with this tool.
- Atoms that terminate the helix/strand may have significant error if they lie outside the actual extent of the secondary structure element.

The likelihood of the element being identified correctly, and hence fitted to the electron density, can be inferred from the “error level” printed to the textport upon completion of the refinement of the Cα trace fitting.

- Values less than 1.0 can be considered very likely correct.
- Values between and 1 and 2 may be correct and should be checked.
- Values > 2 are not likely.
- Values > 3 are certainly not correctly fitted.

Any structure fitted, but obviously not correct can be deleted with the tool **CA Build | Delete current segment**.

On fitting the secondary structure elements, Cα atoms should be deleted from the ends if they are beyond the extent of the secondary structure. Use the tool **CA build | Delete current CA** and append the Cα atoms to the segment with the tool **X-POWERFIT | Next CA as helix-strand**. Upon completion of fitting all the required observed secondary structure elements, the procedure described in the next section can be carried out to place more secondary structure. Since most proteins contain secondary structure throughout most of the molecule, any volumes in the mask not containing secondary structure should be analyzed again.

Adding more secondary structure

Missing strands of beta sheets can be observed sometimes after one round of automated Cα tracing. These can be added manually with the tool **X-**

POWERFIT | Place edit strand by picking two points from the bones. Since the directed refinement has a high radius of convergence, the original placement need only be approximate, but the length may have to be extended due to a size limitation imposed by the refinement algorithm.

Structure in local regions can be fitted by moving the pointer to the center of the regions to be searched (with either the mask pointer, or rhomboid pointer), setting the map mask radius to a smaller value to cover just this region completely (from the Options... dialog), and reducing the bones start value. Use **Pointer | Go to pointer** to reset the display with the desire position. The search can now be carried out again to see if any more structure can be observed in this region automatically, and if so, added as Cα trace with **X-POWERFIT | Vector -> CA trace**.

Searching the PDB for similar motif patterns

Once the secondary structure has been found as vectors, it is possible to search the protein databank using a maximal sub structure alignment of secondary structure. The tool **X-POWERFIT | Search and Browse DB** allows you to run a structure motif alignment program that can, in about five to ten minutes, carry out the alignment of eight secondary structure elements against a selection of 7,000 proteins in the protein databank.

At this stage of the de-novo building process the structure direction is not known; the alignment is carried with the vector elements in both directions.

The results can be browsed as a Cα trace superimposed on the map/bones, and the Cα trace can be loaded into the tracing application and edited if it is found to be close enough to use.

It is sensible to caclulate a new set of proteins for this alignment program. The alignment program has a mode "-generate" that creates the secondary structure motig library. I also suggest that a SCOP library of non-degeenerate proteins is used this is about 6000 proteins. This will lim

Building general structure

Once the secondary structure has been placed, it is now possible to extend this with the tool **X-POWERFIT | Auto extend CA**. The extension is carried out from the current Cα atom in the current Cα trace. This can be set with the tool **CA build | Current res seg.** You should note:

- The starting Cα atom from which the building is to be continued should be well fitted.

- The building process is very critical of its progress and often stops with errors — these should be checked.
- The results from automated building should be carefully reviewed.
- It is possible to interrupt the auto building process by a click with the left mouse button.

The user should experiment with using the tool from each end of the secondary structure placed, using the pie chart in the bottom left hand corner of the display to view the progress. If some structure has been correctly built, while the end is wrong, it is possible to cut up the Cα trace with the tool **CA build | Unjoin 2 CA**, and rejoin them with **CA build | Join 2 segments**. The tool **CA build | Check CA direction** can also be used to indicate the quality of the build if the pie chart is not visible long enough to determine the quality graphically.

Checking the trace for knots

You can check for knots in the Cα trace you have generated by using the tool **X-POWERFIT | Knot finder**. This will take the current segment of Cα trace (select using **X-POWERFIT | Current res. and seg.**) and indicate in the text port whether a knot was found.

Cα refinement

Segments of placed Cα trace can be refined with **X-POWERFIT | CA refinement**, particularly where secondary structure elements have been added in places where the density indicates deformation within the secondary structure. There is a limit of 250 Cα atoms in the refinement because of a limitation in the algorithm to remove the correlation along the Cα chain which is unable to handle large fragments.

Geometric restraints on Cα–Cα bond lengths and on some angles are used to maintain better geometry during the refinement.

Generation of all atom models

We recommend that where reasonably high-quality maps are available, the tool that places the atoms in an “all-atom” model by real space refinement should be used (**CA build | Fit seg. by RSR**). Where the map quality is less

good, or the resolution low, then it is recommended that you use the tool to fit the main chain by theoretical modeling and the side chains by real space refinement is used (**CA build | Fit seg. by CA correlation**). You should add all the sequence information to the Cα trace before building an "all-atom" model, using the sequence assignment tools provided (**Sequence...** palette). The building tools use the sequence information assigned to the Cα trace to automatically place the side chain atoms.

8. X-AUTOFIT:X-BUILD Tools

Introduction

This chapter provides a quick look-up reference for selection on the X-AUTOFIT:X-BUILD palette. This palette provides the tools to generate and modify bones, skeletons and sequences.

To start X-AUTOFIT

X-AUTOFIT is accessed from the **Applications** menu in the QUANTA main menu:

1. Create a new directory in which to run X-AUTOFIT. Move to that directory, and start QUANTA.
2. Select **X-AUTOFIT** from the **Applications** menu. The main X-AUTOFIT:X-BUILD palette and the Pointer palette appear, as well as a graph window containing the allowed regions for a Ramachandran map, and any points representing the phi-psi angles for any known coordinates. If the map table is not open, then open it using the **Map show table** (select **Map table | Show map table** from the **Draw** menu) tool.

   An Object Management table will also open and will fill with objects as they are generated for the following: bones, mask, symmetry atoms, Cα-trace, and 3D text.

   The Object Management table can be used to toggle the relevant information on and off. Objects can be deleted from this table, but X-AUTOFIT will generate these again if required.

> **Note:** If you exit X-AUTOFIT without using **X-AUTOFIT | Finish**, the next time you start X-AUTOFIT, a prompt appears asking if you want to recover from the last building session. If you select **Yes**, this recovers the changes made to the coordinates from the last building session automatically.

When you start X-AUTOFIT, the X-AUTOFIT:X-BUILD palette is displayed as illustrated, along with the Geometry palette and X-AUTOFIT dials emulator. The Ramachandran map opened is described in full in Using X-BUILD (page 77), and the Cα plot is described in Using X-AUTOFIT (page 49). The presence of these plots and how to control their use, such as

switching between the different displays, is described in Graph windows (page 47).

On first use of X-AUTOFIT, if a map is currently open, then a map contoured with a radius of 9 Å will be displayed at the center of the screen display. On subsequent entry to X-AUTOFIT, you will find that any other palettes that were open on the last exit from X-AUTOFIT are also opened. Also, any other addition information such as bones, Cα trace, sequence information, etc., are displayed on the screen if they were present on exit from the last X-AUTOFIT session.

The following information is saved between X-AUTOFIT sessions and is effective on subsequent use:

- The current working position
- The current open palettes.
- The map display radius.
- The coordinate display radius.
- The symmetry display radius.
- If bones are active.
- The bones start parameter.
- The bones delete and trim values.
- Whether bones are smoothed, and side chains are visible.
- The current map used for real space refinement and bones.
- Non-bond parameters and optional restraints in use.
- Regularization parameters.
- The type of alternate conformations being handled.
- The rotamer library used.
- The action on picking the Ramachandran plot.
- The range of residues to draw on the Ramachandran plot.
- Whether to draw the Ramachandran/DNA circle plot.
- The next step value.
- The Cα packing radius.
- The current residue number.
- The atom picking radius.
- The memory pie chart display/hide.

- Dialog box field values.

The majority of the parameters described here are set up on the X-AUTOFIT | Options dialog box. For a description of these parameters refer to *X-AUTOFIT Options dialog box (page 204)*. The memory pie chart flag and the picking radius can only be changed by editing the file xfit.pack.

Map options

This tool sets up the type of map to be calculated and also the reflection file to be used. This is based on the map calculation dialogue from the CNX interface. Note that the bottom three options should be ignored. Map calculation should always be “immediate”.

Calculate map

This tool launches a series of calculations which result in a map being generated and displayed on screen.

The map calculation will be performed taking into account any changes in topology or conformation of the molecular system under study. See Managing Maps (page 25) for more information.

Symmetry palette

The Symmetry palette allows the definition of symmetry and NCS (noncrystallographic symmetry). Several symmetry objects can be created with tools on this palette.

A separate graphical object is generated for each symmetry element, labeled with the symmetry operator. Individual symmetry objects can be toggled on or off (using the **Displayed** column in the Object Management table) or deleted with the Object Management table.

Each symmetry object has a different name, formed from the translation and symmetry transformation used to generate it. For example, the molecule generated by applying a unit-cell translation of 0,1,-1 in a, b, and c for symmetry operator number 3 is named S+0+1-1:3. If any noncrystallographic symmetry operator is applied, the name also includes this information in the true symmetry operator.

Symmetry atoms can be picked and the information about parent real atoms is returned along with the symmetry detail for the symmetry atom picked.

You can delete all symmetry objects by using the **Delete Sym.Ob** tool in the Object Management table, which deletes all graphical objects whose name starts with S and contains a colon (:). If you want to retain a particular symmetry object, change its name (using the **Name** cell in the Object Management table) before using **Delete Sym.Ob**.

Define symmetry

Define Symmetry allows definition of the symmetry of the current molecule or can be used to change the symmetry of the current molecule. This information is not written to the MSF until **X-AUTOFIT | Finish** and then the new MSF files contain the new symmetry information.

1. A dialog box that prompts for either the space group number or space group name. For example, r3 is 146.

2. A dialog box for the cell dimensions. The number of lines to enter on this dialog box will depend on the point group, as some entries on cell dimensions and cell angles are degenerate. If a map is currently open, the default cell dimensions will be read from the map header, so may not need changing.

3. A dialog box for the ncode. This defines the axis order for the cell. Normally the ncode is one.

The symmetry atoms will be updated based on the new symmetry specification.

Define NCS symmetry

The Non Crystallographic Symmetry dialog allows the specification of the real-space rotation matrix and translation that defines the position of NCS related atoms. A dialog box appears:

Next allows the specification of further matrices.

Exit exits the NCS definition. If the current matrix is a unit matrix then this is *not* added.

Quit makes no change to the current NCS symmetry.

The matrix provided by the application is for the identity matrix. The values that specify the NCS should be typed into the matrix. If you do not provide a matrix with a unit value determinant you will *not* be allowed to enter

further matrices with the **Next** option or to **Exit**. A warning appears in the textport to indicate the presence of a undefined matrix.

The NCS information is written to the key worded free format symmetry file *molecule-name*.sym on exit from the NCS tool editing tool. The NCS symmetry matrix is written out with a NSYM keyword and the 12 matrix values on 4 lines of the file (the first 3 on the same line as the NSYM card). It can be easier to add NCS symmetry information directly to this file as multiple NSYM cards. Some programs define the NCS information as the transpose or rotation part of the matrix depending on the internal representation. The transpose of the NCS matrix can be added to the symmetry file with the keyword NTSY.

On exit, the NCS atoms will be calculated and drawn as red (color 3) atoms. You can define up to 60 NCS matrices. To delete all the NCS, exit with unit matrix as the first matrix.

Unit cell

Clicking his tool generates a unit cell box. The box can be removed again by deleting the object from the object management table.

CA Packing Diagram

The **CA-Packing Diagram** tool generates a number of symmetry copies of the current visible and active molecules in the molecule management table, where the copies are shown as Cα traces. This is irrespective of the currently displayed atoms within the molecules. Each molecule is drawn a different color, using colors 1 to 14 in cycles of 14. The number of copies generated is defined as the number of generated molecules that have a centroid position within the **Packing radius** distance from the centroid of the active displayed molecules. A separate graphical object is created for each symmetry related molecule.

Packing Diagram

The **Packing Diagram** tool generates a number of symmetry copies of the current visible and active molecules in the molecule management table, where the copies are shown as all-atom representations. This is irrespective of the currently displayed atoms within the molecules. Each molecule is drawn a different color, using colors 1 to 14 in cycles of 14. The number of copies generated is defined as the number of generated molecules that have a centroid position within the **Packing radius** distance from the centroid of

the active displayed molecules. A separate graphical object is created for each symmetry related molecule.

…Packing radius

The …**Packing Radius** is used for generation of the Cα-packing and packing diagrams and defines the number of copies to make. The number of copies generated is defined as the number of generated molecules that have a centroid position within the **Packing radius** distance from the centroid of the active displayed molecules.

CA Filled Cell

The **CA Filled Cell** tool generates objects that constitute all the symmetry copies that fill the unit cell, regardless of whether the original molecule was within the unit cell (0-1,0-1,0-1). Hence for the space group r3, nine copies of Cα-traces are produced.

Filled Cell

The **Filled Cell** tool generates objects that constitute all the symmetry copies that fill the unit cell, regardless of whether the original molecule was within the unit cell (0-1,0-1,0-1). Hence for the space group r3, nine copies of all-atom models are produced.

Delete Sym. Ob.

This tool deletes all symmetry objects that have been previously drawn.

Write Sym. to PDB

This tool writes out a PDB file containing all the displayed symmetry copies.

Re-arrange atoms

This tool re-arranges waters, ligands, and heavy atoms based on the best non-bonded interactions, center-of-mass placement, or within a map mask.

Hide this menu

Exits the Symmetry palette.

Pointer palette

This controls the display and positioning of the pointer. The pointer provides a visual cue for marking the location where certain commands will occur and allows you to step through a series of atoms or residues.

Show pointer

This turns on and off the display of the pointer, whose default appearance is a white tetrahedron in the molecule window.

Show Ruler

The **Show Ruler** tool displays a ruler at the bottom of the molecule display. The ruler provides a helpful scale for experimental maps containing scant molecular data, and allows you to determine the size of structures seen in the electron density maps. The tool is a toggle, and picking the tool again will remove the ruler from the display.

Map table

The **Map table** tool toggles the map display. Display of the map table can also be controlled from the main menu bar or by entering **MAP TABLE** [**ON** | **OFF**] on the command line.

Interactive contour

The **Interactive contour** tool opens the Interactive Contour palette, which allows you to interactively change the level at which an electron density is contoured (Interactive Contour palette).

Active residue on and Active residue off

These two tools change the modality of X-BUILD by defining whether a tool prompts you for the residues/atoms to act on, or whether a tool acts on a defined *current* residue. The current residue is the same as the definition for the Ramachandran highlighted value and is set by picking a displayed atom, or any **place by** tool defining an atom/residue.

If active residue is on, the *current* residue is highlighted (as well as the current atom indicated), and any subsequent tool acts on this *single* residue without prompting. You can change the current residue by picking any atom within a desired residue. This allows you to more easily select residues, without the prompt/select step. The aim is to make model building more efficient, but this option does generally allow only one residue to be edited at a time.

It is not possible to define selections of multiple residues (i.e., more than one active residue), because this leads to ambiguity when using some tools. To work around this, two new tools are supplied on the Structure palette allowing range and volume minimization from a single pick.

Pointer dials

This changes the dial palette to show an xyz move dial set. Because different parts of the program assume the default mode of the dials (for example, placing Cα atoms, normal dial set, cursor mode), it is sometimes necessary to tell X-AUTOFIT which dials you wish to use. This palette option provides the dial set for the pointer movement. Selecting the pointer dials also results in the Ramachandran plot being drawn, if it is not already present.

Go to pointer

This causes the display to center at the current pointer position. The pointer becomes the center of the display and is the center of the rotation of the display. If the map is active then the map is moved to this position. If the bones are active then these are calculated around the same point.

Place at next residue

This tool moves the display to the next residue in the molecule. If the residue is an amino acid then the display will be centered on the Cα atom of this residue, otherwise the display will be at the first atom of the next residue. This point becomes the center of the display, and results in moving the

map and recalculating the bones here. Specify the increment value for the next residue in the **Options... | Next residue step** value field.

Place at previous residue

The **Place at previous residue** tool behaves like the **Place at next residue** tool except it moves the center of display one (or more) residues backwards along the polypeptide chain. All components of the display, the map, bones, and symmetry where applicable, are regenerated centered on this point. Specify the increment value for the next step along the polypeptide chain in the **Options... | Next residue step** value field.

Place by atom

A dialog box prompts for an atom name, segment name, and residue number. The default values are for the first atom in the data structure. When you press OK, the center of display moves to this atom. If you press **Cancel**, then nothing happens.

Place using bones

This is only active if bones are on and displayed. When you select this tool, the program asks you to select a bones point. The selected point becomes the center of display, and results in moving the map and recalculating the bones. To abort the bones pick, select the tool again, and no movement occurs.

Place using coordinates

Prompts for an atom at which to center the display. This point becomes the center of display, and results in the moving the map and recalculating the bones. To abort the pick, select the **Place using coord** tool again, and no movement occurs.

Place by 3D position

A dialog box appears that allows you to enter specific x, y and z coordinates to place the pointer.

Suggestions

This item launches a help browser dialog which allows you to search the online HTML help documents using multiple keyword searches. The default help book is the Xfit how to, such as how to edit masks (Thesaurus.html). You can search the help page titles using a multiple keyword search which shows a subset of the available pages that include the keywords selected. This book also contains answers to frequently asked questions associated with the Cα-tracing functionality and model building features.

Type a word or multiple words separated by spaces in the search string line and press <Enter>. The large scroll list at the bottom of the dialog will show all the target hits for *all* the search words.

To view a help option, click the scrolling-list line that is of interest to you — the html browser opens with the help information on this topic.

The same tool can be used to browse any of the help books by picking these from the top scrolling list, so this tool can be used for general help book reading.

Hide this menu

Puts away the Pointer palette.

Interactive Contour palette

The Interactive Contour palette allows you to interactively change the level at which the electron density is contoured. Only one contour level of one map is displayed while using this tool, regardless of the display status before you open this palette. By default, the first contour level is shown when the palette opens. The palette allows different contours and maps to be selected so that they can be manipulated interactively one at a time using a <SHIFT><CTRL> mouse action.

Next Map

Moves to the next map (if any) in the list. The first contour level of that map is shown.

Previous Map

Returns to the previous map (if any) in the list. The first contour level of that map is shown.

Next Contour

Moves to the next contour level (if any) for the displayed map.

Previous Contour

Moves to the previous contour level (if any) for the displayed map.

Accept

Stores your contouring changes, updates the display, and closes the palette.

Quit

Discards any changes, closes the palette, and returns to the display shown before the palette was opened.

Notes

For faster interaction, the map radius (in the X-AUTOFIT Options dialog box) should be set to a value appropriate to the computational power of your local machine. Also, since the bricked map file is read continuously during the interaction, it is better that these files be on your machine rather than accessed over a network.

You should have the map table visible when using the interactive tool, so that you can see the contour levels for all maps and levels. The map table retains the previous values until the map or contour level is actually changed when you are using the interactive tool.

The current map, contour level, and actual map contour value are displayed on the message line of the main molecule window.

Bones palette

The Bones palette is used for setting up the bones calculation. If the bones are turned on then all positioning of the screen origin results in the bones being calculated around this point at the current map radius.

Bones on/off

When on, the bones are calculated about the current position. Because bones take a significant time to calculate, turning on this option slows down the display.

Mask bones by mask

The **Mask bones by mask** toggle option automatically trims the bones that lie outside an open mask. On subsequent displays the bones outside the boundary are excluded from the display and from all calculations. This means that the mask becomes integral to the use of X-AUTOFIT and prevents the building of any structure in a region of the map that is symmetry related. This is particularly a problem when the crystallographer is first tracing a map and no symmetry atom can be generated since they have not yet been built.

When the toggle is on the bones are immediately trimmed to the boundary of the open mask, and any subsequent calculations are of trimmed bones. If the tool is then toggled off, trimmed bones are *not* automatically regenerated. Use one of the centering options to regenerate them if desired. If no mask is active when this tool is toggled on, trimming does not take place and an error message is written to the textport.

When the tool is off, no change is made to the bones.

Set bones/RSR map

This option displays a list of the currently open maps and allows you to choose the map for the bones and RSR calculations. If there is only one map, this tool does nothing. If you select **quit**, the map selection does not change. If you select **OK**, the new map is used for subsequent bones calculations and real-space refinement calculations. The newly calculated bones appear immediately.

Find nice area of map

This tool searches the entire map for the region that contains the largest amount of significantly high density. Once found, this region becomes the center of the display, and the map and bones are contoured here. Use this option when you are first looking at a new map and no coordinate information is available.

Map quality from bones

This selection sends to the textport information about the number of connected bones trees and their sizes, in percent of total bones. This data indicates the level of under-connectivity of the map. The textport also lists the number of false links, which indicates the level of over-connectivity.

Smooth bones

This option is off by default, and bones appear partially smoothed. Use this selection to display bones as cubic spline curves that pass through map points. Each curve is one segment of the bones tree, so branch points and termini are exactly as found on the map. Intermediate points are approximated. You are only able to select those points that are actual points of the map.

Side chains on/off

You can show or hide the sidechains in the skeleton by toggling this tool. Sidechains are displayed in color 6 (default value is white). Combine this selection with scaling to observe a larger portion of the skeleton. This selection can be reversed by selecting **Undo last delete**.

Main <-> side

With this selection, you can change the type of bones strand between sidechain and mainchain. When a bones section has a status change, its color changes from yellow for mainchain to white for sidechain. Use the **Undo last** selection to undo your last selection. If sidechains are hidden and you change a mainchain to a sidechain, the section disappears from the display.

Delete 1 section

This selection allows you to remove bones points from the bones skeleton. When you make this selection, the message prompt at the bottom of the molecule window instructs you to select a bones point by clicking with the mouse. The bones points that are deleted are multiple points that lie in a single chain extending from a branch point or terminus to another branch point or terminus. Any point selected from a section results in the entire bones section being deleted. You can undo your last selection by selecting **Undo last delete**.

Any recalculation of bones (for example, through changing the bones parameters or moving to a new bones box) overrides the delete modifications. Also, deleting a branch changes the smoothing function so some branch points move when a section is deleted.

Delete fragments

This deletes one fragment of bones, where all the bones in the fragment are joined to give a single tree from the point selected. The tool activates the delete fragment action which remains active until another palette option is selected (such as **Delete fragment** again). To delete multiple fragments from the display, continue to pick points. The bones display is redrawn to show the deletion. The **Undo last delete** undoes the entire deletion process, and selecting the **Undo last delete** tool again removes the deleted points again.

Delete all fragments

This tool allows automatic deletion of bones by removing bones fragments by size. The tool removes the bones points on the basis of fragment size as an aid to map mask generation. All the bones fragments in the current map are calculated, and the total number of bones points is determined. The tool then removes the fragments of bones (that is, connected sections of bones points), that are smaller than the current fragment delete threshold. The initial threshold of deletion is set to 1% of the total number of bones points. Each subsequent use of the tool first doubles the threshold value and then deletes the bones based on this new threshold value. You should therefore use this tool repeatedly until a section of the map you actually wanted to keep is deleted. Then use the **Undo last delete** tool to reverse only the last deletion. This is probably the quickest method of deleting most of the small, unwanted parts of the bones for mask generation. The tool **Reset delete all** resets the threshold value to 1%.

The tool does not remove small sections of map that have no mainchain component. Use the **Delete fragment** tool to delete these.

Reset delete all

The **Reset delete all** tool resets the threshold value for deletion to 1% for the **Delete all fragments** tool.

Undo last delete

This returns the most recently deleted fragment or section of bones. Clicking this tool a second time reverses the undo operation, thereby repeating the deletion of the bones.

Create bones atoms

This tool creates a PDB file called bones.pdb containing single atoms defined by bones points and displays bones pseudo atoms from points separated by 1.2Å. These points are generated from the full bones point list by progressively deleting points until only the highest points in the bones trace are left. The output is an object in the object table and can therefore be deleted from the object table.

The bones.pdb file is a standard PDB-format file and can be used as part of a refinement procedure without restraints. This type of atom generation has been found to be a very good starting point for maximum likelihood unstrained refinement, since the atoms are already well placed.

Change start value

This tool opens the Set up bones parameters dialog box. This dialog box allows you to manually set the start value for the bones calculation. When you click **OK**, the bones are recalculated. Recommended values are:

- For a MIR/SIR/MAD map1.2
- For a 2fo-fc map1.2
- For a 3fo-2fc map1.8
- For a sigma A weighted map 1.0 to 1.2

Change trim/analysis

This opens a dialog box to change the delete value for bones trimming and the sidechain detect value. A value of 1–4 is good for the trim, and 10–30 for the sidechain detect value.

Pieces of bones smaller than the trim value are deleted. Increasing this value makes a cleaner display, but you lose detail.

The **Side chain detect** value is the maximum length of a branch that is classed as a sidechain. The program executes (**Side chain detect**)/4 calls to the detect function, which causes branches to be continually trimmed back. Increasing the **Side chain detect** value increases the number of sidechains detected, but results in some mainchains being classed as sidechains. Smaller values detect fewer sidechains. Recommended values are 18–20 for proteins that have more than 50 residues and a moderately connected map. You need smaller values for small peptides or where the density is badly broken. Sidechains smaller than the sidechain detect level are displayed in color 5 (default value is white).

Calc bones symmetry

When the **Calc bones symmetry** option is selected, the bones generated by symmetry are displayed. The symmetry-related bones are drawn in color 2 (default value is blue) and also are generated in a reduced representational form. Although the reduced representation of the bones is not as elegant as the real bones, this prevents overloading of the graphics with a very large symmetry-related object. The symmetry bones are provided to show mask overlaps when creating a solvent mask. For Cα tracing, the symmetry is shown more effectively by the symmetry-related Cα trace. The symmetry of the bones is not updated on a bones calculation, they are only regenerated by clicking this tool. This is because there is a significant delay (1–2 seconds for a large bones region) for the symmetry generation. Therefore, to update the symmetry-related bones, select this tool again.

Symmetry off

Turns off symmetry bones.

Hide this menu

Puts away the Bones palette.

Map mask palette

The Map mask palette controls the generation of a map mask for solvent flattening. If generating the mask from bones, then use the Map mask palette options in conjunction with the Bones palette. The Map mask palette allows the generation of masks from bones and coordinates in the current map unit cell. It also allows interactive mask editing and has a progressive reduced representation form that allows the study of large masks even on low-powered systems. X-AUTOFIT currently reads all forms of O masks, (New, Old and Compressed) and writes O compressed format masks.

Calc. mask from bones

This calculates a mask from the current bones using a masking radius (default value is 4.0Å), defined by the option **X-AUTOFIT | Map Mask | Mask delete radius**. The maximum extent of the mask is defined by the displayed volume of the *currently* active map as set by the **Map radius** on the **X-AUTOFIT | Options** dialog box. You can edit the bones from the Bones palette, and see the symmetry overlap when the bones symmetry is on. After the mask calculation has completed, a spherical mask pointer appears that you can use to edit the mask surface.

Calc. mask from coord

This will calculate a solvent mask that covers all currently active and displayed coordinates using a masking radius (default value is 4.0Å), defined by the option **X-AUTOFIT | Map Mask | Mask delete radius**. The extent of the mask is defined by the displayed volume of the *currently* active map as set by the **Map radius** on the X-AUTOFIT:X-BUILD | Options... dialog box. If you want the mask to cover the entire map, set a large map radius before using the **Calc. mask from coord** option.

After the mask calculation has completed, a spherical pointer appears that you use to edit the mask surface.

Mask off

This removes the mask and deletes any reference to the mask from the object table.

Mask dials

This tool forces the dial set to switch to mask editing mode only if a mask is present. This tool will also change the pointer to the spherical mask pointer.

Mask delete radius

This provides a pop-up box to change the value of the mask delete radius, which defines the volume about each atom coordinate or bones point for which the asymmetric unit is masked as protein. The default value is 4Å.

Solvent content

This tool provides an estimate of the solvent content based on the volume of the current mask, the number of symmetry and NCS operators, and the volume of the unit cell.

Add mask at pointer

This changes the mask extent. If the mask pointer is moved to a region that is covered by the map (and so by the mask grid), but the mask does not cover this region, then this tool will increase the mask to cover this volume. If you use this option at the surface of an already-present mask, the mask volume will *increase* in size. The mask surface is recalculated as part of the process and may take between 0.1–5 seconds, depending on the extent of the mask and the system running X-AUTOFIT.

Del mask at pointer

This changes the mask extent. If the pointer is moved to a region that is covered by the map (and so by the mask grid), but the mask covers this region, then this tool will decrease the mask extent so that this volume is solvent. Therefore, if you use this option at the surface of an already-present mask, the volume will *decrease* in size. The mask surface is recalculated as part of this process, which may take between 0.1–5 seconds, depending on the extent of the mask and the system running X-AUTOFIT.

Check for voids

This tool provides an automatic method to remove small regions within the mask (that is, protein) that have been flagged as solvent. Such small regions represent high resolution detail that should not be present in the mask. This tool does a void calculation on the mask grid and determines if there are any regions within the mask that are not accessible from the surface of the mask. These buried solvent regions are removed so that the points in the grid that form this void are changed to be *not* solvent points. The display changes to reflect the results. Depending on the size of the mask, the calculation will take between 1–20 seconds.

Increase resolution/ Decrease resolution

If a mask is very big, or the machine being used has only moderate graphics facilities, you may wish to use these two tools to adjust the number of points used to describe the surface. When the mask is first calculated, every point on the surface of the mask is shown by a point. If the resolution is decreased or increased, then 1/*n* of the points are randomly left in the surface, where *n* has a value between 1 and 10. Therefore, initially there are 1/1 points in the mask (all of them). If you decrease the resolution five times, then 1/5 of the points will remain in the surface. This reduces the interpretability of the surface, but allows the mask to be more easily manipulated, since the object is five times smaller. These commands will *only* affect the number of displayed points on the surface, and will have no effect on the stored mask or any calculation on the mask. You should also note that when using a reduced resolution mask surface, any re-display of the mask (such as when removing voids, add or deleting mask with the pointer), causes the distribution of points on the surface to change. This is because the points are randomly selected, and this selection will change on each calculation.

Save mask to file

The mask currently displayed is written to disk as a compressed O format file. If the mask was read from a file, the cell and extent are determined by the original mask file. If the mask was generated in X-AUTOFIT then the cell and extent are determined by the current map and its extent.

Read mask from file

A mask in an old or new or compressed O format file can be read into X-AUTOFIT. If no mask is present, a new object is created and displayed.

If a mask exists, the mask that has been read from the file replaces the old mask. This allows existing masks to be edited.

Hide this menu

Puts away the Map mask palette.

Text palette

The Text palette controls the creation and use of annotations associated with points in the macromolecular structure. You can create and place text, delete text, and go to a text position. You can add up to 500 text notations.

New text at pointer

When you select this tool you will be prompted for a text string (up to 80 characters long). This text appears at the current pointer position. To place text at a different position, move the pointer to that spot before creating the text.

Delete text

This deletes the current text. You select the text to delete from a scrolling list of all the current text strings. Pick one string and select **OK**. If you pick **Cancel**, nothing is deleted.

Goto next text

This moves the current display center to the next text in the list. If the last text is the current text, you are not advanced to another string and a message prints to the text port. The display center, map, and bones are updated.

Goto previous text

This moves the current display center to the previous text in the list. If the current text is the first text, you are not advanced to another string and a message prints to the text port. The display center, map, and bones are updated.

Goto defined text

This creates a palette that contains a scrolling list of all the current text strings. You can sort the list by using the **Sort** button, and unsort the list by clicking on this button again. The default order of the items is the order in which they were created. You can pick a text entry from the list and the program will center on this text position. The map and bones will be displayed at this position if they are active.

...Fix Validate error

This tool works in conjunction with the **Validate** tool on the main X-BUILD palette. Once validation is complete, the text markers created are linked to the error correction technique required to fix the errors (that is, bond errors are fixed by regularization). Any other text strings added by any other functionality are ignored by this tool. The current text error is fixed with this tool, and the text marker is changed to show that the error/warning has been fixed. Other validation errors may also be fixed, and these will also be updated to show the fix.

Load Property

This brings up a scrolling dialog box of property information that can be loaded into the text editor. Currently these are:

- **User text labels**: text strings that you entered.
- **Termini**.
- **Ligands/ions/etc.**: whatever is not water or protein.
- Water molecules.

You can pick an entry from this list, and the program will load in the relevant information. For example, if you load the water molecules, you can then use **Goto next text** and **Goto previous text** in the text editor to look at each water molecule in turn. The program does *not* write any of the property lists to the session file. When you exit the program, the list of user text labels is saved. You cannot add to or delete from the property list. You must use this option to load in the user text labels list of text items before you can add or delete text markers.

Hide this menu

This puts away the Text palette.

CA Build palette

Load CA coordinates

This tool reads *only* the Cα atom coordinates from MSF information into **X-AUTOFIT | CA Build**. This allows homologous information to be read into **X-AUTOFIT | CA Build** to form the basis of the new structure while using the auto build algorithm to build new all-atom information only from the map. You might need to do this when a molecular replacement solution is to be generated. The new Cα trace is added to the current Cα trace information, so if you are loading an entirely new structure from an MSF, you should first delete all the old Cα trace information. The tool loads in the Cα coordinates for all visible and active molecules.

Next bones box

This selection generates a new skeletonization calculation centered on the current alpha carbon. You can select any alpha carbon atom as the active one to specify where the next bones box will appear. The origin of the bones box is shifted to this alpha carbon, and a new skeleton and map contours are displayed. Use this selection when your alpha carbon trace reaches the edge of the current bones box.

Next CA

This selection accepts the position of the currently built alpha carbon and provides you with another alpha carbon to position. Next, Cα automatically executes a **Guess next CA** calculation and attempts to position the newly added alpha carbon. When you select **Save changes**, the currently active carbon is included in the saved set. If you cannot position the current alpha carbon, use **Delete current CA** to remove it before you save. You can also use the **Undo last build** selection to remove the current alpha carbon.

Repeatedly using **Next CA** does not have the same effect as using the **Auto Extend** tool in the X-POWERFIT palette, because auto-extension can build on information from previous Cα placements to improve the fit of the next Cα, while manual extension cannot.

Delete current CA

This selection deletes the current alpha carbon and makes the previous alpha carbon currently active. If only two atoms are left in the segment, no more can be deleted.

New segment

This selection starts a new segment of alpha carbons at a new starting point on the skeleton. The Accept or Quit popup is displayed with selections that you can use for choosing the new point (see Pointer palette (page 127) for a description of palette selections). Selecting **Undo last build** will delete the last new segment.

If you want to start at a point in the electron density map that is not currently skeletonized, you must recenter the skeleton using either the bones to pick a new point on the map, or the pointer to place the map center and then use **X-AUTOFIT | Pointer | Goto pointer**. If there is already at least one segment present, the new segment becomes active (default color red) and the previous segment becomes inactive (default color blue). You can return to the previous segment by selecting **Current res | seg**.

Delete current segment

This selection deletes the currently active segment and makes the previously built segment the currently active segment. If only one segment has been built, it is deleted. Selecting **Undo last build** replaces the deleted segment.

Reverse chain

This selection reverses the order (polarity) of the current segment. The current alpha carbon becomes the origin of the chain. The chain can be reversed at any time. This allows the current segment to be built in either direction. Selecting **Undo last build** will reverse the chain again.

Current res/seg

Use this selection to change the alpha carbon and segment you currently have selected. The current alpha carbon is yellow and the rest of the segment is displayed in red. All other segments are displayed in blue. When

you make this selection, the message line instructs you to select an alpha carbon. The nearest alpha carbon to your screen selection becomes the active carbon and the segment containing this carbon becomes the active segment.

If you pick a terminal alpha carbon, all selections on the palette and the dials remain active. If you pick an alpha carbon that is not a terminal alpha carbon, then the dials change so that the xyz position of the alpha carbon can be changed. The Cα plot changes to indicate the conformation of the alpha carbon you have selected.

Join 2 segments

Use this selection to join two segments of a Cα chain. The selection is grayed until at least two Cα traces have been generated for the bones skeleton. When the segments are joined, the single segment that is formed becomes the current segment.

Unjoin 2 CA

This tool allows you to break up a segment. This is necessary if you have read your Cα atoms in from a file generated by O. O treats all fitted Cα atoms as a single segment regardless of the connectivity, but X-AUTOFIT treats the different fitted sections of Cα atoms as different segments. This incompatibility causes X-AUTOFIT to

join up all the fitted segments with very long bonds. The **Unjoin 2 CA** tool prompts you to pick a Cα-Cα bond. The bond nearest to the pick will be deleted. This splits the single segment into two segments; one becomes the active segment. If any sequences had been assigned, a sequence alignment is carried out for the two new segments, and any changes are updated.

Add helix/strand

This selection displays a pop-up to define the addition of a helical or beta strand Cα-trace whose length is defined by the number you type.

Select the secondary structure element to fit, and its length. The default is a 10-residue helix. The dial box will change to allow you to adjust the orientation and position of the element, the Accept/Pick dialog box will appear, and an idealized helix/strand object will appear at the current pointer position. You can use the dials to move the object until it is in the correct position in the density. The position can then be accepted from the Accept or Quit dialog box. No other tool is available while this positioning

is in process. If you select **Quit**, no changes are made; if you select **Accept**, this idealized secondary structure element will become the current Cα-trace and the last Cα atom will become the current Cα atom. This trace can then be edited as any other trace of Cα atoms, or can be used as a template to aid in the building process, and then deleted at a later time with **X-AUTOFIT | CA Build | Delete segment**.

Move current seg

This tool allows the movement of the current segment as a rigid body. The dial box changes to allow the movement in x/y/z and rotation about the x/y/z screen axes. The Accept or Quit dialog box will appear. If you select **Quit**, no change is made to the coordinates. Otherwise the current segment is moved to the new position.

Refine current seg

This tool carries out real space rigid body refinement of the current segment. This is most useful to improve the fit of a secondary structure element that was fitted with **X-AUTOFIT | CA Build | Add helix/strand**.

During refinement, a white line model of the segment is shown, and on completion, the Accept or Quit dialog box appears to allow you to either accept the new position or reject the changes.

Check CA direction

This tool assesses the probability of the polarity being correct and reports that information in the text port. If you have defined any residues for any segments in the molecule, sequence alignments are marked in the molecular sequence table at the top of the molecule window. Blue arrows indicate forward alignment and red arrows indicate reverse alignment.

When you select this tool, X-AUTOFIT attempts to fit polyglycine to the trace in both directions, checking geometries. The following information is then reported in the text port:

- The percent fit.
- A statement that the chain is the correct/wrong way around.

- A fit ratio that gives a probability on the correctness of orientation.

> **Note:** The percentage likelihood values for the forward and reverse directions are independent values. The sum will not necessarily equal 100%.

The polarity of the active alpha-carbon trace can be reversed (**Reverse chain** option) so that building can be carried out at either end of the chain. You should evaluate the polarity of the alpha-carbon trace before you generate a peptide backbone.

Show points ~ 3.8Å

This selection toggles a set of markers on and off. They highlight all the points that are 3.8 ± 0.3Å from the current alpha carbon. Use these as possible positions for placing the next alpha carbon.

Guess next CA

This selection calls an assisted Cα building routine that selects the best conformation for the next alpha carbon. This selection is based on connected density, branch points near the test position, a reasonable Cα-Cα-Cα angle, and main chain/side chain bones.

Repeated picking of this tool will move the current Cα atom to a new guessed position. The Textport will print out the point number fitted and the rank order of likelihood. For example:

```
Number of OK points =            4
Point  1  main chain : impossible angle  No branch
Point  2  main chain : OK angle         near a branch
Point  3  main chain : impossible angle  near a branch
Point  4  main chain : unlikely angle    near a branch
Point   2 of   2 : order no. =   2
```

Note, as seen from this example, that “impossible” conformations (where the Cα trace would fold back on itself), are not counted in the number of valid points, and are not displayed by the **Guess next CA** tool. So in the above example, there are four theoretical points for the current Cα, only two are valid; the second is the currently shown on the display, and this is the most likely point.

Fit seg by RSR

Using the real space refinement procedure, this selection generates a polypeptide chain based on alpha-carbon coordinates that fits the electron

density map. Color indicates the quality of the fit of the atoms. A green (by default) atom has been fitted well, a blue atom indicates that the atom has not been fitted because there was no density in the region. Other fit qualities scale between blue and green. If there is a sequence fitted to the Cα trace, then the all atom model for these residues will be build. If no unique sequence assignment has been made, then only a polyalanine model will be built.

If the sequence has been aligned, the sequence number for the built segment is taken from the sequential position of the Cα atoms in the sequence. If no sequence alignment has been set, the sequence is numbered beginning with 1.

Fit seg by database

The tool **Fit segment by database** allows the generation of all atom models based on database fragment building The fragment building algorithm uses the Cα distance matrix to determine five matches/fragment within the database of protein structures. For each five residue fragments, the five Cα-trace fragments are tested from the database to find the most likely conformation of backbone atoms. The backbone atoms are then merged together to form a contiguous polypeptide chain of polyalanine. If sequence information is known then the side chain atoms are added using grid real space torsion angle refinement into electron density.

Note that **Fit seg by RSR** is preferable to this method since it is statistically valid over a wide range of conformations and should always be used where the density is good enough to fit the side chain atoms. It is recommend that, although this type of building method is well known, you should use the tool **Fit seg by CA correlation** as this produces very precise results with very few errors.

If the sequence has been aligned, the sequence number for the built segment is taken from the sequential position of the Cα atoms in the sequence. If no sequence alignment has been set, the sequence is numbered beginning with 1.

Fit seg by CA correlation

This tool uses a superior algorithm to the **Fit seg by database** tool as a method of theoretical building of polyalanine polypeptide chains from Cα traces where there is little density. The tool uses a matrix of values that directly correlate values of Cα conformation found in the protein databank and the Ramachandran angles associated with these Cα conformations. The first use of this tool reads in the matrix of correlation values. Subsequent

uses of the tool access the matrix from memory and are almost instantaneous. The algorithm determines the local Cα conformation, and then directly builds in polyalanine using the Ramachandran values within the correlation matrix. The algorithm reproduces polyalanine coordinates from Cα traces with RMSD values of 0.3A to 0.5A in test calculations. The very low residuals result because the entire information within the protein databank is used, but statistically removing the outlying results from the lower resolution crystal structures. Hence loops are generated correctly.

After fitting a polyalanine chain into the Cα trace, the tool fits the side chain atoms if the sequence information is known using real space torsion angle grid refinement.

If the sequence has been aligned, the sequence number for the built segment is taken from the sequential position of the Cα atoms in the sequence. If no sequence alignment has been set, the sequence is numbered beginning with 1.

Fit seg by D.E.E.

It is not possible to fit main chain and side chain coordinates when the experimental information is very poor or non-existent. No map fitting occurs. The **CA-build | Fit seg by D.E.E.** tool fits main chain atoms using the Cα-Ramachandran correlation table and models the side chains by adjusting the conformation of the side chain chi angles and of multiple neighboring residues. The energy is computed for each possible conformation and the lowest conformation is retained at each residue. The conformation used in the analysis are rotamers taken from the current rotamer library. We recommend that you use the *old* library available on the Options... palette since it is extensive and contains values for chi 3 and chi 4 variations.

If the sequence has been aligned, the sequence number for the built segment is taken from the sequential position of the Cα atoms in the sequence. If no sequence alignment has been set, the sequence is numbered beginning with 1.

Delete fitted seg

This tool deletes a currently selected fitted segment.

Undo last build

This selection will undo your last selection with the exception of **Guess Next CA** and **Save changes**.

Save changes

This selection writes the current Cα trace to a session file, saving all the fitting done up to that time. This session file will be read in the next time X-POWERFIT is used, and therefore can be used as a backup file if necessary. Note that a save to the session file is made automatically on exit from X-AUTOFIT.

X-POWERFIT

This application provides smart tools for Cα tracing based on secondary structure searching and assisted tracing tools. X-POWERFIT is based on the calculation and placement of vectors that represent the principal components of the secondary structural elements.

Find sec. struct.

This tool is designed to determine the secondary structure from an electron density map. Before the tool is used, the bones must be active, and the start value for the bones set to a level that indicates useful information. The default value for the bones "start" is indicated in the table below for experimentally phased maps, but the use of the tool **Bones | Map quality from bones** may indicate more suitable values.

- For 3fo-2fc maps: start = 1.6 to 1.8 σ.
- For 2fo-fc maps: start = 1.1 to 1.2 σ.
- For Sigma A weighted maps: start = 0.8 to 1.2 σ.

The tool can take several minutes to run. Generally, expect it to take approximately one minute for every 30 residues in the structure. Using highly over-connected maps (bones start value too small) significantly increases the calculation time! It should reliably find stretches of secondary structure, especially at lower resolutions (~3–4 Å).

This tool initializes the vectors, removing any previously calculated vectors. The tool is *not* available if a map or bones are not present.

Delete vectors

Allows you to delete vectors placed using the **Find sec. Struct**. tool or placed manually or read from an external file. On selection of the tool, the

application prompts you select a vector to delete. Pick the vector to be deleted. The tool continues to prompt for vectors to delete until there are none left, no vector was close to the screen position picked, or a palette tool was picked. Selecting the palette tool again aborts the command.

Place-edit strand

The **Place-edit strand** tool has three applications.

1. Reclassifying a helix as a strand.
2. Editing the position and length of a strand already present.
3. Adding a strand by picking two bones points.

When you select this tool the application prompts you to **Pick bones or vector** and the Accept or Quit dialog appears. While this Accept or Quit dialog is open, you can exit the tool only by picking **Quit** to reject any changes or **Accept** to accept the changes made.

If a helix vector (color 5) is picked, it is reclassified as a strand vector. If a strand or helix vector is selected, the dials change to allow the translation, rotation, and changing of the length of the vector.

If a bones point is selected, the application prompts you to pick a second bones point on the screen. The tool is aborted by any other pick. On picking a second bones point, a strand vector is placed so that the line joins the two bones points.

The new vector placed by this command is added to the list if **Accept** is clicked.

Place-edit helix

The **Place-edit helix** tool has three applications.

1. Reclassifying a strand as a helix.
2. Editing the position and length of a helix already present.
3. Adding a helix by picking two bones points.

On selecting this tool, the application prompts you to **Pick bones or vector** and then the Accept or Quit dialog appears. While this Accept or Quit dialog is open, you can exit the tool only by picking **Quit** to reject any changes or **Accept** to accept the changes made.

If a strand vector (color 5) is picked, it is reclassified as a helix vector. If a strand or helix vector is picked, the dials change to allow the translation, rotation, and changing of the length of the vector.

If a bones point is selected, the application prompts you to pick a second bones point on the screen. The tool is aborted by any other pick. On picking a second bones point a strand vector is placed so that the line joins the two bones points.

The new vector placed by this command is added to the list if **Accept** is clicked.

Save vectors

This tools allows you to save the current set of vectors to a file. The filename is secondary_structure.vec. This file is read using the tool **Read vectors**, described next. The vectors are stored in the following ASCII format, as 4x4 matrices.

```
Title card (format = a)
Number of vectors to follow (format = i6)
loop i = 1 ; number of vectors  {
     loop j = 1 ; 4 records {
              matrix[i] [j , k = 1 ; 4 fields] (format=4f12.5)
     endloop (j)
} endloop(i)
```

Read vectors

This tool allows you to read in a saved set of vectors from the file secondary_structure.vec. After reading the vectors, the display is changed to show the vectors read from the file.

CA Search/Browse

This tool allows you to take the current Cα trace and search a set of pdb files for matching geometry. You must specify the number of Cα atoms in the trace, the cutoff for the RMSD between your trace and a candidate pdb file, and the directory path for the set of pdb files to be searched.

Vector Search/Browse

The **Vector Search/Browse** tool opens a dialog to allow you to search the protein databank by structure motif. The search uses vectors from the auto search of the secondary structure in X-POWERFIT or vectors placed manually.

If you have done a previous search, you are shown a list of previous hits and given the option to show the current hit. The hits are sorted by a

weighted combination of rmsd and the number of motif equivalents in the scroll list.

You may start to browse the solutions before the run completes, since each time the dialog is opened, the current list of hits is read into the results box and sorted by quality of hit.

You can specify both the number elements to align and the allowed rmsd deviation over all the motif elements. The database filename can be changed if you want to search a private list. The PDB file path must be valid if you want to view the results.

$HYD_LIB/vector.base is the default list of vectors used during a search and contains all the proteins in the Protein Databank as of 1 November 1998.

Generating a new database for searching

You may recalculate the predefined vector list used in the alignment by using the motif searching program in generate mode. The program is found as $HYD_EXE/motif_align and should be run in the following way:

```
>  $HYD_EXE/motif_align -generate -list (user-file-
   list) -database (user-vector-file)
```

The **-generate** option causes a database file to be created.

The **-list** option specifies that the next argument is a file that contains a list of protein filenames. Normally this is the current Protein Databank, which should contain the explicit path and filename of each protein on a separate line.

The **-database** option specifies that the next argument is the filename of the new database vector file. This *cannot* be overwritten by the program.

Generating a new database can take several hours.

...Next solution

If valid solutions have been read in using the **Search | Browse DB** tool, it is possible to view the next solution using the **...Next solution** tool**.** This immediately reads in the next solution in the list and displays the results as a Cα trace.

...Previous solution

If valid solutions have been read in using the **Search | Browse DB** tool, it is possible to view the previous solution using the **...Previous solution** tool. This immediately reads in the previous solution in the list and displays the results as a Cα trace.

...Solution to CA

This tool converts the current solution displayed on the screen into an X-AUTOFIT Cα-trace segment(s). This Cα trace can then be modeled using the **CA-build** and **X-POWERFIT** tools.

Current res. and seg.

Use the **Current res. and seg.** tool to change the alpha carbon and currently selected segment. The current alpha carbon is yellow, the rest of the segment is red, and all other segments are blue. When you make this selection, the message line instructs you to select an alpha carbon. The nearest alpha carbon to your screen selection becomes the active carbon, and the segment containing this carbon becomes the active segment. If you pick a terminal alpha carbon, all selections on the palette and the dials remain active. When you pick an alpha carbon that is not a terminal alpha carbon, the dials change so that the xyz position of the alpha carbon can be changed. The Cα plot changes to indicate the conformation of the alpha carbon you selected.

Vector to CA-trace

The tool **Vector to CA-trace** converts the reduced representation vector into a Cα-trace and hence increases the amount of information determined in the building process. The tool prompts you to pick a vector from the display. If no vector is picked or a palette tool is picked, then the tool does nothing. The tool uses a directed refinement algorithm to fit a “standard” helix or strand into the electron density. Since each vector is classified as either strand or helix (color 5 = helix, color 14 = strand), the tool can automatically determine the secondary structure type to fit. The length of the vector defines the length of the secondary structure element to fit, and hence the number of residues that will be fitted to the electron density.

Electron density must be present in the region of refinement for this command to produce a correct, refined solution. This command is not available if no map is present.

Cycles of gradient refinement are carried out along the principal components of the strand/helix. When fitting helices, the tool also carries out a directed helical refinement, in which the Cα trace is rotated and translated simultaneously so as to find Cα positions along the helix. As the orthogonal components of refinement gradient are separated for at least part of the refinement process, the radius of convergence is very high, on the order of 3 Å.

The tool returns a residual value to indicate whether the Cα trace fitted really does fit the electron density. The fit value is returned to the textport and shows the following information.

- A residual less than 1 indicates a very likely fit.
- A residual between 1 and 2 is OK.
- A residual greater than 2 is not a likely fit and indicates that the Cα-trace fitted is not present in the electron density map.

The refinement is very sensitive to the quality of fit and diverges rapidly when an incorrect structure is fitted to the map. Hence the residual is a very reliable indicator of fit. Alpha helices in proteins tend have a very restricted range of conformations, and therefore helices are placed with a great deal of precision by this command. Strand structure in proteins is very variable, and therefore this tool has less success in fitting this type of structural element. When fitting strands, make sure that the refined coordinates actually fit the map. In some cases it is more useful to place strands with the **Auto extend CA** command.

On picking a vector to place, a new Cα trace appears. After a short period of refinement, the Cα trace jumps to a new position and the residual is displayed in the textport. The new section of Cα trace becomes the current segment, and any previous Cα is colored so as to reflect the changes. If the new Cα trace is the first Cα trace, then all the masking of the palettes is changed to reflect the presence of the Cα trace atom, and the Cα plot appears if it is not already displayed. The new Cα-trace atoms can be edited, appended, and deleted by any of the commands on the CA Build palette. If a sequence table is open, then the new Cα atoms will have residue type “UNK”.

Next CA as helix/strand

This tool adds another Cα atom to the current segment if the current Cα atom is a terminal atom and *only* if the previous 4 Cα atoms lie in either a helix or strand structure. The tool places the new Cα atom in a helical conformation if the previous four atoms are in a helix and in a strand conformation if the previous four atoms are in a strand. Hence the tool is used to extend the secondary structure elements placed by the tool **Vector -> CA trace** but can be used on any general segment of structure if it is a recog-

nized conformation (see the Cα plot). If the previous four atoms are not in a recognized conformation, then the tool stops and prints an error message to the textport.

Auto extend CA

The **Auto extend CA** tool is designed to add Cα atoms to the end of the current segment using continuous calls of the routine to add Cα atoms based on the bones and map. Hence this tool requires the presence of the map and bones before it can be used. The presence of vectors created with the **Find sec. struct.** tool provides the algorithm with additional information, which can improve the results (the vectors do not need to be displayed). The routine is very conservative and stops adding atoms if any of several conditions is not fulfilled.

The tool checks all the possible pathways for the next Cα, based on the bones, using two methods. For each pathway the Cα atom is refined into the density. The tool checks the Cα conformation for each pathway, whether the bones network is main or side chain in each direction, and for the presence of branch points at or near the test positions. The current secondary structure is also used to weight the pathways. The tool then checks the local and global build of the Cα trace by refining polyalanine into the Cα trace and checks the unrestrained geometries generated in the polyalanine chain. Since this building process is highly correlated, the local and global build are different. The routine stops if the build quality significantly falls in the local region or for the whole Cα trace, whenever a clash occurs, or if no pathway can be found in the map. If all conditions are satisfied, the new Cα atom is added and the process is repeated. The build can also backtrack to try alternative paths.

Cα building proceeds at about 2 residues/sec on an R4000 computer but does depend on the current size of the segment, since the global polyalanine test depends on the size of the segment. During building, feedback is provided on the quality of the build as a pie chart in the bottom left corner of the molecule view window. The filled portion of the pie chart represents the probability of the global fit of polyalanine in the best of the forwards and backwards directions. The pie chart remains displayed while building is occurring and is removed when the process ends.

Auto-building can be aborted by clicking the left mouse button at any time.

A large amount of output is written to the textport to allow analysis of the building progress, including the current quality of fit, whether the fit is improving or getting worse, and the reason for abortion of the building process. The coordinates of the automatically extended chain should be carefully reviewed.

CA refinement

The **CA refinement** tool allows a method of improving the fit of the Cα coordinates into a map based on refinement of the pseudobond lengths, angles, and torsions of the Cα trace. The refinement is highly restrained to retain reasonable Cα geometry. Bond length and bond angle restraints are included in the calculation. The tool also carries out rigid-body refinement of the segment. Significant improvements in the fit of the Cα trace to the electron density can occur.

The refinement is successful with large stretches of Cα trace, but due to the inherent correlation within a long segment of Cα atoms, the refinement can take tens of seconds to complete. Therefore some time can elapse between refinement cycles (and screen refreshes) as the algorithm removes this correlation. This tool is limited to 250 Cα atoms since correlation along the chain makes refinement very slow for large chains. Longer Cα chains must be cut with the **CA build | Unjoin 2 CA** tool.

Auto-trace Low

This tool performs Cα tracing for lower resolution or poorly phased data. It is necessary that secondary structure vectors exist (**Find sec.struct.**) before this tool can be used. The tool then checks the helical vectors to determine the best quality helix within the map. A Cα-helix is placed using this vector, and the Cα trace is automatically extended from both ends of the helix. The tracing continues until an exit condition occurs. Exit conditions include **No Path**, **Join point for CA trace**, **User defined abort** or **low quality fit**.

The tool repeats this operation with each helix vector in the sorted order of quality until no more helix vectors are left or the map is traced. If there is still a map to trace, the same sorted order placement and extension occurs for the strands.

The build rate is approximately 1-2 residue/second and the number of residues fitted depends on the resolution and quality of phases.

It should be noted that the tracing method is a pathway analysis algorithm. This means that the method cannot jump a gap in the bones trace. The vector search algorithm works for a much greater range of bones start values, and generally a sigma value of 1.2 or 1.3 should be used with 2Fo-Fc and SigmaA weighted maps. The tracing algorithm, as it requires connected bones, should still be used on bones slightly over connected, for example, about 1.1 sigma for the start value.

It should also be noted that density modification sharpens a map, i.e., good bits get better, and bad bits get worse. Hence, DM maps are often more

fragmented and with highly DM maps, a sigma level of 1.0 may be required if you want to trace the map in one go.

Auto-trace High

This tool performs Cα tracing for high resolution, well phased data. The tracing method is designed to determine a single solution from a simultaneous analysis of all the bones. This tends to be an all or nothing routine.

This tool uses an entirely different algorithm than previously available in QUANTA, although it is still a pathway analysis method. It is designed to be used on maps 2.0Å or better. There is no necessity for vector search, only that the bones be turned on and trimmed to the required molecular boundary. The routine is fast and generates a Cα trace that can be manipulated with all the tools in X-POWERFIT and X-AUTOFIT.

Turn on the bones, trim these and click this tool. Tracing occurs at about 50-100 residue/second.

The routine tends to fail to produce results rather than entirely wrong results. For data of the prerequisite quality it is sensible to use this tool first to check for reasonable results, and then use the **Auto-trace Low** tool if the data are not suitable. Of course the **Auto Extend CA** tool can be used on the results to continue tracing. It should be noted that at modest resolutions the routine can produce helices with not enough residues/turn. Be aware of this.

Consensus tracing

This tool carries out a multiple trace analysis of a protein to determine the optimal interpretation of a map. You must first calculate the secondary structure vectors using **Find sec. struct.** from the X-POWER-FIT palette before using this tool. The tool traces a map nine times and returns the best trace as the highest sum of atomic occurrence with variance 0.7Å.

> **Warning:** Do not abort this tool.

Knot finder

This tool checks the current active segment of the Cα trace for knots. It indicates in the text port whether a knot was found. Click the tool to draw a graphic of the analysis progress and final knot path; click the tool again to close the graphic. Use the tool: **X-Autofit | CA Build | Load CA Coordi-**

nates to import a Cα trace from an already built protein to check this. For example, the protein 1yve.pdb contains a knot.

The literature indicates that the process of creating folded protein from the DNA code means that any protein containing a deeply embedded knot is extremely unlikely. In the context of map interpretation, this tool is therefore used to identify bad connectivity.

Hide this menu

This closes the X-POWERFIT palette.

Sequence palette

This palette is only for use with the CA Build palette.

Load sequence

When you select this tool a dialog box opens to allow the selection of a sequence format and filename from which to read the sequence information. These formats are described in detail in the *QUANTA Protein Design User Guide*, Appendix F.

The Dialog box contains a file browser, a text field for a file name, and a multiple option list for different formats. The default format is PDB, and all the PDB files will be listed in the file browser.

1. **From PDB:** This option will read the sequence from the ATOM card information of a PDB file. The file browser will show all files with a .pdb extension.
2. **From MSF**: This option allows you to fill the sequence table from a MSF file. The file browser will show all the files with a .msf extension.
3. **Free format**: This is the original sequence reader and allows you to read the sequence from a file that has no particular formatting. The file should contain a single letter code of the sequence in either upper- or lower-case, with any number of sequence entries per line, up to 100. Spaces are read in as part of the sequence and can be used to delimit segment structure in the sequence because no alignment can be made across these spaces. Do *not* add spaces that are not segment delimiters. The file browser will show all files with the extensions .aa and .seq.
4. **FastA**: The first line of the file is the title and begins with a “>”, and the rest of this record is the title. The sequence is read until a “*” or “eof”

marker is encountered. Spaces and punctuation are *ignored*. The file browser shows all files with the extensions .aa and .seq.

5. **GCG**: The GCG file may contain an arbitrary number of lines of comment at the start of the file. These are followed by a blank line and then a title line. The sequence is written with 50 residues per line, each line beginning with the sequence number. These sequences are obtained from the GCG package. The file browser will show all the files with the extension .gcg.
6. **NBRF**: The first line contains a ">" in the 1st character position, and a ";" in the 4th character position, followed by the sequence ID. The second line is the title. Subsequent lines represent the sequence, terminated with a "*" character. Spaces are ignored. The file browser shows all the files with the extension .nbrf.
7. **EMBL | Swissprot**: The file begins with lines of comment which have two-letter keywords at the start of the line. The sequence is preceded by a line beginning with the keyword "SQ". The sequence is terminated by a line keyworded with "//". Spaces are ignored. The file browser shows all the files with the .embl extension.
8. **Quanta sequence builder**: The sequence builder option reads the sequence from the QUANTA sequence builder format file. The file browser will show all the files with the extension .qua.

If a sequence is successfully loaded into X-AUTOFIT, it will be displayed at the top of the main molecule window in lowercase letters. After loading the sequence table you can assign sequence information to Cα-trace atoms and any weighted fit will be displayed next to the sequence.

If a Cα trace is already present with sequence information (for example, after using **X-AUTOFIT | CA Build | Load coordinates**), then an alignment will automatically be carried out with the new sequence loaded. The alignment will be shown on the new sequence that is read in. If there are multiple identical domains in a protein (i.e., the ab+ab structure of the insulin dimer), then it is recommended that the sequence read in consist only of one dimer unit. Then use **X-AUTOFIT | Sequence | Unique sequence** to mask out already assigned sections during the sequence assignment.

Delete sequence

This tool removes the displayed sequence and the alignment arrows, but has no effect on the Cα trace assignment.

Show/Hide sequence

This tool allows you to toggle on/off the display of the sequence.

Current res/seg

Use this selection to change the alpha carbon and segment you currently have selected. The current alpha carbon is displayed in color 4 (default color yellow) and the rest of the segment is displayed in color 3 (default color red). All other segments are displayed in color 2 (default color blue). When you make this selection, the message line instructs you to select an alpha carbon. The nearest alpha carbon to your screen selection becomes the active carbon and the segment this carbon is in becomes the active segment.

If you pick a terminal alpha carbon, all selections on the palette and the dials remain active. If you pick an alpha carbon that is not a terminal alpha-carbon, then the dials change so that the xyz position of the alpha carbon can be changed. The Cα plot changes to indicate the conformation of the alpha carbon you have selected.

Unique sequence

This selection is grayed until X-AUTOFIT assigns a unique sequence to a segment. When this happens, the selection is unmasked and highlighted. The sequence is written in upper-case letters in the sequence table and in not used in any further sequence alignment. If you toggle the selection, the sequence is returned to lower case and can be used in subsequent sequence alignments for other segments.

Clear sequence

This tool returns all Cα assignments to unknown (UNK).

Return fuzzy sequence

Use this selection if you decide at any point that the fit is wrong after a unique alignment is made by X-AUTOFIT. When you make the selection, X-AUTOFIT returns to the last fuzzy sequence that you defined. This selection is grayed until X-AUTOFIT assigns a unique sequence to a segment.

Guess Sequence

The **Guess Sequence** tool provides a starting point for assigning sequences for the current Cα segment. The quality of the results is very sensitive to the quality of the electron density, so this tool should only be used when the electron density is reasonable and at least 50% of the protein is traced.

The tool searches for matching patterns of aromatic residues in the density and the sequence (the aromatic residues in the sequence are colored white). A protein with no aromatic residues thus cannot have a sequence assigned with this tool.

Guess sec. struct.

The **Guess sec. struct. tool** will predict the secondary structure from the loaded sequence data. The tool will color the sequence table on screen using red for helices and blue for strands. This tool can be used to assist/ verify the tracing of helices or strands in the structure.

Show/Hide Amino acids

This selection opens a palette listing the 20 common amino acids and the residue type “unknown”. This palette is used to assign a specific residue type to the current Cα, which is color 4 (yellow)]. Selecting any residue from this palette will cause a sequence alignment to be carried out against the loaded sequence (if one is present), and the results of the alignment shown against the sequence.

Show/Hide fuzzy

This selection opens a palette listing the 10 fuzzy descriptions of density. This palette is used to assign a fuzzy residue type to the current Cα which is color 4 (yellow). Selecting any description from this palette will cause a sequence alignment to be carried out against the loaded sequence (if one is present), and the results of the alignment shown against the sequence.

Undo last change

This tool will undo the last assignment of sequence information.

Hide this menu

Exits the Sequence palette.

Build atoms palette

Refine 1 residue

This carries out a real space refinement on the specified residue. This is a particularly good method of placing waters back into density. If you have specified a RTD file (see Defining torsion angles for unknown residues (page 95)), then this tool will actually do a torsion angle real space refinement of the picked residue; if the picked residue is an amino acid, the refinement will refine the side chain torsion and the N-Cα-C-O torsion. If a ribonucleotide is picked, the backbone torsions are refined and the single base torsion.

Restrictions

- A disordered residue is split up for gradient refinement and clamping is effective after editing if active.
- Any residue can be refined, but only known (.GSD) residues and RTD residues are torsion refined. Others are only rigid body refined.
- Incomplete/incorrect residues cannot be refined until built.

Geometric conformation

This tool provides a box of options from which you can select a Ponders & Richardson rotamer library conformation, or a trans conformation. (It is also possible to use the Sutcliff rotamer library or a new list of rotamers designed for the dead end elimination side chain fitting tool). The list of conformations is ordered by percentage of occurrence. You can also place as a trans conformation. For each conformation, the residue is drawn in white, with all the non-bond contacts from the residue. If you click on **accept**, you will get the new conformation, while **quit** will return the old conformation. This tool will also rebuild templated structures if incorrect or missing atoms are found.

This tool is not valid for DNA/RNA.

Restrictions

- A disordered residue is split up for geometric conformations and clamping is effective after editing if active.
- Only amino acids can be set to geometric conformations.
- Amino acids are rebuilt if they have incorrect atoms.
- Bumps are shown during the edit.

Fit side chain by RSR

This will fit a side chain to electron density by searching the chi angles, and allowing some tweaking of the angles in the side chain. Bonds are not refined, and impropers remain correct. This may result in a slightly distorted residue, especially if there is not enough density for the residue. If there is no density the result will be undefined. If you pick a Cα atom then all atoms in the side chain are fitted, even the CB position will be moved to give a correct chiral center and approximately correct chiral volume. If you pick the CG atom of a lysine, then only chi 3 and chi 4 will be refined, there will be *no* change to the N, Cα, C, O, CB, CG atoms. If you pick an atom in the side chain beyond all rotatable chi angles you will get an error and the tool will abort. The routine also aborts for glycine and alanine residues.

Proline is treated slightly differently. The residue is fitted by rotating the imino ring about the N-Cα bond, and checking that there are no clashes with the oxygen atom on the main chain. The single allowed ring pucker is also refined between ±30° from planar.

If the residue has missing atoms, or extra undefined atoms, the residue will be completely rebuilt from the template. This is the best way to build incomplete residues, even from just a Cα atom.

This tool will fit the single base torsion in DNA/RNA.

> **Note:** If using this tool seems to produce an obviously incorrect result, this is usually because the main chain is so far from the “correct” position that it prevents the side chain’s proper fitting to the density. In other words, the Cα chiral center defines the direction of the amino acid side chain. This is particularly common in regions of the map where the density is poor and the main chain trace is far from correct. To produce a better fit, select the tool **Build atoms | Move atom + RSR** and then move the Cα atom to the desired position. Often, as the Cα atom moves, the side chain flips into the correct conformation.

Restrictions

- A disordered residue is split up for side chain fitting and clamping, if active, is effective after editing.
- Only amino acids can be fitted by this tool.
- The residue is rebuilt if it has incorrect atoms.

Fit main chain by RSR

This prompts you to select a peptide bond. The tool will refine the torsion Cα[i]−Cα[i+1] if you pick the C[i]−N[i+1] peptide bond. The omega angle is also refined to values in the range 168° to -168°. This option is not available for residues other than amino acids.

Restrictions

- A disordered residue is split up for side chain fitting and clamping, if active, is effective after editing.
- Only amino acids and nucleic acids can be refined by this tool.
- The residue is rebuilt if it has incorrect atoms.

Move atom + RSR

This option continuously refines all the side chain chi angles that follow a side chain atom that you are moving. For example, if you select this tool and move the Cα atom of a phenylalanine residue, the side chain will be continuously refined: the chi 1 and chi 2 angles change and the bond angles in the side chain are tweaked in response to the movements of the Cα atom. The moving side chain is drawn in white. The side chain will tend to stick in density until is suddenly fits into a new place. An **Accept or Quit** pop-up box prompts you to either accept or reject the changes you have made. You cannot pick any other palette option while in this mode of refinement. The same restrictions apply as with the Fit side chain by RSR (page 163) tool for the last fixed atom, which is the one you are moving. Proline can also be fitted with this tool.

If you wish to move the N-Cα-C=O backbone group using this tool, then pick either the N, C, or O atom of the residue, and the backbone atoms will move as a rigid body. Pick the Cα atom to refine the whole side chain while retaining the position of the main chain N,C, and O atoms.

This tool will fit the single base torsion in DNA/RNA as the C1´ atom is moved.

Restrictions

- A disordered residue is split up for side chain fitting and clamping, if active, is effective after editing.
- Only amino acids and nucleic acids can be refined by this tool.
- Amino acids are rebuilt if they have incorrect atoms.
- Bumps, if active, are shown during the edit.

Edit backbone tor

For polypeptides...

This tool allows the peptide plane and omega to be changed with the dials. The peptide plane is rotated about the Ca[i] Cα[i+1] pseudo bond. Omega cannot be rotated by large values without creating bad geometry. Therefore, do not use this tool to create cis peptide bonds; instead, set (**X-AUTOFIT | Options | Omega restraint**) and regularize (**X-AUTOFIT | Build atoms | Regularize**) the cis restraints.

> **Warning:** The resulting positions of the side chains for residues [i] and [i+1] are not checked after this edit. It is possible to produce impossible chirality, etc., with this operation. You should correct the CB chirality by selecting **Fit main chain by RSR** and picking the Cα atom of both residues, or regularize the resulting residues.

For DNA/RNA...

This tools allows you to edit a set of eight torsion angles along the nucleic acid chain. Two (deoxy)-ribonucleic acid monomer units are edited at a time with this tool, with a result that a total of eight torsion angles can be changed during one edit. These torsion angles are β–γ–δ–ε–ζ–α–β–γ for the two residues concerned. If the residues i and i+1 are edited, then the phospho-ester bond i-1/i remains connected and residue i+1 is always the moving residue with the results that the phospho-ester bond i+1/i+2 becomes detached. You must complete the edit with closure of the i+1/i+2 phospho-ester bond. It is possible to refine the length of this bond subsequently using the regularize tool.

The tool requires that a bond be picked to indicate the two residues to edit, this bond is always part of the i residue to edit, hence picking the bond P(i)-O5´(i) apparently edits the wrong set of bonds as this particular bond is not itself editable.

The torsion δ is part of a five membered ring of the sugar, and hence determines in part the type of pucker of the sugar ring. This torsion is therefore highly restricted in the values allowed and large changes may produce inconsistent ring geometry. Changing the value of δ does not move the base.

The values of the torsions edited are shown in the DNA plot window as moving circular markers, where residue “i” are green and the “i+1” residue torsions are blue. There are two moving symbols for the torsion angles β and γ. All torsions are labelled on the model window using the two central atoms of the edited torsions.

Restrictions

- A disordered residue is split up for side chain fitting and clamping, if active, is effective after editing.
- Only amino acids and nucleic acids can be edited.
- Residues are rebuilt if they have incorrect atoms.
- Bumps, if active, are shown during the edit.

Edit chi angles

This tool allows you to manually edit the side chain chi angles. You are prompted for a residue to edit. The current chi angles are written on the message line as you change them. The tool will abort if there are no chi angles in the residue, or if the backbone is incomplete for this residue. The dials are changed during this option; use the dials to change chi torsions. If you have an RTD file, a ligand can be edited or up to 100 rotatable bonds can be edited. If there are more than eight rotatable bonds then the top dial becomes a toggle for the current active dial set. The changed atoms in the side chain are drawn in white. The **Accept or Quit** pop-up box prompts you to either accept or reject your changes. You cannot use any other palette items when editing the side chain chi angles.

Restrictions

- A disordered residue is split up for side chain fitting and clamping, if active, is effective after editing.
- Amino acids, nucleic acids and residues with torsion defined in the RTD file can be edited.
- Bumps, if active, are shown during the edit.

Move atom

This allows you to move a single atom in x/y/z. The tool will prompt for an atom to move, which you pick from the main window. The dials are changed for this option to allow the movement of the atom. The atom is marked by a white cross. You are prompted to **Accept pick** box for this edit, and no other palette item can be picked during this edit.

If the bumps are turned on you will get all the bumps to all atoms.

Restrictions

- Any atom can be moved.
- Bumps, if active, are shown.

Move zone

This allows you to move a zone of residues in x/y/z and about x=0, y=0, and z=0. The tool prompts you to pick two residues, the first residue in the zone and the last residue in the zone. If the same atom is picked (and hence only one residue), then the atom picked becomes the center of rotation. If different atoms are picked (either in the same residue when editing a single residue, or in different residues), the initial center of rotation is the center of mass of the atoms picked. At any time during the move zone edit you can pick an atom and this becomes the center of rotation during subsequent editing of the zone. Note that if the atom picked as the center of rotation is translated, then the original position of the atom remains as the center of rotation, and not the translated coordinate.

The dials are changed for this option to allow the movement of the residues. The residues are shown in white during the edit. You get the **Accept pick** box. No other palette item can be picked during this edit. If bumps are turned on, you will get all bumps from the moving zone to the rest of the structure. Do not try to move a zone through the middle of the protein with bumps turned on as everything will grind to a halt!

Restrictions

- Disordered residues are kept together if clamping is turned on, otherwise the disordered residues move depending on which residue (A/B) is picked as the first residue in the zone. If the first residue in the zone has no alternate conformation, all the A conformers in the disordered residues are moved.
- Bumps, if active, are shown.

- Incomplete/incorrect residues can be edited if clamping is on, and but cannot be edited if clamping is off.

Model terminal res.

For polypeptides...

This tool prompts you to select a terminal residue. You can use the dials to edit the last three phi and psi angles. The current phi psi values are shown on the Ramachandran plot. The moving molecule is shown in white, and bumps are displayed, if active.

For DNA/RNA...

This tool only edits the first two and last two residues within DNA/RNA due to the large number of torsions in the nucleic acid chain. The order of the torsions depends on whether a 5´ or 3´ end picked to be edited, as the moving residue is “i” for a 5´ edit and “i+1” for a 3´ edit.

When editing a 5´ end the changeable torsions are γ–β–α–ζ–ε–δ–γ, and β where applicable, for the residues “i+1” and “i” respectively where the residue “i” is the terminal residue. When editing a 3´ end the torsions β–γ–δ–ε–ζ–α–β–γ are editable for the residues “i” and “i+1” where the residue “i+1” is the terminal residue.

The values of the torsions edited are shown in the DNA plot window as moving circular markers, where residue “i” are drawn as green circles and the “i+1” residue torsions are drawn as blue circles. There are two moving symbols for the torsion angles β and γ. All torsions are labelled on the model window using the 2 central atoms of the edited torsion.

Flip torsion 180 degrees

The **Flip torsion 180 degrees** tool allows the flipping of recognized rotatable bonds in proteins, DNA/RNA and ligands. The routine will prompt for a bond from the currently displayed and active molecules. If a peptide bond is selected from a protein molecule then the peptide plane (N-C=0) is rotated 180 degrees about the Cα-Cα pseudo bond. This is commonly known as a pep-flip. If an amino acid or nucleic acid side chain bond is picked, and if this bond is rotatable, the atoms along the side chain will be rotated 180 degrees about the bond picked. This is useful when the validation procedure indicates that an HNQ error is present, and therefore the last chi angle of the respective residue should be rotated 180 degrees. It is also possible to rotate torsions within a ligand if this torsion is rotatable (as determined by the algorithm for calculating these), or the torsion is defined

in the “lig.rot” file. A rotatable bond can be added or removed from a ligand using the tool **X-AUTOFIT: X-BUILD | Build atom | add delete | define torsion by bond**.

It is not possible to flip the phi, psi or omega torsion in proteins, or any backbone torsion in nucleic acids. This type of flip would result in connectivity problems within the polymer chain.

Mutate residue

For polypeptides...

You can mutate any residue to any other, except for itself. The new residue is added regardless of any missing atoms in the residue to be changed. The tool prompts for a residue to change (pick any atom in the residue), then a palette containing the 20 amino acids appears, from which you pick the new residue. The new residue atoms are added so that the value of a chi angle that is equivalent in the residue before and after mutation is maintained. Any non-equivalent chi angle is set to the most likely rotamer.

For DNA/RNA...

The tool allows the mutation of a residue to one of the five allowed bases: cytosine, adenine, guanine, thymine, and uracil. The base torsion is set to the same value as found in the residue before mutations. No restriction is made on the Thymine/uracil pair for the DNA and RNA models.

Restrictions

- Only amino acids and nucleic acids can be mutated.
- Disordered residues are replaced with no disorder.
- Incomplete/incorrect residues are replaced complete.

Add/delete...

Opens the Add/delete palette described in the next section.

Hydrogen bonds

This tool calls the QUANTA hydrogen bonding routine. The hydrogen bonds between the real, symmetry and NCS atoms are also calculated and added to the symmetry atom object. This is so that the hydrogen bonds are hidden when the symmetry atoms are hidden using the object management

table. The tool does not calculate the hydrogen bonds between symmetry atoms and symmetry atoms. This is to reduce the amount of visual information on the molecular display so that you can concentrate on the interactions only between the real atoms and symmetry.

The tool acts as a toggle button. When toggled on, the hydrogen bonds are recalculated whenever there is a change in the coordinates. This can slow down the display of the screen after any edit. When this tool is toggled off, no hydrogen bonds are calculated or displayed. To reduce the time of calculation of the hydrogen bonds after editing a residue, you are recommended to reduce the **Coordinate display radius** on the Options... palette. A value of 20 Å is suggested.

Move atom + reg. res.

This allows a residue to be dragged around using one atom in the side chain. When you pick this tool, you will be prompted for an atom. Pick an atom in the molecule. This residue is regularized to completion, after which you can move the picked atom. The dials change to allow the movement of the atoms in x/y/z directions. The moved residue is drawn as a new set of coordinates in colors ranging through green-yellow-orange-red, depending on the error in the geometry.

The **Accept pick** box appears so you can either accept or reject the changes. No other palette is available while this tool is active. A new atom in the moving residue can be selected at any time. Picking this atom will result in this atom being the moving atom.

Restrictions...

- Residues (for example, amino acids) in the .GSD file are regularized to template coordinates. Residues not in the .GSD file cannot be used with this tool. (Use **X-AUTOFIT | Build atoms | Move atom + reg. zone** for these.)
- A disordered residue is split up for side chain fitting and clamping, if active, is effective after editing.
- Disulfide bridge detection is **not** active for this option.
- Fixed atoms are in effect (see page 173).
- The optional parameter of dipeptide restraint is not applicable.
- The optional parameter of phi/psi restraint is not applicable.
- The optional parameter of non-bond interactions can be used.

Move atom + reg. zone

This tool prompts for two residues that define the ends of the zone to be regularized, and then an atom to be the first moving atom. After making these three selections, the zone will be regularized to completion, which may involve waiting one or two seconds. The dials are changed so that the moving atom you picked can be moved in x/y/z directions. The third atom selected can then be moved and all the atoms are regularized about this moved atom. The moving atoms are drawn as a new set of coordinates in colors ranging through green-yellow-orange-red, depending on the error in the geometry. At any stage a new atom can be picked and the previous moving atom is freed. All fixed atoms previously defined remain fixed *unless* picked as a moving atom. Note that the first residue N atom is fixed, and the last residue C atom is fixed unless either is a terminal residue. You can move the atoms at any speed, but give the program a chance to catch up occasionally. This can be determined when the moving atoms become green again.

The changes can be accepted or rejected using the Accept or Quit dialog box. All other palette options are ignored while this tool is active. You can pick any atom in the moving zone at any time.

Restrictions...

- Residues (for example, amino acids) in the.GSD file are regularized to template coordinates. Parameters of all other residue types are determined using type definitions and a PSF table of parameters. This can result in a rather slow start to the regularization process for non-templated residues.
- A disordered residue is split up for side chain fitting and clamping, if active, is effective after editing.
- Disulfide bridge detection is active for this option.
- Fixed atoms are in effect (see page 173).
- Dipeptide restraints apply as prescribed on the dialog box **X-AUTOFIT | Options.**
- Phi/psi restraints apply as prescribed on the dialog box **X-AUTOFIT | Options.**
- Non-bond restraints apply if active (**X-AUTOFIT | Options**), but are *not* recommended with this tool unless using a very powerful machine.

Regularize

For polypeptides...

This allows you to regularize a zone of atoms to completion. The tool will prompt you to pick two residues that define the zone to be regularized, after which the atoms are regularized and updated. You will see the progress of the regularization as a set of coordinates that range in color through green-yellow-orange-red, depending on the quality of the geometry. The N atom of the first residue and the C atom of the last residue in the zone will be fixed. If the first or last residue in the range is a terminal residue, then it is left free to move.

For DNA/RNA (for all regularization tools)...

Note that there are no explicit chiral centers set up for the sugar ring, which means that all forms of sugar pucker can be regularized, and that final coordinates will be that of the initial pucker. If you wish to change the pucker of the ring, use either:

- the **Move atom** tool before regularization, then use the **Refine 1 residue** tool, which may redefine the pucker based on the electron density (if the density gradient is large enough),

  or

- the **Move atom + reg res** tool to quickly pull a single atom in the ring, which redefines the pucker. Then regularization occurs for the new ring definition.

Fixed atoms and non-bonding are supported for nucleic acids, while the options of **dipeptide restraint**, **phi/psi restraint** and **disulfide restraints** are not applicable and are ignored.

Restrictions

- Residues (for example, amino acids) in the.GSD file are regularized to template coordinates. Parameters of all other residue types are determined using type definitions and a PSF table of parameters. This can result in a rather slow start to the regularization process for non-templated residues.
- A disordered residue is split up for side chain fitting and clamping, if active, is effective after editing.
- Disulfide bridge detection is active for this option.
- Fixed atoms are in effect (see below).

- Dipeptide restraints apply as prescribed on the dialog box **X-AUTOFIT | Options.**
- Phi/psi restraints apply as prescribed on the dialog box **X-AUTOFIT | Options.**
- Non-bond restraints apply if active (**X-AUTOFIT | Options**).

...fix atoms

This will open a dialog offering the following four options:

Fix all CA atoms

Fixes all the Cα atoms in a protein.

Fix all CA, N, C, O atoms

Fixes all the main chain atoms in a protein

Fix atoms by picking

Allows you to add to the list of fixed atoms by picking atoms from the molecule. If, however, a fixed atom (marked by a white cross) is picked, this atom is *removed* from the fixed atom list. For example, it is possible to fix all the Cα atoms and then re-edit this list to free some of the Cα atoms.

Clear fix atoms

Deletes the list of fixed atoms.

Color atoms...

This displays a palette that allows you to color atoms by property. The Color atoms palette is described starting on page 179.

Edit atom info

This prompts for a residue to edit, and then displays a palette containing information about that residue. The sequence ID and insertion code can be changed for the residue, while the occupancy and temperature factor for each atom of the residue can be changed in the dialog box. Two fields allow you to set *all* atomic values in the residue to the same value. If the toggle is set to **TRUE**, then on exit the application will set all atomic B-values and occupancies to the values specified in the fields next to the toggle, regard-

less of the atomic values. If the toggle is set to **FALSE,** then the atomic values of the two parameters are taken from the values in the atom list.

If you pick **Cancel**, any changes entered are ignored; if you pick **OK**, all the changes you entered are updated in the data. If you pick a residue with more than 15 atoms (such as a ligand), then there will also be **Next** and **Prev** boxes to allow you to scroll up and down the list of atoms.

Show residue bumps

This prompts you for a residue. All the nonbond contacts to this residue (not including 1-2, 1-3, and 1-4 bonding) are displayed as lines with distance markers. The markers do not go away until:

1. You pick another residue to show bumps.
2. You turn off the bumps from the Options palette
3. You move a zone, chi angles, terminal, atom, or atom + RSR.

Save changes

This writes all changed coordinates to disk.

Undo last fit

This changes all edits back to what they were at the last save. You can check the edit status of the atom using **Color by progress** (see page 179).

Hide this menu

Puts away the Build atoms palette.

Add-delete palette

The Add/delete palette allows the addition and deletion of residue data from the data model. It also allows the reorganization of the peptide chain by applying the **Create bond** tool when a peptide or phospho ester bone is selected.

Delete bond

This breaks the bond selected. If a peptide bond or phospho-ester bond is picked, then the residues become (temporarily) termini, which you can edit using the **X-AUTOFIT | Build atoms | Model first last 4 res** tool. You can

also insert residues using the **X-AUTOFIT | build atoms | Add delete | Add res. to termini** tool, or change the connectivity.

If a disulfide bond is picked, it is deleted.

If any other bond is picked, it is deleted temporarily.

Create bond

This creates a bond between atoms. The tool prompts for two atoms to be picked which are then joined by a bond. If the bond is a peptide bond, you can use the **Regularize** option to join distant residues. You can only join an N-terminus to a C-terminus (or vice versa).

If a disulfide bond or any other bond is created, they are temporary bonds. There is no distance check for these newly created bonds.

Note: If you join two residues that are not consecutive, the order of the residues is changed, but the sequence number of each residue is *not* changed. Because this change moves residues within the data structure, you may need to use the option **X-AUTOFIT | Build atoms | Add delete | Renumber sequence ID** to allow the regeneration of the correct sequence numbers after you have used this tool.

The **Create peptide link** tool along with **Delete peptide link** allow you to change loop connectivity by cutting peptide bonds and then creating new ones (which may be very distant). You can then use the **Regularize** option to drag the residues to a new position.

Add torsion by bond

This tool allows the definition or removal of a torsion angle in a residue that is *not* an amino acid or nucleic acid. The residue can then be rotated manually (**Build atoms | Edit chi angles**, **Flip torsion 180 degrees**), and also refined within any the real space torsion angle refinement algorithms. (**Build atoms | Refine 1 residue**, **Structure | refine zone**)

The tool first prompts you to pick an atom, which identifies the residue that should gain/lose a torsion definition. (If you pick an amino acid or nucleic acid, the routine aborts.) After the residue is specified, the tool then labels the residue with any currently defined torsions. These defined torsions are read from the file lig.rot. If this file is not present, it will automatically be generated using the rotatable bond analysis algorithm. If a bond is selected that is not currently defined as rotatable (and is a valid torsion: non-ring,

non-terminating), then the application will add this to the list of rotatable bonds. If the bond selected is currently defined as rotatable, then this torsion is removed from the list.

The tool remains active until the tool bar is picked again.

On completion of the editing of the rotatable bonds in a residue, the file lig.rot is generated if not already present, or modified if already present. The definitions in this file will be used in any subsequent calculations. Note that the definitions for other residues are retained in the lig.rot file.

Add alternative conformation

This tool allows you to assign a residue up to four alternate conformations. The second conformation is fitted to the second most distinct density pattern in the residue side chain, while the third and fourth conformations are slightly different from a trans conformation (to keep them distinct for addition).

Occupancies are reset so that the sum of occupancies for each atom is one, and B-values are reset to 20. The main chain and CB coordinates are identical — that is, they have no disorder if the clamp flag is set to true, and the main chain atoms will have slightly different positions if the clamp flag is false. For more about handling disorder, see the notes on clamping (B-conf clamped to A backbone (page 205)).

Delete residue

This prompts for a residue to delete, and then removes that residue. If you select a disordered residue, only one of the disorder pair will be deleted, depending on the atom selected.

Delete range

This prompts for the first and last residues, which define the range to be deleted. The residues within this range will be deleted from the data structure. You cannot delete an entire molecule with this tool.

Add res. at termini

For polypeptides...

The program prompts you to pick a terminal residue. If you do not pick a terminal residue, this tool aborts. The program will then display a pop-up

palette of amino acids. When you pick a residue from this list, it is added to the terminus by real space refinement. The algorithm used to fit the terminal residue will go back to the previous residue and fit the phi and psi angles from this residue, and also the phi/psi (depending on whether an N or C terminal was added) of the new residue. Therefore you must have electron density contoured about the old terminal residue. Then, the side chain of the new residue is fitted. The addition of residues by this method is very accurate for well-defined density, there is normally no limit to the number of residues that can be added in a chain. You should not use this tool to carry out de novo density fitting; the Cα-trace method is far more powerful in cases where density is ambiguous.

For DNA/RNA...

The **Add Residue** tool will allow the addition of a nucleic acid to the terminus of a DNA/RNA structure. The tool will first prompt for the selection of a terminus to which to add a residue, and if a nucleic acid terminal residue is picked, a palette appears so you can select one of adenine, cytosine, guanine, thymine, or uracil. If this palette is not picked, the tool aborts. The residue picked from this list is then added to the terminus of the polynucleotide in the conformation defined on the **Options** dialog box (beta/alpha/Z1/Z2). No restriction is made to the addition of uracil and thymine to DNA/RNA. If DNA is being extended, a deoxyribose sugar is added, otherwise a ribose sugar is added. The residue is not fitted to the experimental data, but set to a predefined conformation set in the Options dialog.

Add water at pointer

This places a water at the current pointer position. Then you can refine with the rigid body refine routine. If there are already waters in your data, the new water is placed at the end of the current list of waters in the coordinates, and is given the same segment ID and a residue number 1 greater than the previous water in the list. If you have no waters in your data, then a new segment, W, is created, and the new water is given the residue number 1.

Add atom at pointer

This opens a dialog box with a list of cations and anions, from which you can select one to place at the current pointer position. Once added, the new atom(s) can be refined with the tool **X-AUTOFIT | Build atoms | Refine 1 residue.** If there are other metals/ligands in your coordinates, each new ligand/metal is added at the end of the coordinate list and given the same segment name and a residue number 1 greater than the previous metal/

ligand in the list. If this is the first ligand in your coordinates, the new one is added to the end of your data, and a new segment is created with the name Z.

Re-patch terminal

For polypeptides...

X-AUTOFIT supports various standards for C-terminal oxygen atom naming. This tool allows you to change the C-terminal COO group from/to any of the following conventions: O only, O/OXT, OCT1/OCT2, OT1/OT2, O/OE, O/OT. On picking the tool, the application will prompt for a terminal residue. If an N-terminal residue is picked, then, for polar hydrogen and all-hydrogen modes, three hydrogen atoms will be added. For a C-terminal residue, a dialog box will open that allows selection of one of the terminating types.

For DNA/RNA...

The tool to repatch a terminal in DNA will change the 3´ end of DNA/RNA so that an H3T will be added to all hydrogen or polar hydrogen DNA/RNA structures. If a 5´ end of a DNA/RNA fragment is selected to repatch, a dialog box with these options opens:

1. **No phosphate**: If present, the phosphate group is removed from the 5´ end of the polynucleotide chain.
2. **Phosphate - O5T**: If not already present, a phosphate group is added to the selected 5´ end, and *no* terminating oxygen (or hydrogen where relevant) is added. This will leave an unsatisfied phosphate group.
3. **Phosphate + O5T**: If not already present, a phosphate group is added to the selected 5´ end, and an O5T oxygen is added to the phosphate to complete the terminating phosphate. If hydrogen modes of **All Hydrogen** or **Polar Hydrogen** are used, an H5PT atom is also added.

Show B or U

This tool is used to show B-values or U-matrix values for an atom, a residue, or for all the displayed atoms. The display consists of three orthogonal circles/ellipses that define a one sigma deviation about the centre of an atom for isotropic/anisotropic motion. For the anisotropic display the principle axes of the ellipse are along the principle directions of motion of the atom as defined by the U-matrix.

A dialog box gives you the choice of B-values or U-matrix display, and whether the display shows an atom, a residue or all displayed atoms.

The B-values are taken from the atom B-values stored in the MSF file while U-matrix values are *only* read from a PDB file of the same name as that shown in the molecular table MSF file. If there is no PDB file of the same name as the edited molecule, then no display is added.

A scale value scales the spheres and ellipses drawn. The default value is 1.

Renumber sequence ID

This allows the simple consecutive renumbering of sequence IDs for each segment after insertions and deletions. You can interactively specify a range of residues to renumber or specify the starting sequence ID.

Hide this menu

Puts away the Add/delete palette.

Color atoms... palette

Color by atom

Colors the atoms by element type. The default scheme for view coordinates is:

- Carbon = Color 1 (pale green).
- Nitrogen = Color 2 (blue).
- Oxygen = Color 3 (red).
- Sulphur = Color 4 (yellow).
- Hydrogen = Color 5 (white).
- Everything else = Color 6 (cyan).

Color by progress

- Blue = residues not edited.
- Yellow = residues edited but *not* saved.
- Red = residues edited and saved.

Color by (ρ) fit

Colors the atoms by fit to the map. The color range is a scale of:

- Green-Yellow-Orange-Red/Blue.

  The green atoms fit well and the red atoms fit badly. The blue atoms are those outside the currently loaded map extent.

Color by B-value

Colors the atoms by the value stored in the B-values, where an absolute scale of $Å^2$ is used. For example:

- Green = 0–15.
- Green/yellow = 15–30.
- Yellow = 30–45.
- Orange = 45–60.
- Red = 60–75.
- Blue > 75.

Color by occupancy

Colors the atoms by the occupancy value, where the occupancy data is normalized to lie between 0 and 1, and the following colors are used for 4 different relative ranges:

- Green = 0.75–1.00.
- Yellow = 0.50–0.75.
- Red = 0.25–0.50.
- Blue = 0.00–0.25.

Color B-alt different

Colors the B conformer of an alternative conformation pair a single color (Color 14, pink).

Hide this menu

Puts away the **Color atoms** palette.

Structure palette

Fragment fitting

On picking the tool for fragment fitting, all currently open X-AUTOFIT's/ X-BUILD palettes will be hidden and the Fragment fitting palette appears. Note that the molecule that is edited in the fragment fitting application is the first visible and active molecule.

The Fragment fitting palette is described in Creating a Fragment Database (page 331).

On completion of the search for fragments, the previously open X-AUTOFIT | X-BUILD palettes are re-displayed. If any changes have been made to the structure using the fragment fitting application, then the molecule changed will be marked as edited, and you will need to save it when you exit from the X-AUTOFIT | X-BUILD application. Changes made in the fragment fitting can be undone using **Undo last fit** on either the Build atoms palette or Structure palette.

Rigid body fit

This allows the rigid body refinement (in real space) of entire sections of structure. If a complete domain is selected then the whole domain will be refined. The radius of convergence is not as large as in reciprocal space, but it may be able to improve some fitted regions. The algorithm uses gradient refinement on the six degrees of freedom defined by the rotations and translations of the zone to be refined. Occupancy information is used as in all refinement techniques. The refinement resolution factor defined on the Options palette is described on page 211. The tool prompts for the first and last residues in the refinement zone to be refined and, on completion of the refinement, asks whether to accept changes or quit.

Restrictions

- Atoms present in the zone are refined; missing atoms and disorder are not explicitly treated.

Refine zone

The refinement algorithm uses torsion angle and rigid body refinement on a residue basis for each residue in the refinement zone, with cycles of regularization including non-bonding. The radius of convergence of this type of refinement has been found to be very high (on the order of 1.5/2.0 Å), and allows rapid refinement of regions of the protein. Although the algorithm can be used for refinement of an entire protein, you should not do the final refinement of proteins with this protocol.

The refinement is carried out in real space to the current map. The map must cover the entire region of refinement, since atoms with no experimental data contribution will move only as a function of geometrical restraints to other atoms. The algorithm will move atoms about torsion angles. The geometric restraints always include bonds, angles, impropers, and non-bond overlaps. If the phi psi restraints are active then these are used also. The non-bonds will only be defined for atoms displayed, so if there are symmetry overlaps, then the symmetry atoms must be visible for these contacts to be used in the refinement. (See the options palette, and the tool to change the symmetry radius). The refinement factor on the Options palette can be used where the density is poor due to lack of high resolution data, and this protocol has been shown to be useful on data of up to 4.0 Å. Occupancy information is used as in all refinement techniques. The refinement resolution factor defined on the Options palette is described on page 211. The tool prompts for the first and last residues in the refinement zone to be refined and, on completion of the refinement, asks whether to accept changes or quit.

Restrictions

- Only the A conformer of disordered residues is refined.
- Residues not in the .GSD file are rigid body refined only.
- Incomplete residues will abort the refinement.
- Nonbonds *always* apply regardless of the **X-AUTOFIT | Options** dialog parameter for nonbonds.
- Dipeptide restraints apply as prescribed on the dialog box **X-AUTOFIT | Options.**
- Phi/psi restraints apply as prescribed on the dialog box **X-AUTOFIT | Options**.
- Disulfide bond detection is active.
- Fixed atoms are effective.

Loop fit & Terminal fit

The loop fit and terminal fit algorithms both use a Monte Carlo protocol to determine loop conformations that match electron density. The tool prompts for a start and end point for the search; all the residues in the selected zone must be covered by the electron density. The density for all atoms not part of the selected zone is masked to act as a precalculated non-bond list. The algorithm then generates random conformations defined by the phi and psi angles in the selected zone, and, if running a loop fit, checks that the end-to-end distance of this conformation will approximately fit in this region, and then carries out a density fit.

The conformations are generated at a rate of approximately 2000 per second, and the best ten fits are retained. Initially the search is slow, as the display of the first few new hits will be relatively slow compared to the search speed. Because of this, these algorithms accelerate over the first few seconds. This is particularly apparent with the terminal fitting. The ten best fits are displayed during the search and colored by fit to density. Green is a good fit; red is a bad fit. The calculation continues until you interrupt it by clicking in any QUANTA window with the left mouse button.

After interrupting the search, if there are some solutions to the search, the tools on the Fitting palette will become active, and all other X-AUTOFIT tools will be inactive. The solutions found to this point are sorted by fit to density, so the best solution found so far (by density fitting), becomes the active loop.

Fitting/Continue search

The continue search tool allows the search to be resumed, with the solutions found up to this point still active.

Fitting | Show next conformation Draws the next best conformation in the list of solutions, as defined by fit to density. If the current conformation is the last conformation in the list, then show next conformation shows the first conformation.

Fitting | Show previous conformation Draws the previous best conformation in the list of solutions, as defined by fit to density. If the current conformation is the first one in the list, then show previous conformation will show the last conformation in the list.

Fitting | Show all conformations Displays all the conformations found so far. This allows a check to see if there is a cluster of possible solutions.

Fitting | Accept If the **Accept** tool is selected, the current active loop becomes the actual coordinates for the main chain atoms, and the side chain atoms are added by real space refinement from the Cα positions found by the search.

Fitting | Quit aborts the loop fit option and does nothing.

The loop search will always fit a poly-alanine backbone trace to the electron density regardless of the sequence, but, for glycine, the CB atom will have zero occupancy.

The loop searching will only allow searching for complete residues, so if the loop has not been built, build a random conformation to complete the loop (see **XFIT | Build atoms | add-delete | add-residue-at-termini** on page 176) and then do loop fitting.

Do all…

This tool allows a refinement protocol to be used on a range of residues. This tool is most useful for automatically refining the position of all the waters in a macromolecular structure, while X-AUTOFIT checks the changes made to each water by refinement. If any changes occur that are outside user-defined limits, the **Do all** tool stops and centers on this residue so that this water can be edited.

The **Do all** tool opens the Do all… dialog box.

Do all | Change all This allows the selection of the residue type or zone of residues to be refined. If **Change all Water** is selected, X-AUTOFIT checks all MSF files currently open; if any residue is a water molecule, then this is refined. If the **Protein** option is selected, the range of residues is defined by the amino acids in the first displayed and active molecule. The **Start from Sequence ID Segment ID** option is set at the current residue by any of the following:

- **X-AUTOFIT | Pointer | Place at next residue.**
- **X-AUTOFIT | Pointer | Place using coord.**
- Picking a point on the Ramachandran plot.
- If the last **Do all** tool aborted because changes were too large.

The **Select zone by atom name** tool allows a specific range of residues to be refined. Two pop-up boxes appear. The first asks for the segment name, atom name, and sequence ID for the first residue in the range. The second pop-up prompts for the last atom specification in the range. If any one of the segment name, atom name, or sequence ID is not found in the MSF, then **Do all** aborts.

Therefore, your first use of **Do all** would normally be to apply one of its first two options. When this fails because the residue has become badly built by refinement, the current residue will be set to the next residue after this. The next use of **Do all** requires only that the **Start from Sequence ID Segment ID** option be selected. Note that if a badly refined residue (for example, if a water had no density) is deleted, then the residue that follows the deleted residue becomes the current residue.

Do all | By Protocol… The **By protocol** parameter allows the use of:

- **Refine**—a single residue/atom is refined by gradient refinement.
- **Fit side chain** — an amino acid side chain is fitted to density.
- **Fit main chain** — an amino acid peptide bond is fitted.
- **Find Alt.Conf.** — Search all residues for alternate conformations as defined by fit to electron density. Each amino acid (not GLY/ALA), is searched for a significant second side chain conformation. Variations in main chain conformations are not found with this tool. The pucker of the CG atom in the proline residue is also searched for alternate conformations.
- **Regularize** - Each residue is geometrically refined so that the bond, angles and other geometrical terms are optimized.

Some combinations of **Change all** and **By protocol** are not valid (for example, sidechain fitting of waters). For these conditions, **Do all** aborts.

The abort conditions are:

- **Stop if max deviation > 1.00**
  The **Do all** tool will abort if any atom in a residue (or water) moves more than the distance specified.
- **Stop if bump energy > 0.00**
  The Do all tool will abort if any non-bond contact distance with any other atom of an MSF is less than the specified distance from any atom in the residue (or water) after refinement.
- **Stop if sigma < 0.50**
  The **Do all** tool will abort if any atom of the residue (or water) moves to electron density where the sigma level is lower than the sigma level specified.
- **Min accept RMSD**:
  You can specify a lower limit for how far the refinement has moved - for rigid body refinement, side chain fitting and main chain fitting. The default is set to 0.01 Å. Specifying a larger number (such as 0.1 Å) will mean that if the refinement results in changes that are less than 0.1 Å RMSD, the changes are not kept.

- **Alt. angle diff.**
  This defines how different the conformation has to be for a B alternate conformation to be generated for a residue. When an alternate conformation is generated, the sum of differences in the chi angles between the A and B conformations is computed. If any of the angles exceeds the minimum angle defined with this command, then the alternate conformation is added. (Only valid for finding alternate conformations).
- **Alt. fit fract**
  This is the minimum density value required for a B alternate conformation is to be added. For example, a value of 0.9 specifies that, for the B conformation to be added, the electron density value for the B conformation must be more than 90% of that for the A conformation. (Only valid for alternate conformation finding).
- The **Do all** tool will also abort if the amount of swap space available to the program is less than 10 Mbytes. Repeated display of different areas of the map (for large maps where many residues/waters are present), will result in this abort condition.

  You can also abort the **Do all** tool by clicking the left mouse button in any QUANTA window.

  For each residue of the required zone, X-AUTOFIT will move the display to the residue (or water) that is to be refined, refine this residue (or water) by the protocol defined, and then check the abort conditions. If no abort condition is registered, X-AUTOFIT will move to the next residue and repeat the operation. If an abort condition occurs, then the **Do all** tool will write a warning to the text port and return control to you.

Refine all water

The **Refine all water** tool automatically places all the water molecules into available electron density for the first visible and active molecule. The tool carries out ten cycles of unrestrained real-space refinement of all the water molecules, with nonbond energy refinement to all other atoms that are visible and to the symmetry-generated atoms. This tool supersedes the **Do all... | water fit** since it simultaneously fits all the waters, checks water error correlated to other water positions, and is much quicker.

The tool returns 3D text notes on each water molecule where a problem is encountered. Several errors can occur:

Clash distance (non-bond clash)— Any water molecule under refinement will stop shifting along a gradient if a nonbond clash occurs with any other atom in a visible molecule or its symmetry-generated atoms. The water will be labelled with a text string to indicate that a nonbond clash has occurred and hence refinement has not converged. No account

is taken of partial occupancy waters since water pairs (or higher orders) with 0.5 occupancy often converge to the same point. The tool will therefore warn of nonbond clashes even for partial occupancy waters so that you can handle these water molecules separately.

Move distance (Shifts too large) — Any water molecule that moves more than 0.2 Å during the final refinement cycle will be marked as having shifts too large. This generally occurs for water molecules that begin the cycles of refinement a long way from any minimum position. It highlights water molecules that may not have been fully refined.

Low density — A water molecule is marked if it remains in poor density.

Ref. steps — Any water molecule that has not converged during cycle 10 because the number of internal steps is exceeded will have this error label. This occurs when the gradient of the map is very small or complex, and hence convergence is not complete

Repeatedly using this tool will progressively remove the **shifts too large** and **refinement steps**errors without any change to the converged waters or the waters with Non-bond clash. The user should check errors on the unconverged atoms.

The tool will indicate the number of waters to refine, and, if none are found, a warning is printed. The most likely occurrence of this warning is that the first active and visible molecule contains no waters. Non-bond interactions are calculated to all active molecules regardless of their visibility.

A new tool **…Delete bad water** is available. This tool can only be used after the **Refine all water** tool and acts on the error messages generated by that tool. To delete less waters, use less stringent restraints in **Refine all water**.

…Delete bad water

This tool can ONLY be used after **Refine all water** and acts on the error messages generated by this tool. The **Delete bad water** tool deletes all waters with error messages generated by the water refinement. To delete fewer waters, use less stringent restraints in **Refine all water**.

Fit by D.E.E.

It is possible to fit the side chain coordinates of a residue using modeling techniques where the experimental information is very poor or non-existent. The **Fit by D.E.E.** tool will prompt for start and finish residues and carry out the placement of the side chain atoms by adjusting the conforma-

tion of the side chain chi angles of multiple neighboring residues. Note that the neighboring residues are also changed while calculating the zone of residues.

The energy is computed for each possible conformation of a residues and its closest neighbors and the lowest energy conformation is retained for each residue during the search process.

Undo last fit

The **Undo last fit** tool re-reads the session file from the last save of changes. This has exactly the same action as **XFIT | Build atoms | Undo last**, as they both act on the same set of tools and changes. Any changes made since the last save will be undone and the previous coordinates and residues restored. This can be used when a modeling session has gone wrong, and all changes need to be deleted. You can check the changes that will be restored using the **XFIT | Build atoms_color by progress** tool.

Regularise range

This tool regularises a range of residues based on a single residue selection picked by the user. The default for the residue range is +/- 1 residue. The default values can be changed with the tool **Regularise param**. This tool is suitable for use when **active residue** is turned on.

Regularise volume

This tool regularises a volume of residues based on a single atom selection picked by the user. The default for the volume is 6 Å. The default values can be changed with the tool **Regularise param**. This tool is suitable for use when **active residue** is turned on.

Regularise param

This tool can be used to change the values for the residue range and volume used by the **Regularise range** and **Regularise volume** commands.

Refine volume

This tool carries out a real space torsion angle refinement for all residues (including ligands, but not water) within a volume about a selected atom. The tool requests a single atom pick, or uses the last picked atom when

using active residue mode. The radius of the refined volume is set using the tool **Regularise parameter** on the same palette, and the default is 6Å.

> **Note:** All sections of molecular structure that are covalently bound to non-refine regions are automatically restrained with fixed atoms.

Auto build

This tool automatically runs through the entire fitted residues and fits them using a mixed grid and gradient protocol. It take three residues, fits the side chain of the center residue by Grid refinement. Refine Zone is then used on the main chain atoms only of all three residues, followed by side chain fitting of the two edge residues by Grid refinement. This is repeated for five cycles. The final stage is a round of regularization. This tool can work through gaps/missing regions of the structure.

Change all B/O

This tool provides a simple interface to change the value of the temperature factor or occupancy of the first current active and visible molecule. The tool allows the setting of all atoms--just the main chain atoms or just the side chain atoms--to a user defined value.

> **Note:** The temperature factor is bounded to lie between 1 and 99, and the occupancy is bounded to lie between 0 and 1

Mutate all

This tool allows you to automatically redefine the sequence of the first current visible and active molecule. There are four options:

- **To ALA**: changes all the residues to alanines.
- **To GLY**: changes all the residues to glycines.
- **To Seq**. **tab**.: allow the sequence to be changed to that of a loaded sequence table (read in by **X-AUTOFIT X-POEWRFIT| Sequence | Load sequence**). The tool mutates each residue and tries to maintain the current conformation by retaining the current side chain torsion angles where valid.
- **To Seq. tab. & RS**: same as to **Seq. tab.** But also carries out a post-mutate real space torsion angle refinement.

Hide this menu

Puts away the Structure palette.

Tables and Graphs

This palette is used for the advanced analysis of proteins (in particular) and other macromolecules. The tools available in this palette are used to generate tables containing information on atoms, residues, and undefined data. The data tables can be operated on by a number of functions, and plotted in a number of styles. The graphs and tables can be picked to centre the molecular view, while the graphs can be annotated and plotted to a postscript format file. Hence it is possible to use the table and graph information to zoom in on possible problems within a crystal structure solution, and use this for advanced validation.

For more on table usage, see the *X-AUTOFIT:X-BUILD Tools (page 121)* chapter.

Read file

The **Read file** tool opens a directory browser dialog with two sets of options.

1. The Text handle option controls how the text information read from the file is placed/not-placed into the new General table.

- **Ignore all text**: will read numerical data into the table ignoring all text strings. This includes all the titles and any textual information on each data line.
- **Ignore lines with text only**: will ignore any line of data that contains no numerical information. This would normally include titles (as long as they contain no numerical data delimited by spaces). Any data lines containing numerical data will be read and the text information added to the respective data columns. The aim of this is to ignore title lines while reading text cards into the table.
- **Include all text in table**: will read all text information into the table with the numeric data.

2. The data placement option controls where the data will be put into an already existing General table.

- **Create | replace table**: will place the information into a new table regardless of whether a general data table exists or not. All previous data in this table is lost.
- **Append columns**: will append the data read in as additional columns of data. The length of the columns need not be the same length as that originally in the table
- **Append rows**: will append the data read in as additional rows of data. The length of the rows need not be the same length as that originally in the table.

Read molecule

This tool will read the currently displayed and active molecules into the atom table generating the columns: atom name / residue name / Sequence number/ x / y / z / occupancy / B-value.

Atom and residue ID columns are also generated but hidden. This is used for table and plot picking. The current contents of the Atom table is always overwritten by this action, and any reference by picking graphs or the residue table may be incorrect if the atom table is updated with a different molecule.

Del. Current col.

This will delete the currently selected columns of data. If no columns are selected, nothing is done.

Del. Current row

This will delete the currently selected rows of data. If no rows of data are selected, nothing is done.

If rows of data are deleted from the residue table or atom table, data inconsistency can occur. This is a minor problem since, because the remaining rows of data will be correctly marked, there will be no referencing errors between the tables and molecule view.

Del. Table

This will open a dialog box to allow the deletion of a table. This is necessary when the displayed and active molecules have been changed, or when

1 or more residues have been deleted from a molecule. In both cases, a serious data inconsistency can occur.

Atom Selection

This tool opens a dialog box to allow data selection in any of the tools for generating the tables: Read molecule, Column function, Protein Property, and Differences. The toggle boxes allow you to select protein, nucleic acids, water and ligands where the ligand option is defined as any residue type that is not an amino acid, nucleotide, or water. By default, data is generated only for amino acids and nucleic acids.

All currently open tables are *deleted* if the atom selection is changed since inconsistent data would be produced.

Column function

This tool will open a dialog box to select a function to apply to the current selected column(s). If no tables are open then the tool will abort. If no columns of data are selected then the dialog box contains a table and column selection boxes. All data in the column selected is operated on regardless of the activity of the molecules.

The dialog consists of three scrolling lists of functions:

- **Basic functions**: This provides a list of simple functions that all, except probability, return data to the same table as a new column of data. The probability function returns data to the general data table along with a “bin” column to indicate the bin centre value for plotting. Data in the input column that is not a number (NoData, SegErr or Zero 0) is returned as a “No data” entry in the output column. Most functions are self explanatory, but see the list for more information.

```
Sin, Cos, Tan, ArcSin, ArcCos, ArcTan, SinH, Cosh, TanH :Trig functions, probably not very useful.
Square, Root : Takes the square and root of the data respectively
Abs, -Abs : Takes the absolute value of the data, + negative of.
Negate : Calculates “zero - datum”
Ln, Log : Natural log and Log(10) of data
Exp : E^datum
Prob : Determines distribution of data by finite size bins.
Copy : Makes a data copy.
NoDat->0: Changes all occurences of NoData to a zero.
0->NoDat : Changes all occurences of zero to a “NoData” flag.
Zero-Tor, Pos.-Tor, Neg.Tor : These functions apply to torsion data which are calculated to lie
between -180 to  180 degrees.  The zero function changes the data to -180 to 180, the Pos. function
changes data to the range 0 to 360 degrees and the Neg. function changes the data to lie between -
360 and 0 degrees.
```

- **Residue functions**: This list of functions can only be applied to the atom table information and is used to take the mean / median / min / max / SD / skew / curtosis of a residue or main chain or side chain. If the res-

idue table is not open, this is filled with the basic residue information plus the new values. Data in the atom table input column that is not numeric is returned as a “No data” entry in the residue table.

- **Column value**: This returns a value to the text port as the application of the function to the column, and does not create a table entry.

Calculations

This opens a dialog that allows the four basic operators of (+/-/x//) to be applied between two table columns. Both columns of data must be in the same table. The result is placed into the same table as the original data.

Protein property

This allows the generation of data as derived from the molecule. Most of the functions apply to amino acids, though some can be applied to any other type of residue. Two scrolling list of functions are opened from which a function can be chosen. The last function used will already be entered into the text box.

Property values

This allows the calculation of derived values from the current visible and active molecules.

```
Mean bond length to each atom  : atom information
Mean bond angle to each atom  : atom information
User bond : atom, residue, or Undefined data produced
User angle : atom, residue, or Undefined data produced
User torsion : atom, residue, or Undefined data produced
Omega : residue information
Phi       : residue information
Psi        : residue information
Chi 1/2/3/4 : residue information
CA-CA distance  : residue information
CA-CA-CA angle : residue information
CA-CA-CA-CA torsion : residue information
```

Property errors

These functions return the difference between a value calculated from the molecule, and a standard value.

```
Bond Error to each atom : atom information
Angle Error to each atom : atom information
User bond error : atom, residue, or Undefined data produced
User angle error : atom, residue, or Undefined data produced
Omega error : residue information
Rama.Energy : Residue information - energy from a pre-computed energy surface for Ramachandran
angles of (ala-ala-ala)
Rama.OUB.Dist : Residue information - Distance from the hard sphere 90% surface if not-allowed, or
zero if inside the 90% hard sphere surface.
Non-bond clash : Atom information - Sum over all non-bonds to each atom of the distance to close
compared to allowed value. Zero if OK.
Chi 1/2/3/4 Error : Residue information - Torsion angle difference from value determined and the
nearest rotamer library value for that residue
CA-CA Error : Residue information
CA Improper Error : Residue information
Side Chain Improper error : Residue information - sum over all of chiral and prochiral errors in
side chain. No-data if none.
Nomenclature Error : residue information - Torsion value if not valid, Zero if valid, No-data if no
data
HNQ Hbond Error : residue information - Numerical difference in Hbonds between alternate conforma-
tions if wrong, zero if OK, No Data if not valid residue type.
HNQ Bvalue Error : residue information - Numerical difference in B-value if incorrect, zero if OK,
NoData if residue type is wrong.
```

User defined properties in the property list.

The user properties are found in the list of properties and property errors. These allow the user to return data that is bond, angle and torsion based and defined by an atom name/sequence offset list. This is input from a dialog box that will appear. For example consider the calculation of the chi 1 torsion angles as shown in the following dialog box.

The atoms offsets are 0,0,0,0 as all atoms come from the same residue. The atom names are n, ca, ?b*, ?g*. The ? specifier means that any single character can occur at this position. The * specifier means that any number of characters can occur at the end of the string.

For the calculation of the omega torsion, the atom offsets are 0,0,1,1 and the atom names are: ca c n ca.

The two option lists **Average over** and **For...** determine how the data is averaged for placement into the tables.

Average over

This defines whether the data is to be returned as atom/residue or as is data.

- **None**: The data will be placed as calculated into the general table and will not link back to the molecular data.
- **Atom**: The data returned is atom based, and for torsion data the 2nd atom defines the link back to the atom data. Only those rows of the atom table with the link will have the data value, all other values are set to “NoData”. For the chi 1example, only the Cα atoms in the atom table will have data returned, all other atoms in each residue will have a NoData value. You should generally use the atom data averaging for “any” bond and angle data (c* *), and residue averaging for all torsions, and where atom names are exclusively specified.

- **Resi**: The data is returned as residue based, and for the torsion data the second atom defines the residue and is placed in the residue table. For this chi 1 example, all residues except glycine and alanine will have data values. For the omega example, the last residue in a segment will have no data.

For...

This specifies how the data is placed into the data table where there is more than one value per data table row.

- **One**: Only the last value for a multiple results for residue/atom is used.
- **Each**: Average all data for the multiple results for a residue/atom

For example, consider the user bond [c* *], i.e. all bonds to carbon atoms. This will return multiple values to each carbon in the structure. If **one** is selected, the last value found for an atom/residue is taken as the returned data, and placed into the respective table. This is OK where the data is unique for an atom/residue. The **each** option will average all the values found to an atom/residue, and put the mean value into the table. The latter option is more sensible for the [c* *] example.

Differences

This tool opens a dialog box that allows the determination of differences between two molecules or two segments in the same molecule (for example NCS with restraints).

The first active and visible molecule is used as the reference molecule for all calculations. The Match molecule is selected from the top scrolling list of molecule names. If the same molecule name is selected as the reference molecule, segment comparisons are made and the **Seg. align** input strings are relevant. If the molecule selected is different from the reference molecule, differences are calculated between the molecules, and the **Seg. align** input boxes are ignored.

The XYZ option list is relevant for XYZ differences in position calculations only. The options allow deviation plots between the molecules/segments without a LSQ alignment, with LSQ alignment based on all matched atoms, and LSQ alignment based on Cα atoms.

The Difference type scroll box allows the selection of the type of difference calculation:

- XYZ: “standard” deviation in atom position of matched atoms.
- Bvalue: difference between Bvalues of match atoms.

- Chi 1/2/3/4: difference between the respective chi angles for those residues with the particular chi angle under analysis.
- Omega: Differences in the omega angle.
- Phi / Psi: Differences in the phi and psi angles
- Ramachandran: Differences between the Ramachandran angles as defined as the SQRT((phi[a]-phi[b])^2 + (psi[a]-psi[b])^2)

The **Seg.align** entry fields allows the definition of the segment names for the alignment of segments where the match molecule is the same as the reference molecule.

XYZ and B value data are returned to the atom table and the torsion values are returned to the residue table. The table entries filled are defined as the matched atoms in the reference molecule, for the segment deviations, only the first segment information is filled.

Plot data

Opens a dialog box to allow multiple graphs lines to be generated from a set of table columns. If the table columns are selected, then the dialog will be filled in with the details of these columns; this is by far the easiest method of setting the plotting dialog. If no data columns are selected 3 default plot information dialogs are generated, the second and third have a zero Y data value to indicate that nothing is to be plotted.

Only one X ordinate is allowed; the table entry for this is defined by the first selected column, or the atom table if there is no selected column data. The default X data column is “0” to indicate that a numerical count is to be used for the X data.

For each graph to plot:

- The table name is used to specify the table data to use for the Ydata, if the columns have been selected then this field is filled from this information, otherwise the default table is the atom table if present.
- The Y data column indicates the column to be plotted as the Y ordinate. If columns have been selected then this field will be filled correctly, otherwise the default columns are the last column for Ydata 1 and zero for Ydata 2 and 3.
- The Graph color indicates the color to use for the graph.
- The width is the line width used for the graph line
- Line/Sym/Hist/Dash options indicate the style of the graph to be used. If multiple styles are wanted (symbol and line), then copy the data using

the “copy” Column function, and plot both of these in the two different styles.

- **X-Legend**: A string of characters to be placed for the X-axis. (Note that the graph can be annotated in general).
- **Y-Legend**: A string of characters to be placed for the Y-axis. (Note that the graph can be annotated in general).
- **Axis style**: Color and line width for the axes.
- **Axis font**: Font size in “point” used for the axis annotation.

The graph will be plotted if OK is picked -- nothing is done if Cancel is picked. Note that only one graph window is produced, and previous graphs are overwritten.

Postscript settings...

This opens a dialog box to allow the postscript file output to be changed. This effects the plot generated directly from the table and graphs plots, and also for the current session any plot generated from any Graph plot (not molecular view).

- **Settings**: Allows the use of an absolute font size, as defined when the plot was generated (unchecked box), or a scaled font as a function of the size of the output image (checked box).
- **Color**: Allows the generation of Black and white postscript, Color postscript where black and white is inverted (so the background is white), and Color postscript where all the colors are inverted.
- **Picture manipulation**:

  Step [0.5] - step size for any manipulation (inches)
  X/Y [=] - make the size of plot so X = Y
  X/Y [-] - Decrease the X size of the plot
  X/Y [+] - Increase the Y size of the plot
  [U] - Move plot up
  [>] - Make plot smaller
  [L] - Move plot left
  [R] - Move plot right
  [<] - Make plot bigger
  [D] - Move plot down

- **Set**: Make the changes to the postscript setting and save
- **Plot**: Plot the current Graph generated in X-BUILD
- **Cancel**: cancel

Label graph...

This brings up a dialog to allow the annotation of an X-BUILD graph plot.

- **Text / Box / Line / Circle**: Annotation type
- **Label [ ]**: String if type = text
- **Size(pnt)**: String point size if type = text
- **Orientation**: String orientation
- **Color**: Color of all annotation types
- **Style**: Line style if type <> text
- **Attachment**: Figure / Graph - For general use, annotation should be attached to the figure which allows the annotation to be placed anywhere on the plot, but only approximately at a data point. If the annotation is directly associated with a data point then the annotation should be placed as a "graph" type attachment; but it cannot be outside the graph plot area.
- **Add**: Click this to add the current annotation.

> **Note:** For a text, a single pick of the plot is required, and for lines, boxes, arrows, and circles, two mouse picks are required.

- **Delete**: will prompt to delete of an annotation on the graph regardless of any other setting.
- **Cancel**: Do nothing.

Hide this menu

Puts away this palette.

User Defined tools

This palette is designed to allow you to create a custom palette of tools using any of the tools from the rest of X-BUILD, X-AUTOFIT and X-POWERFIT. It is also possible to set up macro tools that carry out more than on operation when picked.

Define new tools

When active, this allows the user to select any tool off the palettes of X-AUTOFIT/X-POWERFIT/X-BUILD (licensing dependent), and these tools are added to this palette.

When picked, the user is prompted to pick tools from any of the palettes. None of the tools (except those that open other palettes) perform their normal action, but rather place the tool picked as the next tool on the User defined palette. If a tool is picked for a second time, or picked on the User defined palette then this tool is deleted from the user defined palette. This may also used to delete a macro from the User defined palette.

When tool picking is active the **...add menu space** and **...Clear all** tools become un-masked.

The define new tools mode is exited by picking **Define new tools** again.

Define-Edit group

This tool is used for generating macros. Macros are tools that carry out more than one function. When active, and the user picks a tool, the program first prompts for a macro group name (via a pop up dialog), and then the tool picked is added to the macro group. The user continues to add tools and finishes the macro group by picking the **Define-edit group** tool again. The macro group folds up and is shown as a <>group-name

For example - pick a residue, fit side chain to density and regularize.

Pick > Define/Edit group

Pick > Place using coord: The application prompts for a macro-group name [Pick atom & fit/reg]

<> Pick atom & fit/reg : appears on the user defined palette

-> Place using coord: appears also

Pick > Active residue on

-> Active residue on: appears on palette

Pick > Fit side chain by RSR

-> Fit side chain by RSR: appears on palette

Pick > Regularize

-> Regularize: appears on the palette

Pick > Active residue off

-> Active residue off: appears on the palette

Pick > Define/Edit group

The palette folds to show only the tool <> ***Pick atom & fit/reg***

If this tool is subsequently picked then all the operations in the macro are carried out sequentially. In this case the program prompts for an atom pick, fits the side chain to density, then regularizes the residue.

Do not provide a blank macro name (no string).

Do not provide a macro name already used.

In both cases the first occurrence of the name will be the tool used.

...Add menu space

Adds a menu space after the current tool entry during define new tool entry

...Clear all tools

Clears all the tools from the palette. If immediately pressed again, the clear is un-done.

X-AUTOFIT:X-BUILD main palette tools

Validate

The Protein validate tool provides some simple validation procedures to check that the current active visible molecule conforms to the requirements of the protein data bank, as well as warning of deviations from expectation.

If a toggle box is checked, the validation indicated will be carried out on the first current active and visible molecule. The validation will generate text marks at each error residue which can be automatically fixed with the 3Dtext tool **fix validate error.**

It is possible to validate a structure at three levels of stringency. There is a options button at the bottom of the validate dialog to choose the required preference for deviation allowance.

- **Bond errors**. This checks for deviations in bond errors, and marks deviations at the centre of the bond that deviates.

- **Angle errors**. This checks for deviations in angle error, and marks deviation at the middle atom of the bond angle in error.
- **Peptide deviation**. This checks the peptide bond Cα-C-N-Cα and marks any torsions that deviate more than 12 degrees from either planar trans or cis conformation.
- **Ramachandran errors**. This checks the Ramachandran angles phi and psi and detects any phi-psi pair value that lies more then 20 degrees from a hard sphere allowed region.
- **Chi 1 Rotamer Error**. This checks for chi 1 rotamer values that deviate more than the 30 degrees from any of the rotamer values in the current rotamer database (see Options dialog)
- **Chi 2 rotamer Error**. This check for chi 2 rotamer values that deviate more than the 30 degrees from any of the rotamer values in the current rotamer database (see Options dialog)
- **CA-CA distance error**. This checks for deviations in the Cα-Cα pseudo bond. Although this is auto fixed by regularization, this type of error is indicative of some other problem.
- **CA-Chiral error**. This checks for deviation of the chiral volume of a Cα atom.
- **Side chain chiral error**. This checks for deviation in the chiral volume of a side chain chiral and pro-chiral atom set.
- **Non-bond clash**. This calculated the non-bond energy of the molecule as a function of atoms (including symmetry), and returns non-bond overlaps that have high energy. The non-bond error is marked at the centre point between the two offending atoms. If the non-bond error is to a symmetry atom, then the non-bond error is labelled at the real atom position.
- **HNQ H-bond check**. This calculates whether a better hydrogen bond network can be formed if His, Asn, and Gln residue side chains are reversed. A full combination is analyzed to check for correlated errors.
- **Nomenclature error**. This checks for the non-standard torsion definition for chi 2 of asp, phe, tyr; the chi3 of glu and the chi5 of arg.
- **Voids**. This carries out a search for geometrical voids in the protein and marks an error at the centre of each void found. It up to the user to decide if there is electron density at this site, or whether this void is present for some other reason. The auto fix is to add a water molecule.

At the bottom of the dialog is a check box to determine whether a log file of errors is written. The errors are collated as a function of each residue, and a list of each residue is made, and all errors associated with a residue are

listed for each residue entry. Hence an error can occur in more than one residue -- for instance, for a non-bond error.

Find Negative density

This tool can be used to find all the places in a map where atoms of the first displayed and active molecule overlap with negative density in the electron density map. Hence this tool is a validation method. You define which map to use with the tool **Set bones | RSR map**.The usual map to use would be difference density.

On selecting this tool a dialog box appears that prompts you for a threshold value to use in searching the electron density.

The threshold value is defined in map sigma, and the default value is -3.0. If you select **Cancel**, no action is taken. If you select **OK**, the application will read in the entire map and search for all peaks in the map that have a minimum below the threshold value. The peaks are clustered to remove errors closer than 1.0Å apart. The application will search all the current open MSF files for the closest coordinate to each peak and create a text label based on the peak depth and the nearest atom to this peak:

```
Neg P -> B Leu 11  CD1   :  -3.14
```

The text labels can then be used in the 3D text editor **X-AUTOFIT: X-Ligand | Text...**. One use for this validation tool would be in the placement of water molecules into electron density. After placement of water, some of the water sites can develop negative difference density after refinement if the site is of low occupancy, or if it actually is not a water site. These problem waters can easily be found with the validation tool and then deleted.

Set bones/RSR map

This tool displays a dialog box of the currently open maps and allows you to choose the map for the bones and RSR calculations. If there is only one map, this tool does nothing. If you select **Quit**, the map selection does not change. If you select **OK**, the new map is used for subsequent bones and RSR refinement calculations. The newly calculated bones will appear immediately.

Options...

This option displays the X-AUTOFIT Options dialog box, which is described on page 204, after the **Finish** option of the X-AUTOFIT:X-BUILD palette.

Color table

The Color table is described on page 211.

Save built CA to MSF

This tool allows you to save the Cα trace built in X-AUTOFIT/CA Build to a MSF file. It does not save information about the X-AUTOFIT all-atom model. This is not normally required, as all Cα trace information is saved in a session file. All residue names are saved as assigned in **X-AUTOFIT | Sequence**, so if Cα atoms have fuzzy descriptions, these will be saved as described. These residue names will not be recognized by most other graphics programs. To load the coordinate information from the MSF file created by this tool, use the tool **X-AUTOFIT | CA Build | Load CA coordinates**.

Save built atoms to MSF

This tool allows an explicit save of the model built in X-AUTOFIT | Build atoms and X-AUTOFIT | Structure, and any coordinates generated by the tool **X-AUTOFIT | CA Build | Fit seg by RSR**. This is explicit save of data is not normally necessary as X-AUTOFIT will record any changes to a structure in the molecular management table, and force a save of the coordinates on exiting X-AUTOFIT.

The tool will bring up a dialog box that allows the generation of a new version of the file, save to a new filename, overwrite the file, or abort the save (see Finish, below). The molecule that is saved by this tool is the first active displayed molecule in the molecule management table.

Save built atoms to PDB

The **Save built atoms to PDB** tool will save *all* the currently active and displayed molecules in the molecular management table to a single PDB file. The tool can be used to merge data into a PDB file for analysis and editing. If there are no active and displayed molecules, then only the symmetry cards are written to the file.

Run external program

This provides a script-based interface to run external programs. The documentation on this is found later in this chapter (see *External program palette (page 217)*).

Last commands...

The last command documentation is found at the end of this chapter (see *Last commands (page 214)*).

Finish

The **Finish** tool will exit X-AUTOFIT and return you to main QUANTA and the modelling palette. On exit from X-AUTOFIT, the application will ask if you want to save any coordinates that have been changed in any way. If multiple MSF files have been changed while in X-AUTOFIT, this dialog box will appear for each MSF changed while in X-AUTOFIT. All molecules not edited are not saved on finishing. A dialog box is opened that allows various saving options:

Choose the MSF Saving Option

The fourth option, to abort, has the same effect as **Cancel**, and returns you to X-AUTOFIT. The first three options, followed by **OK**, will result in the requested action on the MSF, and then a query about the next changed MSF. X-AUTOFIT exits after the last MSF save action.

If you do not wish to exit X-AUTOFIT, use **Cancel**.

X-AUTOFIT Options dialog box

The parameters in this dialog box control the general behavior of various tools in X-AUTOFIT.

Rotamer library

This allows the choice of the **Ponder & Richards** library of geometries, the use of the **Sutcliff** library of geometries or the use of the Oldfield library of rotamers. This option effects the **XAUTOFIT | Build atoms | Geomet-**

ric conformation tool and the tools for placing side chains by dead end elimination (**X-AUTOFIT | CA-build | Fit seg. by D.E.E** and **X-AUTOFIT | Structure | Fit by D.E.E**). If the Ponder & Richards or Oldfield library is active, then the tool **X-AUTOFIT | Build atoms | Geometric conformation** will create a pop-up palette that contains each of the rotamers allowed for the selected residues, and sort these in order of likelihood. If the Sutcliff library is active, then only a list of the allowed rotamers is shown on the palette.

Protein build mode

This allows you to choose the build mode for protein residues at a terminus ((**X-AUTOFIT:X-BUILD | Build atoms | Add-delete | Add res at termini**). Pfit is used to fit the terminal residue by fit to electron density, while the other choices define a geometric conformation.

DNA build mode

This allows the choice of the build conformation when using the add residue at terminal for nucleic acids.(**X-AUTOFIT:X-BUILD | Build atoms | Add-delete | Add res at termini**). The initial conformation of the added residue is defined by the set of backbone torsion angles that define the option type.

Return main dials after edit

This toggle changes the return set of dials after any edit in the X-AUTOFIT:X-BUILD tools. The default is to return the pointer dials after using a X-BUILD tool, for example: **edit chi angles**. If this toggle is set then the main dials are returned after an edit that changes the dial set.

B-conf clamped to A backbone

This option affects the treatment of disordered residues. When it is turned on, it results in all conformations being superimposed onto the main chain atoms of the edited conformation on completion of editing by a tool. This means that when a residue A conformation is edited (for example, regularized), then on completion of the edit, the B conformation moves to the equivalent edited position, relative to the main chain atoms of the A chain. You should use clamping of the B conformer to the A conformer for all single residue disorder, and only turn off this option when a disordered loop is being added. The aim of this tool is to retain the main chain atoms as single

atoms in a disordered residue, while allowing editing of the residue side chain of both A and B conformers.

Center at Ramachandran point for residues...

This tool changes the effect of picking the Ramachandran plot window. When the toggle is on, you can pick a Ramachandran point in the plot window and the residue information is written to the Textport, and the center of the display will move to this residue. If this option is not set, then only the information is printed in the Textport and no centering of the screen is made.

Ramachandran range

The Ramachandran range option allows you to specify the number of Ramachandran points to be displayed in the Ramachandran plot window. This option is also defined for nucleic acids and the data values relevant for this type of polymer.

The default segment type is ***** and is all segments in the first displayed and active molecule. The start and end fields represent the limiting values of sequence number (resID) that are displayed in the Ramachandran plot. Note that the Ramachandran angles for the start and end residue are not displayed because of incomplete data for these residues.

Show bumps less than 3.00

The optional value (default 3.00 Å) sets the distance under which bumps between atoms are displayed. This option affects many of the build tools. When toggled on, X-AUTOFIT will show all the bumps (shorter than the specified distance) as white lines with the bump distance labeled at the center of the line between the bumping atoms. You may need to turn off the bumps if you are making large changes to the structure and the bump markers obscure the molecule.

Show bumps to inactive molecules

This toggle option allows you to determine whether the bumps will be shown from edited coordinates to coordinates of MSF files that are inactive in the molecule management table. For example, if one open MSF file contains the coordinates that are being edited in **X-AUTOFIT | Build atoms** and a second molecule is open that contains coordinates from the last build as a comparison, then you will want to turn this toggle **Off**, because bumps

to the last built coordinates are not relevant—they are only used as comparison coordinates. If other molecules in the management table are other parts of the same structure, such as ligands and water molecules, then this option should be turned **On**, so that bumps are always observed even when these are inactive. Note that bumps are not drawn to atoms not displayed.

Symmetry picking mode

Goes to the real atom when you pick a symmetry atom.

This toggle changes the behavior of the application when symmetry atoms are picked for the purpose of placing the center of the view with the tool **Pointer | Place using coord**. When this toggle is set to **FALSE,** the application places the center of the display (plus the bones and the map) at the symmetry atom picked. When this tool is set to **TRUE**, any symmetry atom picked when using the **Place using coord** tool will *not* place the display at the symmetry atom coordinate *but* at the real parent atom of this symmetry atom. This can be useful when a symmetry-related residue has been found to be in the incorrect position: the actual residue can immediately be found with this tool and then edited. The **Symmetry picking mode** tool can also be used with NCS, so that when both molecules related by NCS are displayed as real atoms, the NCS atoms will overlay these real atoms. Any differences (where the NCS has broken during the refinement if only partial constraints or restraints have been used) can immediately be observed. You can then jump back and forth between the two copies of the NCS when this option is set **TRUE**.

Next residue step

The next residue step allows you to set the increment/decrement for the tools **Pointer | Place at next residue** and **Pointer | Place at previous residue**. If the step is set to **2**, the placement tools will jump along the polypeptide chain, missing alternate residues.

Number of DP

This tool allows the specification of the number of decimal points (DP) used on the non-bond contacts markers.

Hydrogen representation

This has three possible options: **None**, **Polar**, and **All**. When you first start X-AUTOFIT, it checks the first few residues and sets the default hydrogen

representation so that it matches them. If there are no protein residues yet built, or you wish to change the hydrogen representation, use this parameter to specify the mode you want.

Many tools on the Build palette cannot be used on residues that have incomplete or incorrect atoms, and this includes hydrogen atoms. If you change the hydrogen mode, then only the **XFIT | Build atoms | Geometric conformation** and **XFIT | Build atoms | Fit side chain by RSR** are allowed. These tools will build the residue from scratch, adding all the atoms and their correct names. You should use **Edit | Split & Clean** to change the hydrogen representation if there are many residues to change.

Search time limit

This option sets a default maximum search time for any of the “open ended” search algorithms within X-BUILD and X-LIGAND. These are the loop search in X-BUILD, terminal fitting in X-BUILD and the conformation search in X-LIGAND. The default time limit is 10 minutes, and, on reaching this time, the search will abort and provide a prompt box to continue, or finish the search.

Number of steps for regularize

This option sets the number of steps (cycles of refinement) used for the regularize tool under **X-BUILD | Build-atoms | regularize**. For a small number of residues the number of cycles of regularization is usually limited by the tolerance of minimization.

Phi/Psi restraint

This sets the optional restraint towards certain conformations normally found in proteins. The actual value of the restraint is very low. Regularization will take much longer to converge with this option set to any value other than **None**. All residues being regularized will have their Ramachandran angles set to the conformation requested by this option, so large changes in the atom positions can occur if the starting conformation is a long way from the requested value.

The **None** option turns off phi psi restraints, so these torsions are free to take any value that quenches the deviations from ideality for bonds, angles, impropers, and active non-bond interactions.

The **Helix** option refines phi and psi values towards ideal helical values, the **Strand** option will refine phi and psi values towards those found in an ideal

beta sheet, and the **Nearest** option will refine the phi psi values to the phi psi value nearest to the starting conformation, as defined from the Ramachandran allowed regions.

> **Note:** Restraints to phi and psi may change the coordinate positions of atoms to a better expected geometry, to the detriment of the quality of the experimentally determined structure.
>
> When you have a well defined electron map, you should not impose a phi/psi restraint. The restraint will reduce the fit to density in such cases.

Omega restraint

This option allows you to specify the required omega torsion that defines the peptide bond. The **Auto** option will first check to see if the peptide bond is in a trans or cis conformation, and then set a restraint towards the nearest ideal conformation. If a cis peptide bond is found by the regularization tool when in "auto" mode, X-AUTOFIT prints a warning to the textport.

The **Trans** and **Cis** options force peptide bonds to the specified conformation. Note that *all* peptide bonds in the regularized zone will be set to the requested conformation. Therefore, if a single peptide link is required to be in a cis conformation, then only regularize the two residues that make up that bond. During this process, some distortion occurs, due to large changes of atom positions. To correct for this, change the restraint to **auto** and regularize the region again. The auto detect will flag this as a cis bond and regularize towards it, while retaining all other peptide links as trans.

Nonbonding in regularization

The default for this option is **None**. Normally regularization should be considered a method to improve the geometry of a section of the structure as a function of the bonds, angles, and impropers. This option allows a full nonbond description to be used, including atoms not involved in the regularization (such as symmetry). You can turn on full nonbonding for the regularization protocol, but this slows down the calculation, will make interactive modeling difficult due to the decreasing refresh rate. Also, the inclusion of nonbond interactions makes the interactive regularization less responsive because refinement will not go to completion during each step of the interactive editing. The recommended procedure is that you should *not* use nonbonding for *interactive* modeling of structures, but only for **X-AUTOFIT | Build atoms | Regularize**.

There are three nonbond modes, for all crystallographic protocols the **push** function is the suggested mode. The **VDW** and **VDW | E** modes use a full energy function terms for the nonbond interactions, resulting in slower minimization. For crystal structure modeling no advantage is obtained using the full functions, and in fact they do not minimize well as the potential surface is complex for these terms.

Regularize across disulfide bonds

If this option is active, regularization (including **X-AUTOFIT | Structure | Refine zone**) will check to see if there are any cystine residues in the sequence before it proceeds. If there is a cystine then a check is made of the presence of a disulfide bridge to another cystine residue. If the other cystine residue is already contained in the regularize zone, then the parameters for this bridge are added to the parameters to be regularized. If the second cystine of the pair is not in the zone to be regularized, it is added to the list of residues for regularization and the bridge parameters are added to the list. If this option is not active, cystine residues do not receive any special treatment.

Coordinate radius

The coordinate radius sets the display radius about the current screen center for all the MSF files currently open. This is to allow very large structures to be manipulated on machines with poor graphics. We recommend that the coordinate radius should be at least 20 Å and larger than the map radius. If the coordinate radius is set to a value larger than 999 Å (the default is 1000 Å), then *no* checking of the atoms is carried out. This slightly improves the performance of the program when re-displaying a new screen center for structures where all atoms are required to be displayed.

Map radius

This radius value is in effect for any subsequent map calculations or bones calculations. The display immediately changes to the new map radius after a new value is input. The default value is 6Å, which is recommended for less powerful computers. This value is saved in the session file.

Symmetry radius

The display will immediately change to the new symmetry calculation radius after you input a new value for this parameter. All symmetry atoms

within the defined radius will be displayed. The default value is 10 Å. This option's value is saved in the session file.

Resolution factor

The resolution factor is used in the real space refinement protocol, and will result in automatic adjustment of occupancies as a function of how far away a side chain amino acid atom is from the Cα atom. The further the distance from the Cα atom, the lower the occupancy. For all resolution factors less than or equal to 2.0 Å, there is *no* change in the occupancies of any atoms. For resolution values greater in value than 2.0 Å, there is an increasingly steep sliding scale of occupancies as a function of distance from the Cα. A value of 3.0 Å will affect medium and large residues to some degree, while a value of 4.0 Å will have a sizable effect on the occupancy at the end of a side chain. The purpose of the parameter is to lower the weighting during refinement of those atoms further from the main chain, in areas where the density is probably of poor quality. This prevents those atoms away from the main chain from pushing the main chain atoms out of density, or wrapping around towards the main chain during refinement. This option has no effect on the actual MSF. The values of occupancies are only changed during the refinement procedure.

Color table

The **Color table** tool opens a palette that allows the colors and line thickness of all the objects generated within X-AUTOFIT and X-BUILD to be changed. The changes made in this table are saved between sessions. The values here can be used to make an object stand out if you are working with a particular piece of functionality. For example, the default "moving object" color and width is often insufficiently distinctive during editing of a molecule. You can change this option using the Color table.

Map (n)

There will be between 0 and 6 lines of "map" object labels. Each map label will have 1 **Width** field and between 1 – 7 **Color** fields, depending on the number of contour levels defined for this map. This Color table makes changing the color and line width of maps very easy, but can only be used on open maps. The default values are zero line width and color 14, then 13, and so on.

Bones main

The **Bones main** fields allow you to change the line width and color of the bones object that X-AUTOFIT considers to be part of the main chain. The default values are: line width = 3 and color = 6.

Bones side fields

The **Bones side** fields allow you to change the line width and color of the bones object that X-AUTOFIT considers to be part of the side chain. The default values are: line width =3 and color = 5.

Symmetry

The **Symmetry** fields allow you to set the color and line width of the symmetry atoms. The symmetry atoms include the “all atom” symmetry, the Cα-trace symmetry, and bones symmetry. The default values are: width = 3 and color = 2.

NCS

The **NCS** fields allow you to set the color and line width of the NCS atoms. The NCS atoms include the “all atom” NCS, the Cα-trace NCS, and bones NCS. The default values are: width = 3 and color = 3.

CA trace

The **CA trace** fields allow you to set the color and line width for the Cα trace display generated in the CA Build palette. There is a single line width value that sets the line width for all parts of the Cα trace, and three color values for different parts of the Cα trace.

Atom is the color used for the Cα-trace atom that is defined as the current atom. The current atom is the editable atom within the application. You can set it using the tool **CA build | current res-seg**.

seg is the color used for the current segment in the application, and is used for all the atoms in the current segment, not including the current atom. The current segment is set using the tool **CA build | current res-seg**.

rest is used for all Cα trace atoms in the application that are not the current atom or current segment. The default values are: width of 4, colors = 4 (red), 3 (yellow), 1 (red).

Text

Only the color of the text labels can be changed, as the text labels have no line width. The text labels are the character strings generated using the Text... palette, and by the validation routines within the X-AUTOFIT: X-BUILD application. The default color = 4.

Mask

The **Mask** fields correspond the map mask dot surface generated in the Map Mask... palette of the X-AUTOFIT: X-BUILD application. The default values are: width = 2, color = 5.

Vectors

The **Vector**s fields correspond to the objects generated as part of the application X-POWERFIT. These vectors indicate the major axes of the helices and strand elements found in the electron density map. The default width is 5, and colors 5 for helices and 10 for strands.

Moving atoms

The **Moving atoms** field sets the color and line width of any temporary object generated during an edit function in the X-BUILD application. For example, the objects generated by **Build atoms | Move zone**. The default values are: width = 3, color = 5.

Pointer

The **Pointer** fields determine the color and line width of the rhombohedral pointer used throughout X-BUILD and X-AUTOFIT. These Pointer settings also control the appearance of the spherical mask pointer. The default values are: width = 3, color = 5.

Last commands

This opens the last command table where all the X-BUILD editing history is stored. The table can be used to check the progress of the model building process, undo and redo each command, analyze the work carried out, and create log files of the use of the X-BUILD functionality

To hide the table, pick the **Last command** tool again. Whether the table is present or absent does not effect whether commands are saved to the file. As each command is issued (and as the user accepts the changes made by that tool), data is added to the *top* of the table.

Menu bar at the top of the table.

The last command table has a menu bar that contains tools for processing current and previous sessions of commands.

FILE: Write log file

This writes a log file to disk, although this is automatically done on exit from X-BUILD unless command saving is completely turned off.

An example is:

```
Command log file - written by Quanta: X-BUILD
Written at : 16:09:47 on Fri May 29 1998
Index    Command             Residue-range       time : date
1    Save changes(*)     A : 1-E : 6         15:25:09 on Fri May 29 1998
2    Save changes(*)     A : 1-E : 6         15:31:33 on Fri May 29 1998
Comment -> (1) bad fit to density
3    Save changes(*)     A : 1-E : 6         15:43:44 on Fri May 29 1998
4    Regularize          A : 4               15:54:28 on Fri May 29 1998
5    Goto pointer(+)     B : 26              16:01:32 on Fri May 29 1998
```

FILE: Command saving off

This will turn off all saving. To turn the saving back on again the user should pick the last command tool again on the main X-AUTOFIT X-BUILD palette and the application will prompt whether command saving is to be turned back on.

FILE: Browse old command files

This tool will open a scrolling list of previous session files. The list contains multiple entries as:

`(`*n*`) commands: (time written on date)`

The list is sorted by date/time written, where the current open file is read/write and is first. All previous files are READONLY.

You can open the previous file and look at the contents in the table, You can also do analysis, undo/redo etc. to see what you did in previous session. However, you cannot write a tool entry to any READONLY previous command session file. If you use a new tool, the previous file is closed, the current file is reopened and the tool entry added to the current session file.

FILE: Delete old files

This tool lets you delete old files or just delete all the old session files. A scrolling list is shown, and each entry can be kept or deleted. This command moves all session files down one to fill any gaps.

VIEW: Info (+) lines: Show/Hide

This tool controls whether info lines with (+) at the end of the tool name are hidden.

VIEW: User comment: Show/Hide

This tool controls whether the molecule display is labeled with the user comment cards added.

ANALYSIS: Check command file

This will return the percentage of each tool usage in a table to the text port.

ANALYSIS: Make suggestion

This will make simple suggestions as to the usage of X-BUILD as a function of the tool usage.

The last command table

The last command table contains seven columns of data that summarizes the changes that have been made to the coordinate model by editing coordinates with an X-BUILD function:

- **Command**. The command used.
- **Residue(s)**. The residue(s) that were edited by the command.
- **Time**. The date and time of the edit.
- **Undo**. A yes/no column to indicate whether the action has been undone.
- **Redo**. A yes/no column to indicate whether the action has been repeated.
- **Index**. An order to the commands.
- **Comment**. A user defined comment attached to the edit.

By default, the table is sorted by Index so that the last command is at the top of the table, but the tables can be resorted by picking the column headers **Command**, **Residues**, **Time**, **Comment**, so that it is possible to collate the tool usage differently. Note that clicking the **Residue** column sorts the table using the first residue within the edit range. If either the **Undo** or the **Redo** headers are clicked, then the program prints out an analysis of the commands used within the X-BUILD functionality and makes suggestions to the tool usage.

Cell picking

It is possible to pick cells within the last command table, and picking a cell results in an action defined by the column the cell belongs to. If a cell in the **Command** column is picked, then details of the changes made during the edit are printed to the text port. This information includes the RMSD, max and min atomic deviation per residue. If a residue cell is picked, then the program centres the display at edited residues for the particular command and moves the pointer to this position. If a **Time** cell is picked, then a full time/date record is written to the text port, where the table only shows the time information due to space limitations.

If the **Redo** or **Undo** cells are picked, you can redo and undo each tool edit as required. Note that the data stored for each edit consists of the coordinates (and temperature factors/occupancy) for each atom before and after the edit. Therefore an **Undo** modifies the coordinates so that the change made is undone. The **Redo** cell modifies the coordinates so that a particular edit is redone. Note that if multiple edits have been made to a particular residue then the last edit can only be redone from the last edit of that resi-

due, and all edits to a residue can only be recovered from the first **Undo** of the first command containing this residue. Additionally the changes to a residue can be studied at all the intermediated edits.

The command cell, when picked, allows you to add a comment that is relevant to the edit made. This comment is written to the log file of the build session, and can also be displayed on the structure view using the menu option (see below).

The menus allow you to open a previous last command table from an earlier model building session. When this is the case, all the cell picking is active and can be used to recover changes made at an earlier date. New edits to the structure are not added to a previous command table, but the current version is reloaded and change added to this.

The last command table therefore stores all the edits and details about these for an entire build session. This command table allows point in time data recovery, data analysis and a note book facility. It is also possible to use the last command table as a teaching aid as it can be set up to show best practice for map fitting.

External program palette

This palette, when no script files have been written, contains only the **hide this menu** tool. Tools will only appear on the palette if correctly formatted scripts have been written.

Scripts must have names script.# where # is a number between 1 and 20.

A script file is an ascii file of key worded instructions that defines the layout of a dialog box and the generation of a command file to run a particular external program. The script files also contain control key words to export coordinates from X-BUILD, and also to import coordinates and maps back into X-BUILD. The aim of script files accessed from the External program palette is to set up an interface to external programs allow you to run them directly from QUANTA. The script defines the layout of a dialog box, including hidden preference boxes and, using the data entered in the dialog box, runs an external program with the parameters defined by the user.

The scripts can be changed when the dialog is not visible as each script is only read when the dialog box is opened. A debugger is provided when the script is interpreted to provide some feedback on the generation of the script.

Definitions

Script: file containing the described information in this help that defines the look of a dialog and the contents of a command file.

Command file: file produced by a script that is submitted to be run by the computer, and is normally (but not necessarily) an external program command file.

General form of a script

Program (program name) (Text string description of program)
< assignments of variables >
< control statements >
< QUANTA data export statements >
< script definitions for dialog and command file >
< QUANTA data import statements >

Only the program name and description is required, although other statements are necessary for a meaningful script.

Variables

Variables within scripts can be defined and can either be text or number values, no distinction is required. A variable can be initialized using the assign keyword, or are taken from a dialog entry box.

- An initialized variable (assigned) can be used as an initial value parameter for a dialog entry box or a command file definition.
- A variable defined from the a dialog entry box can be used in a command file definition, or a hidden dialog (preference box) state box.
- A variable defined from the dialog entry box *cannot* be used to define an initial value parameter for a dialog entry box because its initial value is undefined.

General formatting

A text string without spaces can be added to a script *as is*, but text strings containing spaces or commas much be enclosed by double quotes to delimit the string.

All fields are required — an empty text field can be placed with “ “quotes; note that a space is required.

A text string delimited by { } curly braces defines a variable in the script that will be replaced by the value of the variable when used.

All short fields are, by default, un-initialized variable names for the entry field they specify.

Program definition line

This line is the only required line and must be the first line of the script. The keyword is **program**, and the first parameter is a text string that is written at the top of the dialog box while the description text is not used.

Assignment of variables

Initialized variables are set using the line:

```
assign data “input_data”
```

The name of the variable is (data) and its initial value is a text string (input_data). This variable can then be used with a command file line as:

```
write “FILIN  {data}”
```

(The **write** keyword writes the line *as is* to the command file).

The variable can also be used to define the initial state of a dialog input:

```
string 1 “Input” “File” “{data}.pdb”  “Input PDB file” for-
mat 40
```

(The string defines a dialog string input).

Un-initialized variables

Uninitialized variables can be used when their value is defined from a dialog entry field when the **OK** button is clicked. Generally they can only be used in a command file line:

```
write “calc u(100) = {Bond}”
```

The quoted string here is written to a command file when the dialog **OK** button is clicked, and the value of the variable ***bond*** is taken from a dialog entry value. Make sure the variable is defined or nothing will be written.

An example dialog entry field to define this variable:

```
dreal 1 “n7” “Bond” 0.10000 “Warning limit (A)”
```

The **dreal** keyword defines a dialog only entry field (does not provide output to a command file - hence can *only* be used to set a variable), and the

variable “Bond” is defined by the short string label for the real entry field whose initial value is 0.1.

Variables can also be used for hidden dialogs such a preference dialog box:

```
dbutton 1 “n1” “Preferences” “ “
dif {Preferences} then
dreal 1 “n7” “Bond” 0.10000 “Warning limit (A)”
dreal 1 “n8” “Angle”  5.0 “Warning limit (o) “
dendif
```

This example shows a dbutton keyword defining a dialog only field button with the short name of **Preferences.** The **dif** keyword defines a dialog only condition statement that will display the contents of the **dif - dendif** statements as a dialog if the dbutton **Preferences** is pressed. (The contents of the preference box would contain two dialog-only real entry fields to define the value of the variables **{Bond**} and **{Angle}**).

Control statement

The control statements define the general behavior of the dialog and command file submission.

```
saveall
saveOK
savecancel
nosave
```

These four control statements are used to define whether the contents of the dialog entry fields are remembered when the dialog is exited. The normal default would be to saveOK, where the contents of the dialog fields are written back to the script file when the dialog **OK** button is pressed. Hence, the value of the dialog entry fields are remembered when the user wants to carry out the action (to run the command file). “Saveall” will always save the entry fields; “savecancel” will only save the fields if the “cancel” button is pressed, and “nosave” will not save the value of the entry fields back to the script.

```
notify
```

The notify control statement will write a <Crtl> G to the end of the command file, so that the completion of the command file will result in a beep.

```
nowait
```

```
wait
```

The nowait control statement results in a submission of the command file and immediately returning control to QUANTA. This option is sensible when the calculation performed by the external file takes some time. The default wait statement pauses any action in QUANTA until the command file action has completed.

```
#
```

The # character as the first character on a line is used to define a comment line, and any subsequent text on the same line is ignored.

QUANTA data export statements

There is currently only one data export statement **output**, though the format allows future additions. If the dialog OK is accepted, all **output** statements are carried out before the script is run regardless of their placement in the script file. It can therefore be assumed that the filename produced will exist when the command file is submitted to be run.

```
output pdb (filename).
```

The output statement is followed by a format (in this case pdb - a file of coordinates in PDB format for active and visible molecules in QUANTA/X-BUILD). No other formats are currently allowed.

The filename is the file name of the output coordinates.

QUANTA data import statements

There are currently only two data import statements, though the format allows future additions. If the dialog OK is accepted *and* a **wait** control statement is provided, all **import** statements are carried out after the script is run regardless of their placement in the script file. If a **nowait** control statement is placed in a script with an **import** statement, the imported information will be undefined.

```
import pdb (filename)
```

The import statement is followed by a format (in this case pdb). The imported coordinates will *replace* the currently active and visible coordinates in QUANTA/X-BUILD. It is necessary that the number of atoms, the atom names and size of residues exactly matches that of the QUANTA data structure.

```
import map (filename)
```

The import statement will open a QUANTA brick map of name (*filename*).

Script definitions key words

The order of the script definitions defines the order they appear on the dialog and also the order of the lines of text in the command file. It is generally most efficient to use definition keywords that define both the dialog and

command file, though this sometimes is not possible -- hence the use of variables.

Script definitions can be divided into three categories:

1. Those that control only the structure of the written command file.
2. Those that control only the structure of the dialog.
3. Those that control both the dialog and command file.

Command file control statement

There is only one keyword that controls the command file; this is used to write a string of text (including numbers) when the command file is generated. It is possible to use variables in the written text to enter variable values defined from the dialog entry fields or assignment statements.

```
write “a string of text {including} variables”
```

The text string in quotes is written *as is* (except for the variables) to the command file.

Dialog only control statements

Dialog only statements define the layout and field entry within the dialog and have no effect on the generation of the command file. They can be used to set the value of variables or to generate hidden dialog boxes. There is no reason to create a dialog only field entry that does not define a variable. *All* dialog entry fields implicitly define variables for each entry field. The variable name is the “short string” of the field. All statements are free format.

- space (No fields)
- dstring (variable field = “string”)
- dinteger (variable field = integer)
- dreal (variable field = real)
- dtoggle (variable field = logical T/F)
- dbutton (No variable field)
- dif
- dendif

Except for the **space** keyword, the general format of a dialog entry field is...

```
(keyword) (Nfields) (Command string) [ (short string) (field
default) ] (long string) (format fields)
```

- (keyword) - This is one of the dialog keywords that defines the entry field type
- (Nfields) - The number of entry fields to follow (shown in[ ]).
- (command string) - Not used in this context (would be written to the command file for a dialog/command file entry line).
- [ ] the 2 data statements inclosed are repeated (Nfields) times.
- (short string) - short string label associated with each field and written next to the entry field, and the implicit name of a variable of value defined by the value in the entry field.
- (field default) - the initial value of the entry field.
- (long string) - a string that labels the entry line of multiple fields.
- (format fields) - see definitions, used to change the look of the dialog.

The following are the allowed keywords.

- space: creates a one line space in the dialog box.
- dstring: creates string field entries.
- dinteger: creates integer field entries.
- dreal: creates real field entries.
- dtoggle: creates true/false toggle button entries.
- dbutton: creates a button(s) - would normally be used for hidden boxes.
- dif-dendif: delimits a hidden box whose initial state is hidden, but appears when a button linked to the if {variable} is pressed.

For example:

```
dreal 2 “bond-angle” “Bond” 1.0 “Angle” 2.0 “Parameters:”
format 8.3
```

A two field real number entry dialog string where the initial value of the bond entry is 1.0, the initial value of the angle entry is 2.0, and the format of the entry field is f8.3. The look of the dialog would be a line.

```
Parameters: Bond [   1.000  ] Angle [   2.000 ]
```

For example: A hidden box (note that there is no default value field here):

```
dbutton 1 “hidden-box” “” “ “
dif {Preferences} then
dreal 2 “bond-angle” “Bond” 1.0 “Angle” 2.0 “Parameters:”
format 8.3
dendif
```

The **dbutton** entry line defines one button labeled Preferences and no preceding long string name as indicated by the empty quotes.

```
[Preferences]
```

The hidden box contains the line in the previous example. When the button is pressed, the current dialog is replaced by the hidden dialog box (a single line with two real entry fields here), and a single [return] button to exit back to the parent dialog box.

Note: Nested dialog boxes are not supported.

Dialog and command file statements

These statements are the most useful to create scripts as they control both the dialog and command file. They have exactly the same format as dialog only lines. Not only can these statements control the look of the dialog and the command file, the fields can be used a variables, but this may result in multiple use of a field. All statements are free format.

- string (variable field = “string”).
- integer (variable field = integer).
- real (variable field = real).
- toggle (variable field = logical T/F).
- button (No variable field).
- option (variable fields = logical T/F).

```
(keyword) (Nfields) (Command string) [ (short string) (field
default) ] (long string) (format fields)
```

- (keyword) - This is one of the dialog keywords that defines the entry field type.
- (Nfields) - The number of entry fields to follow (i.e. inclosed as []).
- (command string) - This string is written as is to the command file when the command file is written. Normally this would be a keyword for the external program.
- [ ] the 2 data statements inclosed are repeated (Nfields) times.
- (short string) - short string label associated with each field and written next to the entry field, and the implicit name of a variable of value defined by the value in the entry field.
- (field default) - the initial value of the entry field.
- (long string) - a string that labels the entry line of multiple fields.
- (format fields) - see definitions, used to change the look of the dialog.

The following keywords are valid:

- string: creates string field entries.
- integer: creates integer field entries.
- real: creates real field entries.
- toggle: creates true/false toggle button entries.
- button: creates a button(s) - would normally be used for hidden boxes.
- option: creates a option list - where only one value is true.

Example:

```
option 3 “REFI METHOD” “CGMAT” T “CDIR” F “CGRADD” F “Refine-
ment method” xstart 2500 xsep 1500
```

Format additions

Format additions are added to a dialog definition that change the look of the dialog box for a statement. They have no effect on the command file.

```
Format (n):(n) depends on string/integer/real fields

xstart (n): Start point of the fields in 1/1000 inch

xsep (n): Separation of the fields in 1/1000 inch

append: Place current statement fields at the end of the pre-
vious statement fields

vertical: Place each field entry on a new line - vertical
arrangement
```

For integer field, **format 4** would allow a four character integer number entry from -999 to 9999.

For a real field, **format 8.3** would allow a eight character number entry with up to three decimal places.

For a text string, **format 20** would allow a twenty character string entry/

Continuation marker

It is possible to add a continuation mark to the end of a line in the command file by adding a “/” at the end of any statement line. This does not affect the dialog box.

Examples:

1. The following example creates a single text entry using a dialog only statement and a variable {file} to pick up the user defined name for the export statement. (Note that the {file} variable is the short text name for

the text entry field. By default the file name is updated in the script when [OK] is pressed. (i.e., default “saveOK”)

```
program “Write PDB” “Write of displayed/active molecules”

dstring 1 “n1” “file” “script.pdb” “PDB name” xstart 2500
xsep 2500 format 40.00

output pdb {file}
```

2. The following script was used to run the program SQUID for the validation of proteins. This external program functionality has now been replaced by the Tables and Graphs functionality within QUANTA and is therefor no longer provided.

```
program squid “Squid Validation”
wait
assign strdir “$HYD_LIB/squid/”
output pdb script.pdb
write “#!/bin/csh -f”
write “setenv SQUIDIO /y/programs/squid/squidio”
write “$HYD_EXE/squid -none -file script.pdb -put 300 300 700 700 << ‘END’”
write “echo”
write “check”
write “xfit open”
write “sel resi protein”
write “excl atom h* d*”
write “calc u(100) = {Bond}”
write “calc u(101) = {Angle}”
write “calc u(102) = {Plane}”
write “calc u(103) = {Clash}”
write “calc u(104) = {Probe}”
write “calc u(105) = {Water}”
label “Refinement parameter check”
toggle 1 “stream {strdir}check_restraints.str” “ “ TRUE “Check bonds etc.” xstart 2500 xsep
toggle 1 “stream {strdir}check_HNQ.str” “ “ TRUE “Check HNQ” xstart 2500 xsep 2500
toggle 1 “stream {strdir}check_cis.str” “ “ TRUE “Check for cis peptides” xstart 2500 xsep 2500
toggle 1 “stream {strdir}find_holes.str” “ “ TRUE “Check for possible voids”
toggle 1 “stream {strdir}check_clash.str” “ “ TRUE “Check for clash” xstart 2500 xsep 2500
dbutton 1 “n1” “Preferences” “ “
dbutton 1 “n2” “Don’t press this !” “ “
dif {Preferences} then
dreal 1 “n7” “Bond” 0.10000 “Warning limit (A)”
dreal 1 “n8” “Angle”  5.0 “Warning limit (o) “
dreal 1 “n9” “Plane” 10.00000 “Warning limit (o)”
dreal 1 “n3” “Clash” 1000.00000 “Warning limit (Kcal)” xstart 2500 xsep 2500
dreal 1 “n5” “Probe” 1.2 “Voids Size (A) “
dtoggle 1 “n6” “Water” FALSE “Use water” xstart 2500 xsep 2500
dendif
dif {Don’t press this !} then
  space
  space
  label “            Boo !”
  space
  space
  space
dendif
write “ “
write “xfit close”
write “end”
write “yes”
write “‘END’”
```

References

1. Greer, J., *J. Mol. Biol.* **82,** 279-301 (1974).
2. Oldfield, T.J., and Hubbard, R.H. *Protein: Structure, Function and Genetics*, **18,** 324–337 (1994).
3. Ramachandran, G.N., and Sasisekharan, V. “Conformation of polypeptides and proteins”, *Adv. Prot. Chem.*, **23** 283–437 (1968).
4. Dickerson, R.E., and Guis, I., *The Structure and Action of Proteins*, Benjamin/Cummings, ISBN 0-8053-2391-0.

9. Using X-LIGAND

The X-LIGAND application automates the process of fitting ligand coordinates to electron density maps with little or no user intervention.

Overview

X-LIGAND was designed for the crystallographic protocol in which ligand binding experiments have successfully produced protein/ligand complexes that form isomorphous crystals with respect to the apo form of the protein. These ligand binding experiments result in isomorphous crystallographic data where the protein part of the model structure can be solved directly, or by molecular replacement, using the apo protein model. The ligand binding sites should be apparent within omit density maps from these experiments as connected density that corresponds in approximate size to that of the ligand. There will also be other electron density due to data and model errors.

This experimental protocol can be relatively rapid with many ligands being successfully bound and data collected in a short period of time. The result is that the model building process becomes the rate limiting step for the structure determination.

X-LIGAND is therefore designed to search for unsatisfied electron density (density containing no molecular coordinates), sort these in order of volume, and fit a ligand to these sites automatically. The application is also able to search conformation flexibility of a ligand by varying any rotatable bonds and fit these rotamers to density at a rate of more than a thousand per second. This entire process, including refinement, can be carried out quickly with X-LIGAND.

Because of the success rate of producing apparently correctly fitted molecular coordinates to electron density, X-LIGAND has been found useful for any general fitting of coordinates to omit/2Fo-Fc/fo-fc data, including small inorganic molecules (such as sulfate ions), polysaccharides, and polypeptides.

Requirements

X-LIGAND requires a set of protein coordinates, one or more sets of ligand coordinates, a map, and symmetry of the crystal system. The map may either be of difference density or (n+1)fo-(n)fc as the program excludes all regions with model coordinates already present in density. If the ligand has rotatable bonds, the initial conformation of the ligand is not important.

Ligands

Because X-LIGAND is designed to use multiple ligands that do not “see” each other (nonbond interactions), then all parts of the molecule for which you want to calculate nonbond interactions must be part of the first MSF. This is not normally a problem with molecules whose waters are not included in the file, because the solutions from the search are sorted by size, and hence the best solution is first.

Rotatable bonds

The application uses a routine to determine the rotatable bonds within ligands, since this is necessary for all the refinement methods used in X-LIGAND. This routine is not based on atom typing and does not require the presence of hydrogens. This means that X-LIGAND is able to determine the correct number of rotatable degrees of freedom for any ligand regardless of the hydrogen mode of the molecule. Since the algorithm is based on bonding of the ligand and geometrical analysis of the structure, the ligand used in X-LIGAND must be close to its energy minimum. Since this is required for the application (as only torsional conformational space is explored), this does not represent a limitation in the application.

Rotatable bonds are rejected in a structure if:

1. The two central atoms are part of a ring.
2. The 2nd or 3rd atom is a terminating atom.
3. The two central atoms are able to form a partial double bond such as two sp2 carbon atoms.
4. The first or fourth atom is a hydrogen.

Since hydrogen atoms (in polar or all-atom models), have little electron density associated with their position, fitting hydrogen atoms increases the complexity of the search problem while adding little to the quality of the final structure. If you must place hydrogen atoms in the search, as with high

resolution structures, then you can add torsions that define their position using the tool **Define 1 torsion**.

Edited rotatable bonds

If the rotatable bonds are changed to include more, or remove some, then the application creates a file of the new definitions. It is not necessary to know the contents of this file, or edit the file, as there are tools within X-LIGAND and X-BUILD to edit this file using a graphical user interface. This file is of the format:

```
Title line

Residue_name  atom1  atom2  atom3  atom4
Residue_name  atom1  atom2  atom3  atom4
etc.
```

For example, for a lysine and phenylalanine residue, the file could contain:

```
Rotatable bonds for the two amino acids lys and phe
LYS N CA C O
LYS N CA CB CG
LYS CA CB CG CD
LYS CB CG CD CE
LYS CG CD CE NZ
PHE N CA C O
PHE N CA CB CG
PHE CA CB CG CD1
```

The title line can be anything and is ignored, but must be present. The residue/atom definitions are free format and delimited by spaces.

Each line defines a rotatable bond for a residue. There can be as many as 100 rotatable bonds for a single residue, and as many residue definitions as required. The ligand must only have a single residue name. The residue name is a string of up to four letters and must be the same as this ligand's name in the MSF. The series of four atom names on each line defines the torsion angle; the rotatable bond lies between the second and third atom. Any lines in this file that lack either the residue name for the current ligand or atom names are ignored.

MSF file requirements

The protein molecule must be the first molecule in the Molecule Management table. If this is not so, close all the preceding MSF files and reopen them, using the append mode before you enter X-LIGAND. X-LIGAND aborts if the first molecule is not a protein.

A ligand to be fitted must be provided to X-LIGAND as a separate MSF file. Multiple ligands can be provided to X-LIGAND so that each different ligand is stored in different MSF files. Hence it is necessary that at least two MSF files must be open, displayed, and active to use the X-LIGAND, but

more than two MSF files can be provided if different ligands need to be fitted to a protein. The protein coordinates must be in the first MSF file.

Symmetry

The X-LIGAND application generates a sphere of symmetry atoms around the working molecule so that symmetry ligand sites are not found as multiple sites. The symmetry atoms display as blue and the NCS atoms as red.

General use

Active site searching

X-LIGAND first finds all sites around the molecule where the electron density is above a user-defined threshold value, but in which no model atoms occur. These sites are then sorted by size, with the largest region of density first. The ligand is placed at this largest site by the application, and if the ligand has no internal degrees of freedom, then it is necessary only to refine the results.

Conformation searching

Once all the sites have been determined by the application, if there are several internal degrees of freedom in the ligand, possible conformations can be fitted to the density at the rate of about 2000/second. The conformations are searched by changing the rotatable bonds, where the rotatable bonds are determined automatically if the molecule has been built in the 2D builder, or provided explicitly as an external file. X-LIGAND automatically determines the precision at which rotatable bonds need to be searched to efficiently give good results. The largest ligand so far tested had 16 rotatable bonds; it took about 10 minutes to solve (about 1,000,000 conformations searched).

Precision

X-LIGAND weights the search precision of each torsion depending on the effect each torsion has on the overall shape of the molecule. Torsions in the middle of the molecule will be searched to the maximum precision of 1°, while those torsions that only affect the placement of 1 or 2 atoms may only be searched at a precision of 180°. This weighting scheme allows the appli-

cation to be computationally efficient, and prevents many solutions with almost identical coordinates.

Once the application has searched the conformations and found a good solution to the fitting problem, it is possible to carry out manual changes to improve the fit. These can be carried out at any time, and may be necessary if the ligand is particularly flexible (more than 10 internal degrees of freedom).

Refinement

The ligand can now be refined using torsion angle real space refinement to the electron density. This has a very high radius of convergence (about 2 Å), and will refine the ligand regardless of errors remaining from the search. The refined ligand can be saved to a new MSF at any time.

For a rigid molecule, the program will immediately fit the ligand to a site. You need only refine the position.

The conformation search tries to simultaneously fit the position, orientation, and rotatable bonds of the ligand, and keeps the results that fit the density the best. If you have a particularly flexible ligand (more than ten torsions), then the program will probably require an omit map so that you do not get fits to other density. For more rigid molecules, you can use a 2fo-fc map, since X-LIGAND checks each site for nonbonds. The refine option is very powerful, and will often fit something a long way from it minima.

Masked tools

Initially, only the first three tools are unmasked, (plus **Exit**), as nothing can be calculated until the sites have been searched for. Once the ligand sites are found, other tools become active. If there are no internal degrees of freedom, the conformation search and the sub-tools under it will not be unmasked, since they are not relevant for this type of ligand. If you wish to change the torsion list, the tool to **add/del 1 torsion** can be used to edit the list of torsions at any time, and the **Select new ligand** tool used to re-read the contents of this file and reset all the necessary options. If you have only one ligand, this process occurs without a prompt, otherwise a dialog box prompts you to select the ligand again.

X-LIGAND palette

Search for ligands

This searches for regions of density that have no atoms and returns the sites, sorted by size. The biggest region of density is the first site. The ligand site size is defined by volume of density that is connected, above the defined threshold (default=1.5 sigma), and not overlapped with the atoms in the first MSF. This will take about 5–10 sec. The program places the ligand, in its current conformation, in approximately the best orientation at the first site.

Adjust site volume

This tool adjusts the volume of the electron density to be searched to be the same as the volume of the ligand. This works by adjusting the sigma value used in flood fill until the density volume found matches the volume of the ligand. The initial value for the sigma level is defined within the **X-LIGAND | Change search setting…| Search threshold** tool. In general, it appears that setting this to quite a high value (3.0) is best. The Adjust site volume tool then searches down to find the correct extent of electron density for the ligand to fit into. Occasionally, there is a problem with the tool not finding the correct volume if the initial value is left too low.

Change search setting...

This opens a dialog of options that allows basic search parameters to be modified.

Search threshold

This sets the map level (in sigma) at which the search for possible ligand sites is made in the map. A lower threshold will result in more sites being found, and will also increase the extent of the volume of each site that does not overlap the protein. A lower threshold will often result in the overlap of sites, giving fewer, larger sites. This is the only parameter that needs to be set by you. The default is appropriate for searching 2fo-fc maps.

Conf search mode

This is automatically set by the program. If the number of internal degrees of freedom results in the total number of searched conformations exceeding 50,000, then a **Random** search is set, otherwise a **Grid** search is set. You can override these defaults by selecting the mode you want.

Distance and positional restraints are supported to increase the success rate of the conformational search during ligand fitting.

Symmetry atoms can be picked and the information displayed.

Electron density truncation

This option allows you to eliminate electron density in the ligand binding site that overlaps with the protein atoms or their symmetry equivalents.

Non-bonding

Defines the type of non-bonding interactions between the receptor molecule and ligand. The following interactions are supported:

Self
Full
None

Partial bond torsion search

This allows you to turn on or off the option of having flexible partial bonds in the ligand conformational search. For example, the peptide bond.

Use 3rd moments

To place the ligand in the density initially, X-LIGAND compares gyration radii and moments of inertia for the density and the ligand to determine the initial position and orientation of the ligand in the density. There is a new option to use third moments of inertia for the shape-matching. If **No 3rd moment** is chosen, the second moments of inertial are chosen, as in the earlier release of QUANTA.

Z-weights

Option to apply Z-weighting restraints to the ligand atoms based on their atomic number. This option is helpful if the ligand contains heavier atoms such as Hg or S. The ligand-placement algorithm will weight both the heavier atoms in the ligand, as well as the regions in the ligand electron density volume.

Regularize

This option allows you to perform or skip regularization on the ligand geometry before and after conformational searches.

Grid search parameters / Max conf to search

In a grid search, this parameter forces an upper bound to the number of grid values to search. The program automatically determines a grid search resolution for each torsion, so if the product of all torsions to search exceeds this number, the grid step sizes will be increased until the total number of search conformations is less than this value.

Max time for search

This parameter defines the length of the Monte Carlo search.

Intra-ligand Non-bonding

This option defines whether nonbond interactions between ligand atoms are to be included (**none** or **self**) and allows you to turn on or off the calculation of nonbond interactions with the protein (**full**). For most ligands and electron densities, this energy term is not needed for a reasonable fit, particularly when the electron density volume is close to that of the ligand volume.

When the electron density volume is smaller than the ligand, or the ligand is to be fitted into a small void of the protein, then nonbond terms are needed between the ligand atoms to aid the fitting process. The ligand will fold up onto itself if these terms are not calculated.

This energy term can be turned on and off, because there is a time penalty for this energy term, the calculation usually being about 3 times slower. Nonbond terms between the ligand and any other visible molecule are

always calculated, since they are pre-calculated before the search and thus provide no time penalty within the calculation.

Grid search parameters / Max tors to search

The default maximum number of torsions to search for a grid search is 12. If there are more than 12 torsions to search, X-LIGAND will not search the least important torsions.

Radii / Map display radius

This parameter sets the radius of displayed map about the ligand.

Radii / Calculation radius

X-LIGAND sets this parameter to cover the entire molecule of interest plus 5 Å beyond the most distant atom. Use this option to change the scope of the search algorithm.

Position search centre

The tool to position search center allows a particular region of the map to be focused on by setting the center of the map search to a particular point in space, and then using a smaller calculation radius (**X-LIGAND | Changesearch setting… | Radii / Calculation radius**). This allows faster use of the application on very large molecules, but is probably not necessary for most problems. On selecting this tool, a pointer will appear at the screen center (the current search center), and can be moved with the dials. A new menu will open, allowing you to accept or reject the new position. See Using the 3D Pointer (page 265), for detailed information on the **Pointer** palette.

If **Accept position** is selected from the Pointer palette, the search center will move to this position, and symmetry recalculated based on this center and the calculation radius. If **Quit** is selected, then no change is made.

Map on/off

This tool provides an easy method to turn the map display on and off from the X-LIGAND menu. When the maps are not displayed, the ligand site can be observed more easily, and on machines with slower graphics, refinement

and ligand editing will proceed more quickly because the display will be redrawn more quickly.

Reset Ligand

This tool resets the position of the ligand to that of the original ligand on entry into the X-LIGAND module. This is needed when scoring the fit of a particular ligand conformation to the density.

Next site

This moves the display to the next potential ligand site and places the ligand, in its current conformation, in the site in its best orientation.

Previous site

This moves the display to the previous potential ligand site and places the ligand, in its current conformation, in the site in its best orientation.

Goto site

This allows you to go to a particular site directly rather than navigating with the **Next site/Previous site** options.

Search conformations

If the ligand contains internal degrees of freedom such as rotatable bonds, then it is possible to search conformations of the ligand as a function of these torsion angles. The program can search up to 100 rotatable bonds simultaneously.

During the search, the program automatically weights each rotatable bond according to the effect rotating the bond would have on the structure. For bonds that will significantly affect the structure, the search step will be 1°. For a rotatable bond that has very little effect on the structure, the search angle will be 180°.

If the total number of search conformations exceeds 50,000, the application selects the monte carlo search method, which is stopped only by an aborting mouse click or when the search time limit expires after 10 minutes. If the number of conformations to search is less than 50,000, the application carries out an exhaustive search, normally in less than one minute. In this

case it would be normal to allow the application to complete the search, but you can abort it with a mouse click if you need to. During the search process, the current best solution is displayed, so if you see a good solution, you can stop the search and this best solution becomes the first of the best 20 displayed.

...Refine all

This tool refines all the 20 conformations found by the **Search conformations** tool and then orders the conformations on the basis of the scoring function.

...next conformation

The program, on completing the search, lists the 20 best solutions in order of density fit. You will often find alternative conformers in this list. You can look through the list using the tools next conformation and previous conformation. Next conformation will show the next best fit, unless the current fit is the last one, in which case the next fit is the first one.

...previous conformation

The program, on completing the search, lists the 20 best solutions in order of density fit. You will often find alternative conformers in this list. You can look through the list using the tools next conformation and previous conformation. Previous conformation will show the previous best fit, unless the current fit is the first one, in which case the previous fit is the last one.

...all 20 conformations

This tool will show the best 20 conformations found by the conformation search. This can show how the conformations cluster and may indicate a single well-fitted conformation when all 20 are very similar, or a weakly restrained conformation when all 20 conformations are similar but show significant variation, or that there are multiple conformations.

Note that when this mode is highlighted then it is not possible to refine or save the coordinates as the applications does not know to which structure the operation refers. Selecting this tool again will turn of the display of all the best solutions and return to the previous displayed solution.

If the search for conformation results in less than 20 solutions, then only those found will be displayed.

…Fix search origin

This tool allows you to turn off the positional fitting of a ligand. This is useful when a monomer of a polymer is being fitted to density for each monomeric site. To use this option, move the ligand manually to the center of the required site, then fix the origin. When a conformation search is subsequently carried out, only the orientation and internal degrees of freedom are searched for that ligand, thereby fitting the ligand only to the position required. This tool does not make the search proceed any faster.

…Fix ligand atoms

This tool allows you to fix one or more atoms in the ligand molecule during the conformational search.

…Clear fixed atoms

This tool resets the atoms fixed by the previous tool.

…fit in stages

This tool fits the portions of the ligand with better density (based on shape-matching) before fitting the rest of the ligand molecule. This option reduces computation time.

After a ligand is fit with this command, one or more previously unconstrained atoms may end up being fixed and several torsions may also be inactivated. Repeated **…fit in stages** sessions can proceed from that state, but you may sometimes prefer to reactivate some or all torsions. In that case, the constraints from atoms and torsions should be manually removed before repeating **…fit in stages**. The commands **…Clear fixed atoms** and **Set active torsions** should then both be used.

Flip tor 180

This tool, which is a copy of the one on the X-BUILD | build atoms palette, flips a selected torsion within a ligand by 180 degrees. The selected torsion within the ligand must be rotatable or nothing will be done.

Set active torsions

This gives a menu with all the active torsions highlighted. The torsions present in the molecule are shown on the ligand molecule for reference. Initially, all the torsions are set as active (denoted by a check mark before the torsion), and can be inactivated by selecting the tool for a torsion. An inactive torsion is not searched or refined.

When a flexible ligand molecule is fitted to electron density using the **X-LIGAND | Search conformations** tool, it will be found that the central core of the ligand will be fitted to the density first. As the search continues, torsions will be refined so that larger parts of the structure will be fitted around the central core of the ligand molecule with peripheral parts of the ligand poorly fitted. It is therefore possible with this tool to inactivate the torsions that no longer need to be refined by the search procedure because this part of the ligand is already fitted. This will improve the radius of convergence for fitting regions of the ligand not yet fitted.

Add/del 1 torsion

Allows you to specify and remove torsions manually by picking a single bond. When selected, the tool will label all the current rotatable bonds (either calculated, or from the file lig.rot), with the torsion number. It is possible to pick a bond that is either already defined as a torsion, which will be subsequently deleted, or a new bond that will be added to the list of torsions, if valid. An invalid torsion is one that is part of a ring or is a terminating bond. It is possible to add a torsion between atoms not normally recognized as rotatable, such as a peptide bond or where the 1st/4th atom is a hydrogen atom. After picking a bond, the tool exits. If you want to continue adding torsions, you should use the same tool again.

Once this tool has been used to modify the rotatable bonds in a ligand, the application creates or modifies the lig.rot file, lig.rot, which contains the definitions for rotatable bonds. This file is then used in any subsequent analysis of the ligand. Note, other definitions for different residues are retained in the lig.rot file.

Edit torsions

This allows you to manually edit the torsion angles of the ligand, up to a maximum of 100 torsions. If there are more than seven rotatable bonds, then the eighth dial allows you to toggle up and down the list of rotatable bonds. If you have no torsions, this option does nothing.

Edit position

This allows you manually move the ligand position and orientation. The dials will change to allow movement of the ligand in xyz and rotated about the xyz axes.

Refine

This does real space refinement of the current ligand conformation's torsion angles in the current site. When you select the refine tool, the position, orientation, and all defined torsion angles are refined to the electron density by least squares.

Score fit

This ligand scoring tool takes the current ligand conformation and assesses how well the ligand fits the density. The larger the score, the better the fit.

Save ligand to MSF

This creates a new MSF of the current coordinates. The application will open a file browser and prompts for a filename for the new ligand position and conformation. The option **SAVE** saves the current ligand coordinates, while **CANCEL** aborts the save and does nothing.

Select new ligand

This tool allows a different ligand to be selected from the currently open MSF files. If more than one ligand was open on entry to X-LIGAND then a dialog box will appear, allowing a new ligand to be selected. If **Select** is chosen, this will become the new active ligand and will be fitted, and sourced in subsequent parts of X-LIGAND. If **Cancel** is picked, no action is taken. This option resets all torsions and recalculates the step size for the search, and checks the search method to use.

Database/list

This tool is used to carry out multiple ligand fits to a single site. For the current site it is possible to:

- Provide a list of ligands (such as molecules in the crystallization media), which are then fitted in turn to the site, and the best fit is returned.
- Provide a database of ligands that will be searched to find the best twenty ligands in the database that fit the site, and a view of all twenty results.

A dialog box is opened that has a file browser and 2 sets of options. A list of skeleton data base files are shown (*.ind) as default. A skeleton file can be chosen to carry out a database search.

The first option list is for the file type. The default is the list of skeleton files (*.ind). If the second option is chosen then all the MSF files in the directory are listed with the file name *001.msf. This option allows a small number of ligands to be placed into multiple MSF files with names **nnn*.msf where *nnn* is a number from 001 to 999. This is designed for simple access to a small number of ligands. Each MSF file must only have 1 ligand in.

The second option defines what data is to be returned. The default is to return the best fit ligand, all ligands currently open (molecules 2 and higher) will be closed, and on completion the best fit ligand is opened. The “Best 20” option will return the best 20 ligands, and on completion of the search the best 20 (or the number of ligands in the databank) will be opened and displayed with the best fit ligand as molecule 2 and the worst of the 20 ligands as molecule 21.

If cancel is picked from the dialog box nothing is done. If OK is picked from the dialog box then the search is carried out.

Exit

The **Exit** tool will exit from the X-LIGAND application and restore all the previous atom selections. If any new ligand positions have been saved while in X-LIGAND (**X-LIGAND | Save ligand to MSF**), then these will be opened and be referenced in the molecular management table. The position of the search ligand will be that on entry to the X-LIGAND unless a new ligand was selected while in X-LIGAND (**X-LIGAND | Select new ligand**) when it will appear at the site on changing the active ligand.

The **Exit** tool in X-LIGAND will not save the current ligand position; these new fitted sites are saved using the tool **X-LIGAND | Save ligand to MSF**.

10. Using X-SOLVATE

X-SOLVATE is a function within QUANTA X-BUILD or X-AUTOFIT which allows you to search an electron density map (fo-fc, 2fo-fc, or 3fo-fc) for water peaks, to adjust the peak positions, and to save them as water coordinates. If you have a license for X-BUILD or X-AUTOFIT, X-SOLVATE is automatically available from the QUANTA **Applications** menu.

This chapter describes the functions and steps that you take to use X-SOLVATE. It covers the following areas:

- Molecular coordinates used by X-SOLVATE.
- Saving water molecules on exit from X-SOLVATE.
- Accessing X-SOLVATE.
- Search for peaks palette.
- Peak search parameters dialog box.

Molecular coordinates used by X-SOLVATE

X-SOLVATE requires at least one MSF file and allows you to supply other MSF files containing coordinate data. When doing non-bonding calculations, X-SOLVATE will use the coordinates from all MSF files that were open at X-SOLVATE's start-up. New water positions will only be determined if there are no atomic coordinates from any of the MSF files within the minimum cutoff distance from an electron density peak. This means that waters that X-SOLVATE previously placed will be used in the non-bonding calculations and new waters will *not* be placed at these previously determined sites.

Saving water molecules on exit from X-SOLVATE

When you exit X-SOLVATE, the waters will be saved in one of several places, depending on the currently open MSF files.

1. If any MSF currently open and active has any water molecules (including those placed previously by X-SOLVATE), then the new water molecule will be inserted in this MSF after the last water already here. The sequence ID of the new water molecules appended to this dataset will

be started at the last sequence ID + 1 to give consecutive numbering of all the waters. The segment ID will be the same as the segment ID of the current waters in this MSF.

2. If none of the currently open and active MSF files have any water molecules, then a new MSF file, called searchwaters.msf, will be added to the molecular table and all the new water positions will be added to this file. The first sequence ID will be 1 and the new segment name will be W.

Accessing X-SOLVATE

Before you begin

Load the coordinate set of a molecule into QUANTA. Also supply an electron density map (fo-fc, 2fo-fc, or 3fo-fc) and the current symmetry for the crystal system of the original data.

You can use the routine without symmetry but, if you do, no peak checking is carried out for symmetry equivalents. This produces water molecules related by symmetry overlapping.

Note that symmetry atoms can be picked and the information displayed.

To begin X-SOLVATE

1. Select **X-Solvate** from the **Applications** menu. QUANTA saves the current selection, color, and display masks. It then generates a selection mask that contains all atoms plus their symmetry equivalents for the map currently loaded.

   The color mask is set so that the working molecule is colored by element, and the symmetry-equivalent atom positions are colored blue and the NCS atoms red.

   A three-dial set is displayed containing dials for xyz positioning of current water molecules. Additionally, the **Search for peaks** palette appears.

Using X-SOLVATE

This section describes the steps for searching, adjusting, and saving water molecules.

Searching for water molecules

A search for water molecules identifies the water peaks in the electron density map and sorts them by size. X-SOLVATE is designed to run searches in shells. The first search finds water around the protein as defined by the max-bond parameter. If all water molecules from this search are saved, a subsequent search finds the next level of water molecules further from the protein.

To run a search

1. Select **Search for waters** from the Search for peaks palette. A search is initiated for peaks that:

- Are more than the **min-bond** from the protein.
- Are less than the **max-bond** from the protein.
- Are separated by more than **min-sep** apart.
- Do not overlap symmetry equivalents.

  Peaks are returned in size-sorted order, where the first peak is the largest in the map. The molecule window changes so that the view is centered on the first peak and a 6.0-Å sphere of the map displayed. Atom contacts to the peak are shown by white dotted lines and numbers indicating distances (up to 3.5Å). The program maintains non-bond information of 6.0 Å for each water, so if you move the water more than 2.5 Å, the non-bond data will be incorrect.

2. Select **Next peak** from the Search for peaks palette to move to the next peak. The **Molecule** window display adjusts accordingly.
3. Select **Previous peak** from the Search for peaks palette to return to the peak you have just finished examining.

You can adjust the parameters for a search by selecting **Change search setting…** in the Search for peaks palette. A dialog box is displayed.

For an explanation of each of the parameters, see Peak search parameters dialog box.

Adjusting water molecule position

After examining the water molecules identified by the search, you can manipulate each of them individually to position them as you want.

1. Zoom in on the water so that you can see its position relative to the atoms around it.
2. Adjust the position of the water using the dial set that is displayed.

The non-bonds and hydrogen bonds to each water appear as white and yellow lines (color 5 and 6). The white lines represent hydrogen bonds to atoms and the yellow lines are non-bond interactions. The non-bonding and hydrogen bonds up to 3.5Å are shown.

Saving the water to MSF

On completion of checking the water site for nonbonds and fit to density, this water position can be saved. The tool **Save as water** on the Search for peaks palette will place a water molecule at the peak site. If this water molecule on subsequent analysis is required to be deleted, then the **Delete water** tool will remove the water molecule associated with this peak in the electron density.

X-SOLVATE saves the water in its current position in a currently open and active MSF file only if this file already contains water molecules. If not, then a new file searchwaters.msf is created.

You can save all water peaks that are identified in a search by selecting **Save all peaks** from the Search for peaks palette.

Search for peaks palette

This section provides quick-reference to information on the Search for peaks palette selection. This palette provides tools for searching for water peaks, examining the peaks, and saving them.

Access the Search for peaks palette by selecting **X-Solvate** from the **Applications** menu in the QUANTA menu bar.

Search for waters

Use his selection to initiate a search of the current molecule for the peaks that:

- Are more than the **min-bond** from the protein.
- Are less than the **max-bond** from the protein.
- Are separated by more than **min-sep** apart.
- Do not overlap symmetry equivalents.

Change search setting...

Use this selection to change the parameters used in the calculation. For information on parameter settings, see Peak search parameters dialog box.

Position search center

Use this selection to move the search center in the display. Place the new center of the calculations using the 3D cursor that is provided. The default center is the geometric center of the molecule.

Map on/off

Use this selection to toggle the current map display on and off.

Previous peak

Use this selection to move the display so that it is centered on the previous peak from the peaks list. This peak can then be manipulated and saved as a water molecule.

Next peak

Use this selection to move the display so that it is centered on the next peak in the peaks list. This peak can then be manipulated and saved as a water molecule.

Goto peak

Use this selection to move the display so that it is centered on a specific peak in the peaks list.

This peak can then be manipulated and saved as a water molecule.

Save as water

Use this selection to mark a peak to be saved on exit from X-SOLVATE to the .msf file at the current water position.

Delete water

Use this selection to mark a peak *not* to be saved on exit from X-SOLVATE to the .msf file at the current water position.

Save all peaks

Use this selection to mark all waters so that they are saved on exit from X-SOLVATE.

Exit

Use this selection to exit from X-SOLVATE and to save the water molecules in the .msf file.

Quit

Use this selection to exit from X-SOLVATE when you do not want to save the peaks. Previous selection, color, and display masks are restored.

Peak search parameters dialog box

This section provides quick-reference information on the Peak search parameters dialog box options. This dialog box provides parameter options peak searches.

Access the Peak search parameters dialog box by selecting **Change search setting…** on the Search for peaks palette. This dialog box allows you to change the parameters that are set for a water search.

Min density level

This is the minimum value of electron density in sigma where a water can be placed. The default value is 2.

Min distance to protein

This is the minimum distance possible for a peak to be from the protein. The default value is 2.0 Å. This resolution is less than the reasonable minimum expected in high resolution maps so that error inherent in the maps at this stage does not prevent determination of the water positions.

Max distance to protein

This is the maximum distance from the protein that a water peak is found. The default value is 5.0 Å. The setting prevents X-SOLVATE from finding peaks in a density region where no other data justifies placing one. The aim of the routine is to find the first solvation shell, then to repeat the peak search with the first solvent shell present, thus allowing the determination of further solvent shells by iteration.

Min sep between peaks

This is the minimum distance allowed between peaks. Peaks that are closer together than the min-sep value are replaced by a single peak in the average position. The default value is 1 Å. The value used here depends on your working practice; either multiple waters are placed close to each other with low occupancy, or single unit occupancy sites can be placed with minimum separation of 2.5 Å.

Calculation radius

This is the radius that describes the region to be searched by X-SOLVATE. The default value is calculated from the current loaded macromolecule so as to cover the entire molecule.

Display radius of map

This changes the map display radius about each peak. The default value is 6.0 Å.

Temperature factor

This is the BVALUE of the temperature factor saved with the water position coordinates. The default value is 20.

11. Using the CNX Interface

QUANTA 2006 includes an interface to features of the CNX program for crystallographic calculations. The main features are:

- Production of a CNX-compatible PSF and parameter file. The PSF includes structural information used by CNX in addition to that available in the coordinate file.
- Rapid computation of a variety of electron density maps, including omit maps.
- Refinement: simulated annealing; constant-temperature dynamics; rigid body, positional, and B value refinement

To access the CNX interface, go to the **Applications** menu (or enter **cnx** on the command line prompt).

In addition, two utility programs are accessible from the CNX interface or from the command line by typing **cnx util**:

- The chtoxprtf program converts an RTF file generated by QUANTA into a CNX-format RTF file.
- The psfgen program takes a parameter file and PSF file and, for the set of residues that are not named in the $QNT_XRAY/known.res file, generates unique atom types for each atom and updates the PSF and parameter files accordingly. The purpose of this is to allow expert users to change the parameterization.

For additional information on CNX, please see the CNX documentation and tutorials, which are published separately by Accelrys.

The CNX Interface palette

Using tools on the CNX Interface palette, you can set up protocols from QUANTA for CNX, generate CNX scripts, and run CNX calculations. Multiple protocols can be set up, but usually only one can be run at a time on a single host. By using different hosts, you can process multiple CNX jobs simultaneously.

QUANTA manages the files needed to execute a CNX job: an input script containing CNX commands; a reflection file (.cv); and the coordinate file (PDB), PSF file, and parameter file necessary to define the molecular system. The input script is constructed by QUANTA from template files

(maps.cnx or refine.cnx) taken from the following directories in order of priority:

- The current working directory.
- The directory pointed to by the QNT_USR environment variable.
- The QNT_XRAY environment variable.

This allows the user or group to maintain their own refinement and map calculation protocols. The template files contain all the instructions necessary for running a particular type of calculation. Lines that need to be replaced by QUANTA parameters are labeled with {Q??> } in the first 6 columns.

The initial part of the CNX script defines all the parameters and files to be used in the calculations. The interface can produce a file describing the topology of the molecule (named *filename*.xpsf) and a parameter file (named *filename*.xprm), where *filename* is taken from the name of the first msf in use in the program.

Palette tools

Set Symmetry

Space group and unit cell dimensions may be obtained by QUANTA while importing a PDB-format coordinate file containing the CRYST1 card, or may be defined by using this tool. This information is necessary for any crystallographic calculations. In using this tool, you are prompted for space group, unit cell information and orthogonalization code. The orthogonalization code defines how the crystallographic coordinate frame is related to the orthogonal coordinate frame used for atomic coordinates. The appropriate value is usually **1**.

Set-up Structure Factors

A file of observed structure factors (F_{obs}) is required for all CNX calculations. This file must be in CNX format (.cv) and include a complete header and the definition of a test set.

The **Set-up Structure Factors** tool displays a File Librarian from which to select a reflection file. You also need to define the resolution limits of the data to be used in the CNX calculations.

Parameter Set

The **Parameter Set** tool opens a dialog box for specifying which parameter set to use with CNX. The choices are the standard set of parameters, the Engh and Huber parameter set, or a combined set.

Engh and Huber developed a set of parameters for use in crystal structure refinement following an analysis of the geometry of a series of molecules in the Cambridge Data Base (Engh and Huber 1991). A set of parameters and topology files based on their work has been an integral part of CNX for some time (the parhcsdx.pro and tophcsdx.pro files). The standard parameter set is based on the CHARMm forcefield. Engh and Huber parameters are the more frequently used of the two for refinement.

Specify the desired parameter set with the **Parameter Set** tool on the CNX Interface palette. You can use QUANTA to construct PSF and parameter files that conform to either the standard or the Engh and Huber parameter set with the **Generate CNX PSF** command. If the Engh and Huber parameter set is selected, then when PSF generation is requested the following occurs:

1. The current CHARMm parameter file and QUANTA dictionary file settings are saved.
2. The current QUANTA dictionary file is changed to $HYD_LIB/ ehxplor.dic. This maps from the standard protein atom and residue names to the Engh and Huber atom types as defined in the tophcsdx.pro file.
3. The current parameter file, to be used in generating a parameter file from QUANTA, is set to the file $QNT_XRAY/EHXPL.PRM. This contains the Engh and Huber parameter set from parhcsdx.pro converted into a format acceptable to QUANTA.
4. The current set of active atoms is passed to the PSF and parameter generator. The $QNT_XRAY/eh_defs.list file is used to describe the torsions for the named residues. This overrides the generation of all possible torsions, which is the normal output of the PSF generator when generating CHARMm psf files. The parameter chooser asks you to supply any missing parameters.
5. The QUANTA dictionary file and CHARMm parameter file revert to those saved in Step 1.

In this way, the limited set of Engh and Huber parameters is used as a basis for the parameter chooser.

Map Generation

Through the CNX interface, you can specify the essential variables to be used in map generation. The program uses those values to fill in the appropriate fields in the maps.cnx file in the $QNT_XRAY directory. You can set your own protocol for these calculations by changing these files. The program first looks in the working directory for the named files, then in the $QNT_USR area, and finally in the $QNT_XRAY directory.

Clicking **Map Generation** opens a dialog box in which you can choose:

- The type of map to be calculated (**2fo-fc**, **fo-fc**, **fo**, or **fc**).
- Whether it is an *omit map*, that is, whether certain residues or a particular segment are to be excluded from the phase calculation for map generation. The symbols ^ and * are interpreted as the beginning and end of a segment.
- The grid size at which to compute the map. For some maps, CNX fails with the default value (0.33 Å) and a finer grid (e.g., 0.25 Å) should be tried.
- Whether the calculation runs in the background or interactively, or whether the interface produces a script to be run later using the **Run CNX...** command.

Tools in the dialog box, and their functions, include:

Omit Residues from Map Calculation Modify the segments and residues used in calculating a map. By default, all atoms in the molecule are selected for map calculations. This can be modified by omitting certain segments or residues.

Segment Enter a segment type.

Residues Enter a residue or residue range.

Define Extent of Map Determine map boundaries, which can be specified in four different ways, following standard CNX usage. The map is calculated to cover the specified volume.

Cover Entire Molecule Calculate a map to cover all atoms in the molecule, with an extra extension (cushion) of 2 Å.

Cover Selected Atoms Calculate a map to cover all displayed atoms, with an extra extension (cushion) of 2 Å.

Fill Unit Cell Calculate a map to fill the entire unit cell.

Fill Defined Box Calculate a map to fill a rectangular volume. The center of the box is defined as the center of rotation of the screen. The box

dimensions are arranged around that center. The x, y, and z values of each edge of the box are in Ångstroms.

Rigid Body Refinement

You provide the essential parameters to be used in rigid body refinement. The program uses them to fill in the appropriate fields in the rigid.cnx file in the $QNT_XRAY directory. You can set your own protocol for these calculations by changing these files. The program first looks in the working directory for the named files, then in the $QNT_USR area, and finally in the $QNT_XRAY directory.

Clicking the **Rigid Body Refinement** tool opens a dialog box allowing you to set up the refinement protocol. The options are:

- To perform rigid body refinement. Specify the number of steps of rigid body minimization.
- To run positional refinement. Specify the number of steps of coordinate minimization.
- To run B-value refinement. Specify the number of steps of B-value optimization.
- The type of initial B-factor correction (scaling) to be used. The choices are none, isotropic and anisotropic.
- The type of bulk solvent correction. The choices are **none**, **Babinet** or **atom mask**.
- To generate maps at the end of the refinement (the file mentioned in the first paragraph contains the specification of the maps to be generated).
- The rootname for the input file. This name is used for all the files generated in the CNX calculation (input command file, log file, coordinate file, and map output file).
- Whether to run the calculation in the background or interactively, or whether the interface produces a script to be run later using the **Run CNX...** command.

Refine Structure

You provide the essential parameters to be used in refinement. The program uses them to fill in the appropriate fields in the refine.cnx file in the $QNT_XRAY directory. You can set your own protocol for these calculations by changing these files. The program first looks in the working directory for

the named files, then in the $QNT_USR area, and finally in the $QNT_XRAY directory.

Clicking the **Refine Structure** tool opens a dialog box allowing you to set up the refinement protocol. The options are:

- To run dynamics refinement, either simulated annealing or constant-temperature dynamics. For either choice, you may specify the initial temperature and the number of refinement steps. For simulated annealing, the decrease in temperature per step is calculated from these two values. You also have the choice of torsion angle dynamics or Cartesian dynamics.
- To run positional refinement. Specify the number of steps of coordinate minimization.
- To run B value refinement. Specify the number of steps of B-value optimization.
- The type of initial B-factor correction (scaling) to be used. The choices are none, isotropic and anisotropic.
- The type of bulk solvent correction. The choices are **none**, **Babinet** or **atom mask**.
- To generate maps at the end of the refinement (the file mentioned in the first paragraph contains the specification of the maps to be generated).
- The rootname for the input file. This name is used for all the files generated in the CNX calculation (input command file, log file, coordinate file, and map output file).
- Whether to run the calculation in the background or interactively, or whether the interface produces a script to be run later using the **Run CNX...** command.

Other CNX Interface tools

Generate CNX PSF. Generate a PSF and parameter file for the currently selected molecular system.

Set CNX Host... Open a dialog box for selecting an external host on which CNX has been installed and which has been referenced in the file $HYD_LIB/applcomm.db. The $HYD_EXE/cnx.bat file contains the commands to run CNX on an external host. This file must be located on the selected host.

Run CNX... Open a dialog box for selecting an already-existing CNX script to run.

Check Log File Open the Choose CNX log file File Librarian. The selected log file is scanned for errors that may have halted or affected the CNX run and reports any errors to the textport.

Update Coordinates Open the Choose the PDB file File Librarian. After an CNX job is complete, a new set of coordinates is saved in CNX PDB file format. This tool reads those coordinates and updates the current molecular system.

CNX Utilities. Access the CNX utilities described at the beginning of this chapter.

Exit CNX Interface. Close the CNX Interface palette and return to QUANTA core functions.

Handling disorder in QUANTA and the CNX interface

The alternative location indicator is held for each atom in QUANTA as a (mostly) hidden part of the atom name. It is blank if there is no disorder. Up to six alternative positions are handled by most of the electron density fitting functionality. When a CNX-format PSF or PDB file is saved, this information is queried so that the correct information is included. In addition, PSF generation includes a valence-checking function that ignores atoms that are not disordered but are connected to disordered atoms. PSF generation correctly replicates the bonding as it appears on the screen.

12. Setting Up Molecular Systems for CNX

This chapter briefly describes how to set up some common types of molecular systems in QUANTA for use in CNX calculations. These examples can serve as templates for setting up your own work.

This chapter describes

Setting up a CNX system for a simple protein

Setting up an CNX system for a protein, solvent, ion, and nucleic acid (see Setting up a system for a nucleic acid, solvent, ion, and ligand).

Before you begin

To complete these exercises, copy these files from the directory $QNT_ROOT/tutorial/files into your working directory:

- Barnase protein system: 1a2p.pdb and 1a2p.cv
- DNA, solvent, and ligands: 102d.pdb and 102d.cv

Setting up a CNX system for a simple protein

This section describes how to set up a simple protein system and perform CNX calculations on it.

Setting up the structure for CNX

1. From the **File** menu, select **Import**. The Import File Librarian is displayed
2. Select the Protein Data Bank PDB file 1a2p.pdb to import.
3. This protein has some residues with multiple conformations (disorder). You will be asked if you wish to “Handle Alternate Location as Disorder?” Click **Yes**. Unless you have turned on the preference that causes the CRYST card to be read automatically, you are prompted whether you want to set up symmetry. Click **Yes**.

4. You are then asked to confirm that the spacegroup P32 is correct. Click **Yes**. (The reason for these prompts is that some third-party programs that output PDB files do not write the CRYST line correctly, so it is not always safe to read it blindly.)
5. The file 1a2p.msf is created and opened.
6. If you import a PDB or other format file that does not contain correct cell and spacegroup information, this can be added to the MSF using the menu function **Edit | Symmetry->Define**.
7. From the **Edit** menu, select **Split and Clean**. You should see the molecule colored according to whether it is protein, solvent, nucleic acid, or other. Note the three red atoms. These are zinc ions. Click **Save to separate MSFs**. You are warned that "Bonding exists between different classes." This warning is triggered because the zinc ions were assumed to be covalently bonded to adjacent protein residues. Click **Continue** to break these bonds. You now have three MSF files each containing one sort of molecule, enabling the most appropriate tools to be easily applied to each for correcting the structure.
8. Click **Clean Options**. This allows control of the hydrogen model to be used. Choose **None** and click **OK**.
9. Click **Clean Protein**. This adds hydrogens when that option is chosen, sets up standard names and atom types, and rebuilds any sidechains that have missing atoms. You are asked about the disorder in the protein. Choose to **Keep all Disorder** and click **OK**.
10. Click **Clean Solvent**. This command serves a similar function for solvent as **Clean Protein** does for protein.
11. Ligand molecules would normally fall within the class "Other". The Molecular Editor is available from the Split and Clean palette to fix bond order or any other details of ligand molecular structure. In this case, the zinc ions are the only member of the "Other" class, and do not require additional attention.
12. Click **Finish** to exit the **Split and Clean** function.

Running a CNX job to minimize the system

This section describes the steps for minimizing the protein using CNX:

1. From the **Applications** menu select **CNX Interface**. The CNX Interface palette is displayed.
2. From this palette, select **Parameter Set...**. Select **Use Engh and Huber parameters** and click **OK**.

3. From this palette, select **Set-up Structure Factors...**. The Select Intensities file librarian is displayed listing the structure factor files.
4. Select the intensity file 1a2p.cv. Set the **High Resolution Limit** to **1.5**. Click **Open**.
5. Select the intensity file dummy.fob; make sure the weight factor **WA** is set to **0**. Leave the other fields at their default values and click **Open**.
6. Select the Set **CNX Host...** tool. Select a local host from the scrolling list and click **OK**. The CNX Job Characteristics dialog is displayed; accept the default values and click **OK**.
7. Select the **Refine Structure...** tool. The Positional Refinement Settings dialog box is displayed. In this dialog box, of the three types of refinement that you can perform, select only **Perform Minimization Refinement**, with **Number of Steps** set to **20**. Select **Initial B-factor Correction**: **Anisotropic** and **Bulk Solvent Correction: Atom Mask**. Select the action at the bottom of the page; **Wait for Calculation to complete**. Leave all other fields at their default values and click **OK**.
8. The CNX job runs on the selected host from the file cnx.bat. You will be notified in the Textport when cnx.bat is started and when it is finished.
9. Select **Update Coordinates...**. Select the refined coordinate file, 1a2p_Protein_refine.pdb. Click **Open**. The three zinc ions need new segment names. Choose 'D' when prompted regarding this issue for the first ion. Use 'E' and 'F' for the second and third ion, respectively, when prompted. In the 'Save data into msf files' dialog box, all three molecules should be checked (default). Choose to **Create a new generation of the file** and click **OK**. The coordinates on the screen and in the active MSFs are updated. If you watch the screen carefully you can see the atomic positions shift as the coordinates are updated.

Several files should be generated from this example:

- 1a2p_Protein_refine.inp - the script that is sent to X-PLORCNX.
- 1a2p_Protein_input.pdb - the initial coordinates of the system.
- 1a2p_Protein_refine.out - the CNX logfile.
- 1a2p_Protein_refine.pdb - the final coordinates.
- 1a2p_Protein.xprm - the CNX parameters file.
- 1a2p_Protein.xpsf - the CNX PSF file.

Running standalone CNX

The PSF (.xpsf) and parameters files (*.xprm) from Quanta, can be used with other CNX scripts, when combined with the input pdb file that is written by the CNX interface (*_input.pdb). Frequently, a topology file and a parameter file are needed to define a ligand for standalone CNX. Quanta can be very useful in this case. The Molecular Editor will write a CHARMm-style RTF file if requested in the dialog that occurs upon exit. That RTF can be converted to a CNX-style topology file (*.top) by using the appropriate command in the CNX Interface/CNX Utilities.. That topology file, combined with the parameter file (*.xprm) that the CNX Interface writes, supply the ligand information needed for refinement in standalone CNX.

Dealing with possible problems

You could encounter several problems when running this sort of calculation. This section discusses three common ones.

If CNX does not run or cannot be found

QUANTA probably cannot find where the executable for CNX is located. Check with your systems administrator or check your installation guide for the locations of cnx.bat and cnx.exe. cnx.bat should be in the $HYD_EXE directory, and it should be referenced in the $HYD_LIB/applcomm.db file for any machines on which you want to run CNX. The cnx.bat file is a simple C-shell script that executes cnx.exe.

Seeing the attached-atoms error message

This message box is displayed during PSF generation if any atoms have too many attached atoms. This could happen if the geometry is poor and the distance-search bonding algorithm is used.

This problem should not occur for a protein-only system, since the Special Protein algorithm should be used automatically.

Using the **Edit | Bond Options** function, you can also disable bonding between different segments. This is useful if you have solvent in close proximity to other structures within the same MSF.

Bad bonding can occur if connectivity has been incorrectly specified in an external file. In such cases, switch to an algorithm other than **Stored Connectivity plus Distance Search**. If bad bonds persist, manual editing is straightforward: the message box allows you to color the problematic bonds red and the rest of the molecule white. Choose this and focus in on

the trouble spots. Use the **Break Bond** tool on the Modeling palette to eliminate the problem bonds. Then you can use **File | Save As** to make sure the corrected connectivity is saved for future use.

Unable to generate a PSF because of unknown atom types

"Unknown atom types" refers to the situation where one or more atoms have not been assigned a valid CHARMm atom type. This should never happen for a protein that has been typed on import using the correct dictionary file or cleaned with the Split and Clean application. It can occur for non-standard groups. Methods for correcting the problem include:

- Use **Apply Dictionary** to retype using a different dictionary file. You can use your own customized dictionary files.
- If RTF descriptions exist for all the residues in your system, perform a CHARMm energy calculation in RTF mode.
- Isolate the problem by placing the parts that work into a separate MSF. This can conveniently be done with the Split and Clean application. Then use the Molecular Editor to correct any problems.

Setting up a system for a nucleic acid, solvent, ion, and ligand

This section presents an example of how to:

- Set up a system containing nucleic acid, solvent, ion, and ligand.
- Set up CNX to perform a minimization and generate an electron density map for this system.

Doing initial setup

Import the pdb coordinate file 102d.pdb, and proceed with Setting up the structure for CNX (page 261) as for the previous example up to and including Step 6.

Preparing the other molecules

1. From the **Edit** menu, select **Split and Clean**. You should see the molecule colored according to whether each molecule is solvent, nucleic acid, or other. Click **Save to Separate MSFs**.
2. Click **Clean Options**. This allows control of the hydrogen model to be used. Choose **Polar** and click **OK**.

3. Within the Split and Clean application, click **Clean Solvent**. This adds hydrogens to the solvent and corrects the atom types if necessary. For the purposes of crystallographic refinement, ideally the X-SOLVATE application should be used to place water molecules in the electron density. The CHARMm **HBUILD** function is also useful, since it checks for hydrogen-bonding partners as it places the hydrogens.
4. Click **Clean Nucleics**. This adds hydrogens and corrects atom types in the DNA.
5. Setting up a ligand generally requires manual intervention to obtain the best chemical representation. If the initial structure has very good geometry, then the **Auto Edit** tool on the Split and Clean palette may be able to perform the tasks automatically. In this case, click **Molecular Editor** and choose **102d_Other.msf**. Select the **Change Bond** tool. A total of 16 bonds should be changed from single bonds. At both ends of the molecule, click each bond in the two rings three times to make these systems aromatic. Each bond should have bond order of 1.5 as indicated by double lines, one of which is solid and one of which is dashed. Also, set the bond between each of the four nitrogen atoms and its adjacent carbon atom to have bond order 1.5 (dashed and solid bonds). Click **Add Polar Hydrogens**.
6. Click **Save and Exit**. In the Save Options palette that appears, choose **Reassign atom types** and **Reassign atom charge**. Click **OK**. The Adjust Total Charge palette now opens. Enter the desired total charge as **2.04** and select Do not smooth. Click **OK**.
7. Click **Finish** to exit the Split and Clean application. To check that the structures are correctly set up, perform a PSF mode CHARMm calculation. The missing parameter box that opens indicates those parameters that were estimated within QUANTA. Quit this tool to finish the energy calculation.

Setting up the CNX calculation

This section describes the steps for minimizing the protein using CNX.

1. From the **Applications** menu select **CNX Interface**. The CNX interface palette is displayed.
2. From this palette, select **Parameter Set...**. Select **Standard Parameters** and click **OK**.
3. From this palette, select **Set-up Structure Factors**. The Select Intensities file librarian is displayed listing the structure factor files.
4. Select the reflection file 102d.cv. Set the **High Resolution Limit** to **2.2**. Click **Open**.

5. Select **Generate CNX PSF** to write the necessary structure and parameter files for your molecular system.
6. Select the Set **CNX Host...** tool. Select a local host from the scrolling list and click **OK**. The CNX Job Characteristics dialog is displayed; accept the default values and click **OK**.
7. Select the **Refine Structure...** tool. The Positional Refinement Settings dialog box is displayed. In this dialog box, of the three types of refinement that you can perform, select only **Perform Minimization Refinement**, with **Number of Steps** set to **20**. Select **Initial B-factor Correction**: **Anisotropic** and **Bulk Solvent Correction**: **Atom Mask**. Select **Generate maps post-refinement**. Select the action at the bottom of the page; **Wait for Calculation to complete**. Leave all other fields at their default values and click **OK**.
8. The CNX job runs on the selected host from the file cnx.bat. You will be notified in the Textport when cnx.bat is started and when it is finished.
9. Select **Update Coordinates...** Select the refined coordinate file, 102d_Nucleic_refine.pdb. Click **Open**. In the Save data into msf files dialog box, all three molecules should be checked (default). Choose to **Create a new generation of the file** and click **OK**. The coordinates on the screen and in the active MSFs are updated. If you watch the screen carefully you can see the atomic positions shift as the coordinates are updated.

Several files should be generated from this example:

- 102d_Solvent_refine.inp - Contains the script that is sent to CNX.
- 102d_Solvent _input.pdb - Contains the initial coordinates of the system.
- 102d_Solvent _refine.out - the CNX logfile.
- 102d_Solvent _refine.pdb - the final coordinates.
- 102d_Solvent.xprm - CNX parameters file.
- 102d_Solvent.xpsf - the CNX PSF file.
- 102d_Solvent_refine_2fofc.map - the 2fofc map calculated after refinement
- 102d_Solvent_refine_fofc.map - the difference map calculated after refinement.

13. Tutorial

Lesson 1: Getting Started with QUANTA (page 269)
Lesson 2: Getting Started with the X-Ray Modules (page 276)
Lesson 3: Manual CA-tracing (page 284)
Lesson 4: Automated CA-tracing (page 294)
Lesson 5: Sequence assignment (page 303)
Lesson 6: Model Building and Real Space Refinement using X-BUILD (page 307)
Lesson 7: Ligand Fitting using X-LIGAND (page 318)
Lesson 8: Water Fitting Using X-SOLVATE (page 323)

Setup for the tutorial

Before beginning, unpack the QUANTA tutorials tar file.

Extract the files needed for the QUANTA tutorials to an appropriate location with the command:

```
>  tar -xvf QUANTA_tutorials.tar
```

Lesson 1: Getting Started with QUANTA

The following material is covered in this lesson:

- Starting QUANTA,
- Loading in a molecule,
- Manipulation of a molecule on the screen,
- Exporting files, and
- Exiting QUANTA.

1. Set up for the tutorial

If you have not extracted the tutorial files from the QUANTA tutorials tar file, see Setup for the tutorial (page 269).

Move into the directory lesson3/.

It is best to start QUANTA in a proper terminal window or winterm. It is not recommended to start the program from the Console window.

At the UNIX prompt, type:

```
> quanta
```

followed by <Enter>.

As the program starts up you are asked for permission to copy some startup files to the directory that QUANTA is running in. Click **Yes** or press enter.

After a few moments you will now see a number of QUANTA windows appear (Figure 2.).

- **QUANTA Viewing Window**: The molecular structure under study is displayed in the viewing area. It is empty at present because no molecules have been read in.

  Menu Bar: The menu bar contains nine menus from which various QUANTA functions and external programs can be accessed. Each entry picked and dragged with the mouse gives a commands menu from which various operations can be selected. In general, as you move from left to right on the menu bar, you will move from housekeeping functions through structure creation and editing functions, to simulation and analysis functions.

  Command Prompt: You can enter specific QUANTA commands on the Command Line located directly under the viewing area.

  Message Area: The Message Line is located directly under the Command Line. During QUANTA operations, this area displays instructions and error messages.

Textport: A scrollable, resizable window that displays QUANTA operational messages and information about displayed structures. Messages can be the result of a selected function. For example, a message may list the atoms contained in the displayed structure, confirm picking of an atom in the viewing area, or transmit an error message.

- **Palettes**: Commands can be picked from the Geometry and Modeling palettes.
- **Molecule Management Table**: The various cells display statistics of the molecule(s) under observation and turn the display on and off.
- **Dial Emulator**: It is an electronic mouse equivalent of mechanically operated buttons or dials.

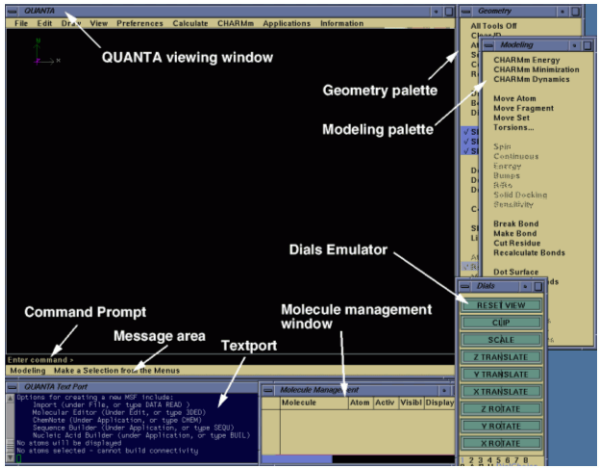


Figure 2. The layout of QUANTA

2. Importing a molecule

Go to the **File** pulldown menu located on the menu bar and drag down to the **Import Single Structure** command and then release the button to select it. Click on the file you want to import (kringle4.pdb), (it will become highlighted) and then click **Import**.

You should now have an image of the protein with the atoms colored by element. The atoms are joined together by lines that represent covalent bonds. Regardless of how a structure is created, information about it is stored in

QUANTA as a molecular structure file or MSF. File names are written as filename.msf. An MSF contains at least this basic information:

- All atoms in the structure.
- Location of atoms in 3D coordinate space.
- Atom types, names, and charges.
- Residue and segment information.

3. Manipulation of the molecule

Dial emulators that control vector quantities like translation, rotation, and scaling are affected by the position of the cursor over the emulator dials.

For example, if you click on the right side of the Z-rotation emulator dial using any mouse button, the structure in the viewing area rotates clockwise. Pressing and holding the button while dragging the mouse to the right results in continuous clockwise rotation that increases in speed as the cursor moves farther right. If the molecule gets lost then pick the dial bar called **Reset View**.

A much more natural method for manipulating the molecule is to use a combination of keyboard keys and mouse buttons as a virtual track ball (summarized in Figure 3.).

Move the mouse so the cursor is in the viewing area. Press and hold the middle mouse button and move the mouse so the cursor moves around the viewing area.

The structure rotates at the same speed and in the same direction as the mouse. The axes, displayed in the upper-left corner of the viewing area, rotate with the structure. Release the mouse button. The structure and axes stop rotating.

Now we will rotate about the screen Z-axis. This axis is perpendicular to the plane of the screen.

Press and hold the right mouse button. Move the mouse so the cursor moves to the left side of the viewing area. While continuing to press and hold the right mouse button, move the mouse so the cursor moves to the right side of the viewing area.

We can move the structure through a clipping zone.

Press the shift key and hold the left mouse button. Move the mouse so the cursor moves towards the top of the viewing area. You have narrowed the depth of the zone along the Z axis where parts of your structure are visible. Move the mouse so the cursor moves towards the bottom of the viewing area. You have increased the depth of the zone along the Z axis where parts of your structure are visible.

We can also translate the displayed structure on the XY axes.

Press the <Shift> key and hold the middle mouse button simultaneously.

A translation along the Z axis is also possible.

Press the <Shift> key and hold the right mouse button simultaneously.

You can scale the structure as well.

Press the <Shift> key and don't press any mouse buttons. Move the mouse so the cursor moves towards the bottom of the viewing area. The molecule enlarges as the viewer appears to come closer to the molecule. Move the cursor to the top of window and watch the structure become smaller.

You can easily display information about atoms.

Many QUANTA operations require that an atom or set of atoms be picked before processing. When an atom is picked, an identification tag (ID) is displayed on the screen adjacent to the picked atom. Place the point of the cursor over an atom. Click the left mouse button. The atom ID is displayed next to the atom, and information about the atom is displayed in the Textport.

Now, we will return the structure to its original orientation.

Select the **Reset View** dial on the **Dial Emulator**. The structure returns to its original orientation. **Reset View** in the Geometry pallet and the **Reset View** dial perform the same function.

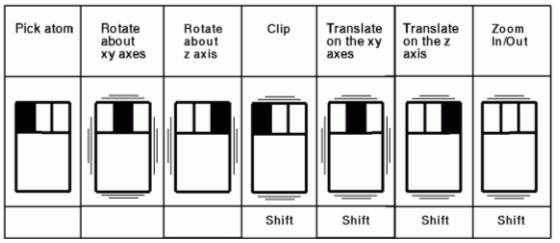


Figure 3.

4. Setting the origin of rotation

To change the origin of rotation click the **Set Origin** tool on the Geometry palette (you may need to bring this palette to the front by clicking on the side on the border). The command will be highlighted and ticked, and a message will appear prompting you to pick an atom. Click on an atom in the protein. The protein will then be re-centered about this atom and the atoms coded name (ID) will appear in green. Click on **Clear ID** on the Geometry palette to remove the atom name from the display.

5. Exporting files

To export files from QUANTA go to the **File** pulldown menu located on the menu bar and drag down to the **Export**. You can select the format required in the **Data Format** box (e.g. CHARMm CRD - you may have to scroll down using the blue/black scroll bar) and enter a file name to be saved in the **Filename** box before clicking on **Export** or hitting enter.

With the CRD format, you are prompted to enter a title or comment for the file. Enter a blank line to complete the title.

6. Accessing Help

The QUANTA manuals and various tutorials are accessible from the menu function **Information/Help....** In addition, there is interactive help for some palette tools and top-menu bar items that can be turned on with the command SET HELP. This can be optionally accessed by clicking the palette or menu with the right mouse button.

By default, online help will launch a copy of the Netscape browser. It is therefore necessary for the command “netscape” to exist in your path.

You can verify that you can use Netscape by using the UNIX command:

```
> which netscape
```

Alternatively, help text can be displayed in the textport. This can be specified using the command SET HELP, but it happens automatically if Netscape cannot be started. Only text is printed to the textport - HTML formatting, graphics, tables etc. will of course be lost.

7. Exiting QUANTA

To exit from QUANTA go to the **File** pulldown menu located on the menu bar and drag down to **Exit QUANTA** or by typing **end** in the command line.

Lesson 2: Getting Started with the X-Ray Modules

In this tutorial, you will learn to manipulate the display and positioning of the main view in QUANTA. The context will be using X-Ray crystal data, a structure with an electron density map.

1. Set up for the tutorial

For this tutorial, the best molecular replacement solution (rigid_k4.pdb) has been used to calculate a 2Fo-Fc electron density map at 1.9 Å (2Fo-Fc_k4.mbk). You will be viewing this map along with the kringle structure that was used to calculate the phase information.

Move into the directory lesson4/which contains the necessary files for this tutorial.

Start QUANTA.

Select **Applications/X-AUTOFIT**. This starts the application.

The main QUANTA graphics window will contain a molecule kringle and a contoured section of map at 1.0 and 2.0.

The following palettes and tables appear:

- X-AUTOFIT:X-BUILD,
- 3D Text,
- Pointer,
- Ramachandran plot, and
- Object Management Window.

While the following pallets or tables remain visible:

- Main QUANTA window,
- Molecule Management Table, and
- Geometry.

2. Symmetry

Click on **Symmetry…** on the X-AUTOFIT:X-BUILD palette. The Symmetry palette appears as the front window.

This palette allows the definition of symmetry and NCS. A number of symmetry objects can be created using tools on this palette. In this case, the symmetry has already been defined.

Click on **Unit cell** on the Symmetry palette.

This tool will generate a unit cell box as an object (UNIT CELL). The box can be removed again by deleting the object from the Object Management Table.

Click on **CA Packing Diagram** on the Symmetry palette.

This tool generates a number of symmetry copies (shown as Cα traces) of the current visible and active molecules in the Main QUANTA window. Each molecule is drawn a different color, using colors 1 to 14, in cycles of 14. Each symmetry copy is created as a separate object that can be turned on/off using the Object Management Table.

The symmetry objects can be deleted simultaneously by selecting the **Delete Sym. Ob.** tool on the Symmetry palette.

There are a number of related tools within the Symmetry palette. Try to find out what each tool does. Before moving on, make sure that all the symmetry objects have been deleted.

To close the Symmetry palette click on **Hide** this menu.

3. Moving the pointer

The default appearance of the pointer is a white tetrahedron in the molecule window. The pointer can be moved in several ways.

- The **Dial Emulators**: the pointer can be translated by controlling the position of the cursor over the dials (if necessary, refer back to lesson 3 here).

Note: The dials can be variously used for the pointer, masks, building or the interactive contouring option. An understanding of these is integral to using the software efficiently and important aspects shall be covered during subsequent tutorials. On accessing the X-AUTOFIT:XBUILD palette the default dials palette allows the pointer to be moved.

- The **Virtual Track Ball Option**: the pointer can be moved by using a combination of keyboard keys and mouse buttons (summarized in Figure 1).

Move the mouse so the cursor moves inside the viewing area. Press the <Ctrl> and <Shift> keys and don't press any mouse buttons. Move the mouse to translate the pointer in the X and Y axes. Press the <Ctrl> and <Shift> key and press the left mouse button. Move the mouse left and right to translate the pointer in the Z axis.

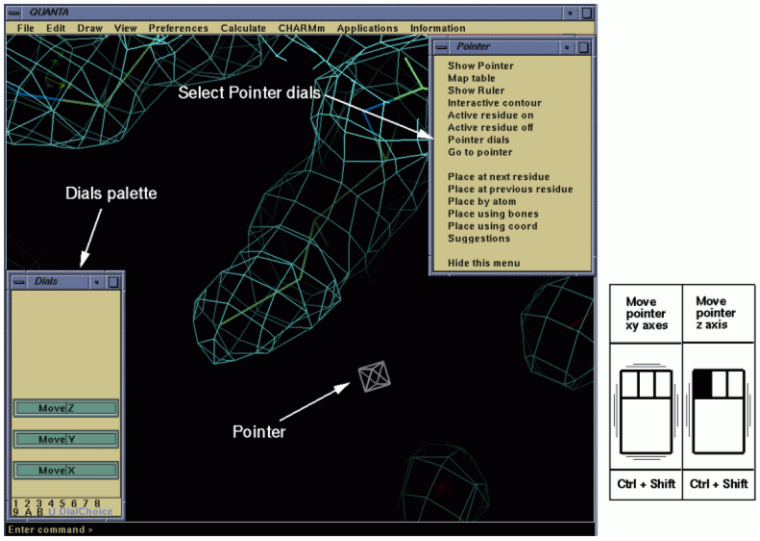


Figure 4.

4. Selecting an atom

To pick an atom move the mouse to an atom position and click the left mouse button. This identifies the atom. If you click a symmetry atom (colored blue for real symmetry or colored red for NCS symmetry), the label appears green and be preceded by a “#” symbol.

5. The pointer palette

There are several methods of changing the view center in the X-Ray applications. In all cases, you will find that the symmetry atoms and map (and bones, if on) are recalculated for the region of interest.

> Make sure you can see the Pointer palette. If you cannot see the palette then click on the top bar of the X-AUTOFIT:X-BUILD palette to move this to the front, and click on the **Pointer...** tool: the palette will appear. Select **Go to pointer**.

This moves the screen center to the current pointer position.

Click **Place at next residue**.

This command varies according to the current residue. It moves the view to the next residue in the sequence (except where the step value has been changed). You can also change the current residue if you click on the Ramachandran plot or use **Place by coord** or **Place by atom**.

Click **Place by atom**.

This prompts for an atom name.

Enter the following:

Segment name: A
Sequence ID: 35
Atom name: CA

Then click on the **OK** button.

Click **Place using coord**.

This waits for an atom pick using the left mouse button while pointing at an atom. Look at the bottom of the graphics window at the message line. Pick atom is displayed, and the **Place using coord** tool remains highlighted until you make a pick.

Note: If you miss an atom when you click, nothing happens.

6. The interactive contouring palette

The interactive contour tool allows you to change the level at which an electron density is contoured.

Click **Interactive contour** on the Pointer palette. The Interactive contour palette appears as the front window.

As soon as the routine is entered only one contour level of one map is displayed regardless of the display state before entering the routine. The default is to show the first contour of the first map initially. In this case, the Next Map and Previous Map tools are grayed out since only one map has been loaded.

The contour level is controlled using the Dials palette or a combination of keyboard keys and mouse buttons: the virtual track ball option (Figure 2).

Move the mouse so the cursor moves into the viewing area. Press the control and shift keys and don't press any mouse buttons. Move the mouse to left and right to change the current contour level of the map. Click **Next Contour** and the display will change to the next contour in the list for the current map. The **Accept** tool will make the changes to the stored values for contouring and exit. The **Quit** tool will reject any changes and return to the original display.

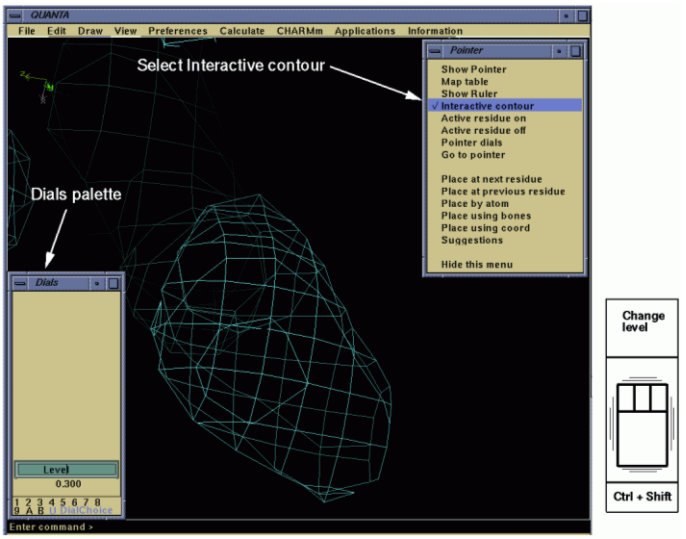


Figure 5.

7. The Ramachandran plot

You can also change the center of view using the Ramachandran plot palette which is located at the bottom right of the screen. It is currently quite small, so make it bigger by picking up a corner with the mouse and dragging. (The non-glycine residues are shown green, and glycine residues are shown blue, and the current residue (if defined) is shown red.)

Pick a blue circle with the mouse. The blue circle turns red, as this is now the current residue, and the molecular display has changed to center the glycine you picked.

8. The text palette

The Text palette controls the creation and use of annotations associated with points in the structure. These text marks allow you to label parts of the model for future reference and also change the center of view to a specific

text position. You will become very familiar with these during the subsequent tutorials.

If the 3D Text palette is not visible click **Text…** on the X-AUTOFIT:X-BUILD palette. Text strings have already added in this tutorial. You can get to the first text string by picking the tool **Goto defined text**. A sorted list is provided. Since the first text string {TEXT STRING 1} is already highlighted click **Goto**. This takes you to a yellow text mark {TEXT STRING 1}. Now pick **Goto next text**, and you will move to another text mark {TEXT STRING 2}.

You can add your own text mark.

Move the pointer using the virtual track ball option somewhere else in the structure, or use one of the other placement options described above (the pointer will be reset to the new screen center). Now select the tool on the 3D Text palette **New text at pointer**. You will be prompted for a text string. Type in a string of characters, and hit **OK**. The new text string will appear. You can go to the various text string in the structure using either of the tools Goto next text, **Goto previous text** or **Goto defined text**.

Try using the **Load property** tool on the 3D Text palette.

9. Controlling the behavior of the program

Select the **Options...** tool on the X-AUTOFIT:X-BUILD palette.

You will find various options that control the functionality of the X-Ray applications.

Change the symmetry radius to 50 and click **OK**. A large selection of symmetry atoms is displayed after about 10 seconds. Now change the value back to 6.

For all the subsequent tutorials, all these parameters will be set to sensible values for the particular tutorial.

10. End Session

To end your session click on **Finish** on the X-AUTOFIT:X-BUILD palette and go to the File pulldown menu. Select **Exit QUANTA** from the menu.

Lesson 3: Manual CA-tracing

Cα-tracing may be done manually using X-AUTOFIT. This tutorial involves the use of the electron density map (2.4 Å) of the protein RNase Sa (1, 2) solved in 1987. This map represents the exact data used to solve the protein structure and forms the basis of the original publication (rnase.mbk). During this session, you will be tracing the a-carbons of a monomer unit in the homodimer RNase.

In this tutorial the following material will be covered:

- Skeletonization of an electron density map, and
- Fitting a Cα trace to a map.

1. Start X-Autofit

Move into the directory lesson5/ and start QUANTA.

Select **Applications/X-AUTOFIT**. The program opens the main X-AUTOFIT:X-BUILD palette, the Object Management palette, the Molecule Management palette and a window labeled **CA angle/torsion**. A skeletonized view of the electron density map is displayed and two items of text are present.

2. Set up for the tutorial

From the **File** pulldown menu, select the **Replay Session** pullright menu. Then select the **Start** command. Highlight rnase.rec and click Open. This runs a script that sets up a map mask and opens the Maps Management table. The map rnase.mbk is already loaded and has been contoured at 0.30 e Å-3 (which is 1.5). To toggle the map on and off, click on the level box in the Map Management table that currently reads 0.30.

3. Bones generation

The bones have already been calculated and are displayed. To toggle the bones on and off, you can click the tool **X-AUTOFIT:X-BUILD/ Bones.../Bones on-off.** Turn off the map with the Map Management table. Currently, just the main chain bones are displayed (colored yellow). To turn on the side chain bones (colored white) by picking **Side chains on-off**.

Examine the bones to determine how to adjust the Start value. The Start value is a cutoff parameter (in) that you can adjust. All density below the defined value is removed from an electron density map and ignored when a bones skeleton is generated. If the bones appear as spaghetti (Figure 1A), the Start value needs to be increased until individual segments become evident (Figure 1B). If no connected bones are visible, or if they appear as single dots, the Start value needs to be decreased (Figure 1C).

To change the value, click **Change start value** on the Bones palette. Experiment with different values.

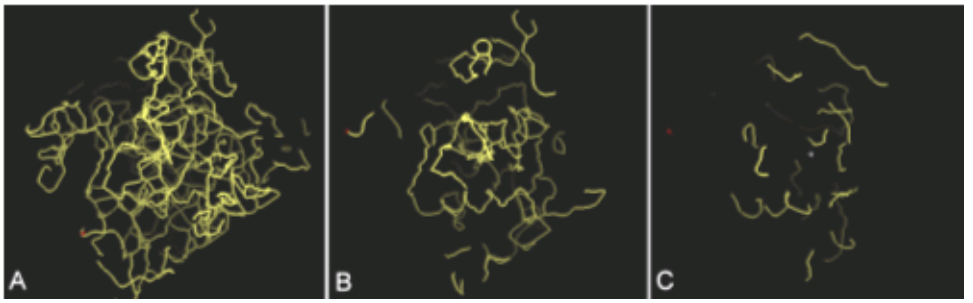


Figure 6.

Before proceeding make sure that the Start value is set to 1.3σ.

4. Adding a helix

Open the CA Build palette by clicking **X-AUTOFIT:X-BUILD/CA Build…**.

There are 2 methods of adding carbon atoms in X-AUTOFIT, you can either add predefined secondary structure elements such as a helix or sheet,

or add carbon atoms one at a time using the semi-automated fitting algorithm.

You should turn off the map so that only the bones are visible. Rotate the image until you can identify the section of bones that can be recognized as a helix - this is also labeled as an item of text Helix (11 residues).

Use the tool **Add helix or strand** on the CA Build palette to add a helix. A small dialog box will appear that allows you to add a helix/sheet of defined length. You should change the value to 11. A white image of a helix will appear, as well as another small palette Accept/Quit. You will not be able to use any other tool in X-AUTOFIT while this Accept/Quit palette is open.

The **Accept** tool from this will take the current coordinates defined by the white movable helix and create a new segment from this. The **Quit** tool will return everything as it was before. Now use the virtual trackball to move the white helix. You will find it initially easier to have the map turned off, and just fit the helix to the bones.

You need not get a perfect fit the first time as moving the helix in 3D can be difficult. It is often easiest to just adjust the Z-rotation, when the picture has been orientated so that the helix is in the plane of the screen. (You will not need to change the x/y/z position). Make sure that the Cα positions of the helix align with the branches of the bones as these represent the side chain positions. You will find that moving the structure while the map is displayed can be rather slow. So when making large changes, for example when adding a helix, turn off the map, and at the final stages turn the map back on. Look at the map for the final adjustment of the Cα trace.

When you are satisfied with the position of the helix select **Accept** from the **Accept/Quit** menu.

After you have accepted the helix your main QUANTA window should look something like Figure 2. The white helix is now colored red (indicating that it is the current segment) and has been created as an object (Cα

TRACE), which can be turned on/off from within the Object Management table.

To improve the position of the helix you can use the tool **Refine current seg** on the CA Build palette. You can abort the refinement by clicking anywhere in the graphics window if things are obviously going wrong. You can also manually adjust the position by clicking on the tool **CA build/Move current seg** and use the virtual track ball.

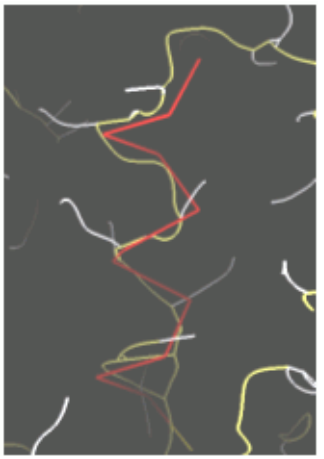


Figure 7.

5. The current Cα segment and atom

There are three types of CA-trace atoms: (1) an active Cα atom (colored yellow) in (2) an active Cα trace segment (colored red), and (3) other CA-trace segments (colored green) placed previously.

The active Cα trace atom (and active segment) is set using the **Current res. and seg.** tool on the CA Build palette. After selecting this menu item you are prompted to pick a Cα atom. Select different atoms on the helix you have just built to get a feel for this tool.

6. Adding and Moving Cα atoms

Cα atoms are added to existing segments by clicking on the tool **Next CA** on the CA Build palette. After checking that either of the terminal atoms of the helix you have just built is the active atom try to add a new atom. You should re-display the map for this.

The new atom is placed in the best possible position as evaluated by X-AUTOFIT. You can, however override this positioning by moving the active Cα trace atom (colored yellow) using one of the mechanisms listed below. Try each method in turn.

Note: Before trying these different mechanisms ensure that the **CA Dials** are active by clicking on the tool **CA build/Show CA dials**.

- By picking a point on the skeleton (active Cα will “move” towards this).
- Using the dials to set the angle (beta) and torsion (gamma) or the virtual track ball (The virtual track ball motion is by far the most efficient method) (Figure 3).
- By making an alternative auto-placement (**CA Build/Guess next CA**).

Clicking on any position within the pseudo-Ramachandran plot. This tool cannot be properly used until the 4th Cα has been placed (see following section of this tutorial).

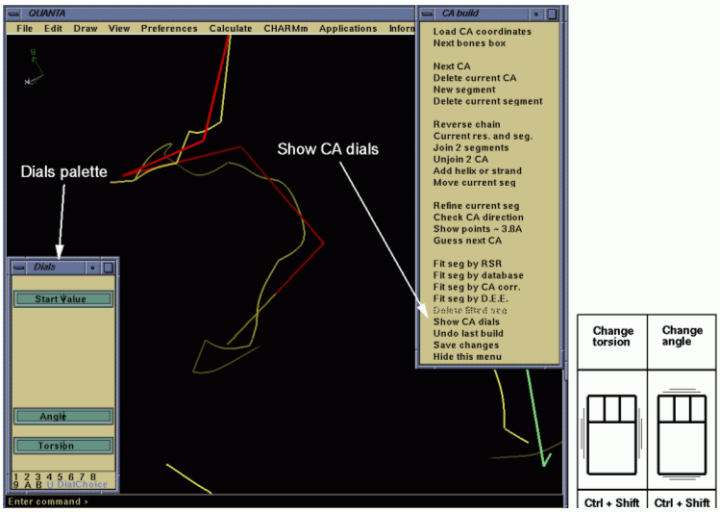


Figure 8.

Note: Because different parts of the program assume the default mode of the dials, it is sometimes necessary to tell X-AUTOFIT which dials you wish to use. From the X-BUILD:X-AUTOFIT palette you can select the following sets of dials:

- Pointer/Pointer Dials - The x/y/x pointer
- Pointer/Interactive contour - Change the map contouring
- CA Build/CA Dials - Change the angle and torsion when CA-tracing
- Mask/Mask Dials - Interactively edit the mask (see following tutorial)

Currently, you should be using the CA Dials mode.

7. Using the pseudo-Ramachandran plot

A pseudo-Ramachandran plot is generated for alpha-carbon geometry using well-resolved protein structures - this plot defines the probabilities of specific Cα geometries. The torsion angle of four consecutive alpha carbons is plotted against the open angle that is defined by three consecutive alpha carbon atoms (Figure 9.A). A probability map of alpha-carbon geometry can then be generated. The resulting plot shows the probability of the alpha-carbon geometry being restricted to certain regions (Figure 9.B). Different areas on the plot correspond to specific conformations of the protein

backbone, including alpha helices and beta sheets. This surface is used to direct fitting of alpha carbons to the displayed electron density pattern. A pointer indicates the current value of the open angle and torsion angle. The plot provides an interactive tool for monitoring and manipulating the conformation of the alpha-carbon trace. The pseudo-Ramachandran plot is displayed in an independent window labeled **CA angle/torsion**.

> Clicking on any position within the pseudo-Ramachandran plot changes the position of the active Cα trace atom (colored yellow).

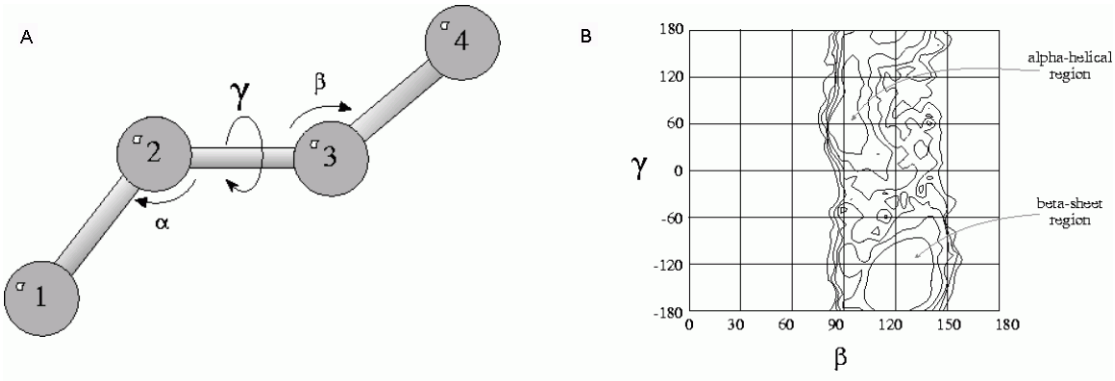


Figure 9.

8. Re-centering the bones and map

The bones are only calculated inside the volume of electron density that is displayed. When you reach the edge of this you will need to re-contour the map by centering on the current Cα to continue tracing.

> The tool **Next bones box** recalculates the bones around the current Cα atom.

9. Reversing the chain

You can change the direction that the chain is being built by clicking on the tool **CA Build/Reverse** chain tool. After this, you often need to re-center the view by clicking on the tool **Next bones box**.

10. Extending the helix

Extend the helix as far as you can either way until you reach a region of electron density that is uninterpretable. Use the tool **Check CA direction** (tool 14) to test the direction of the trace and how well the building has been carried out. Make sure that the maps/bones is covering this Cα trace.

If you get two zero values, the build was not good (or you have not moved the map to this position). If you get at least one non zero value, then you have certainly made a reasonable attempt at the tracing of the chain.

11. Building a beta sheet

When building the beta sheet, it is easier to use the semi-automated map fitting routines rather than adding a strand of defined length. This is because the strands within protein structure are not typical in shape.

First, re-display the map centering on the item of text beta sheet (6 strands). To do this go to **Text…/Goto defined text** and click beta sheet (6 strands).

Use the tool **New segment** on the CA Build palette. The **Accept:Quit** palette will open and the message line will prompt you to pick a bones point from the graphics window. Find a branch point on the bones skeleton that looks like a piece of main chain with a long side chain. A red cross appears on the picked location. This is the starting alpha carbon.

If you are not satisfied with this point, simply click on another location on the bones skeleton. When the required point has been picked, then click on the **Accept** tool from the Accept:Quit palette.

The bones will be recalculated to be centered on this point, and a two Cα atom line segment will appear. The current Cα atom is yellow, and all other Cα atoms in the current segment are red. All other segments are green. Make sure the map is turned on.

Use the **next CA** tool to extend the trace by one Cα atom. When you have built as far as possible in one direction, use the **Reverse chain** tool to start building in the opposite direction. When you can no longer build this segment, repeat the operation using the new segment tool to add a new segment of Cα trace, and the **next CA** tool and **reverse chain** tools to extend the trace in the each direction. If you wish to remove a Cα atom, use **CA build/Delete current CA**.

12. Refinement

To improve the fit of the current segment of Cα trace you can use the tool **Refine current seg** on the CA Build palette. You can abort the refinement by clicking anywhere in the graphics window if things are obviously going wrong. You can also manually adjust the position by clicking on the tool **CA build/Move current seg** and use the virtual track ball.

13. Joining two segments

If you have made sufficient progress during the time allowed it is likely that you will need to join two segments of Cα trace.

To join two segments select **Join 2 segments** on the X-AUTOFIT:X-BUILD palette. The message line prompts you to select two alpha carbons from two built traces.

14. Opening a molecular structure file

You may want to compare your build Cα trace with the final refined coordinates for the RNase.

Click on **Finish** on the X-AUTOFIT:X-BUILD palette. To open the coordinates rnase.msf use **File/Open**. Click on the file rnase.msf (it will become highlighted) and then click on **Open**. The RNase appears in the main model window. If you go back into X-AUTOFIT you will be able to compare the positions. To make the comparison easier you will need to turn off the bones and map.

15. End session

To end your session click on **Finish** on the X-AUTOFIT:X-BUILD palette and go to the **File** pulldown menu. Select **Exit QUANTA** from the menu.

References

1. Sevcik, J.; Dodson, E.; Dodson, G. G.; Zelinka, J. “The X-Ray analysis of ribonuclease Sa.”, Metabolism of Nucleic Acids, including Gene Manipulation. Slovak Academy of Science, Bratislava., **33** (1987).
2. Sevcik, J.; Dodson, E.; Dodson, G. G. “Determination and restrained least-squares refinement of the crystal structures of ribonuclease Sa and its complex with 3'guanylic acid at 1.8 Å”, *Acta Cryst*, **B47** 240 (1991).

Lesson 4: Automated CA-tracing

The X-POWERFIT application is designed to analyze and interpret electron density maps of proteins and to detect the presence of alpha helices and beta sheets (1). The application gives the best results for large maps where the amount of information presented to the user by either the electron density or bones can make the Cα tracing task daunting. The X-POWERFIT application also provides methods to refine secondary structure into electron density and an extension to the semi-automated fitting to automatically trace multiple residues into density.

In this tutorial, the following material will be covered.

- Skeletonization of an electron density map,
- The calculation of a mask boundary from the map,
- Determination of the secondary structure, and
- Fitting a Cα trace into this map

The tutorial involves the use of electron density map (2.4 Å) of the protein RNase Sa (2, 3) that you used in the previous lesson.

1. Opening a molecular structure file

Move to the directory lesson6/ and start QUANTA.

When QUANTA has finished loading, open the coordinates rnase.msf (for reference) using **File/Open**. Click on the file rnase.msf (it will become highlighted) and then click on **Open**. The RNase appears in the main model window.

2. The map management table

Select **Map Table** from the **Draw** pulldown menu; pull across to **Show Map Table**. The Map Management table appears as the front window. Select **Maps/Add a Map** in the Map Management table. Choose the map called rnase.mbk by highlighting it and clicking Open. It is recommended that the map be contoured at 1.5 (which is 0.30 e Å-3). Enter 0.30 in the first **Level** box and click the **OK** button.

3. Start X-Powerfit

Select **Applications/X-AUTOFIT**. The program opens two palettes - the main X-AUTOFIT:X-BUILD palette and a Pointer palette. A Ramachandran plot appears for the structure in a window labeled **CA angle/torsion**. The map is contoured about the center of the molecule. To open the X-Powerfit palette, pick the tool **X-AUTOFIT:X-BUILD /X-POWERFIT**. Hide the molecule by clicking on the visible box on the Molecule Management table.

4. Bones generation and editing

Since we do not know where the molecule is in the map, you first need to identify the molecule in the map. This is most easily carried out using the bones display, as it is possible to remove excess information from the display very easily. First it is necessary to display the entire map.

Pick the tool **X-AUTOFIT:X-BUILD/Options…**. Set the map radius to 35 Å. Click on the **OK** button.

After completion of the contouring, the map should cover most of the unit cell (Figure 1). If you turn on the Rnase molecules using the Molecule Management table, these should lie near the center of the displayed map. Turn off the coordinate display again.

Note: You should NOT cause a recalculation of the bones during this editing session (use any pointer tool or bones on-off), as this will negate all the editing, and recalculate the bones in the volume of the map. After the mask has been calculated it is much easier to mask the bones not required.

To generate the mask, we will assume there is no coordinate information (otherwise it is possible to generate a mask directly from the coordinates). It is necessary for the bones to be turned on:

Using the tool X-AUTOFIT:X-BUILD/Bones.../Bones on-off, turn on the bones, and turn off the map with the Map Management table. Now make the bones easier to visualize. Turn off the side chain bones to reduce the information on display by picking X-AUTOFIT:X-BUILD Bones.../Side chain on-off.

Note: Before proceeding, make sure that the Start value is set to 1.3, by using the command Bones/Change start value.

To edit the bones, you will use a tool to automatically delete bones by volume. This tool deletes fragments that are smaller than a cutoff threshold defined by their size divided by the total bones volume. The working tool

takes a few seconds. The message line indicates the progress of the calculation.

Initially, to get an idea of the automated deletion of the bones, you should pick the tool **X-AUTOFIT:X-BUILD Bones... /Delete all fragments** five times. You should see in the textport the report of a delete percentage. This indicates the percentage of the total bones deleted. After five deletions, any fragment less than 32% of the initial size will have been deleted.

At this point, you should see one large volume of bones with a tail leading off. You may need to rotate the image to see the tail. It is necessary to remove this side fragment.

Pick the tool **X-AUTOFIT:X-BUILD Bones... /Delete 1 section** and pick a bones point on the link. The link will disappear from the screen.

To delete the offending tail we will use the tool that deletes whole fragments.

Pick the tool **X-AUTOFIT:X-BUILD Bones... /Delete fragments** and pick a bones point on the fragment. The fragment disappears from the screen. Note that this tool remains active until picked again. You should watch the message line and wait for the pick a bones point prompt after each pick.

If you delete a required part of the bones, you can use the tool **X-AUTOFIT:X-BUILD Bones... /Undo last delete** to restore it. Now turn on the side chains again using **X-AUTOFIT:X-BUILD Bones... /Side chains on-off**. You will see some small sections of bones that remain, and need to be deleted before continuing. IF you do not see the small bones, click on the **Side chains on-off** command again. Use the tool **X-AUTOFIT:X-BUILD Bones.../Delete fragments** to move around and pick/delete each small section. (Remember to pick the tool again to abort this mode).

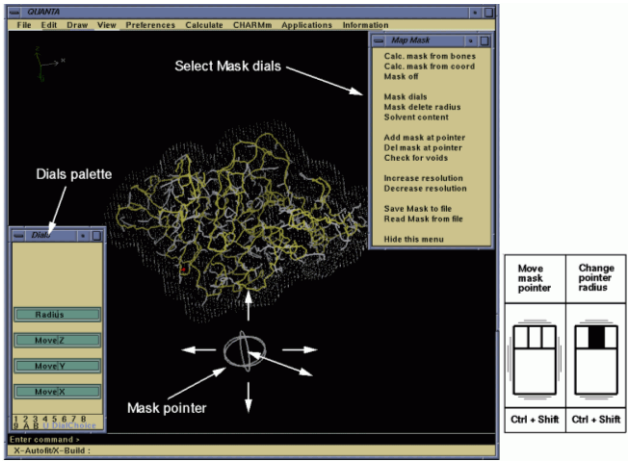
On completion of this process, you should have just the bones that define a monomer.

Note: To check the bones in a real case you can use this tool to show the bones symmetry: **X-AUTOFIT:X-BUILD Bones.../Calc bones symmetry**.

This calculates the symmetry bones and will display these as a reduced representation net of bones. These should NOT overlap the real bones after the completion of the editing.

Pick the tool **X-AUTOFIT:X-BUILD Map Mask... /Calc. mask from bones**. The progress of the calculation is shown on the message line. Upon completion a net of white points indicate the mask.

You may wish to edit the mask with the mask sphere pointer. You can manipulate the mask sphere pointer by using the Dials emulator or a combination of keyboard keys and mouse buttons as shown (Figure 2).



If the graphics on your machine are not very powerful, it is recommended that you hide the bones using the Object Management table and reduce the dot density of the mask by picking **X-AUTOFIT:X-BUILD/Map Mask.../Decrease resolution**. This will make it easier to manipulate the pointer and display.

To remove the deep cavities into the mask, move the mask pointer over these and use the tool **X-AUTOFIT:X-BUILD Map Mask.../Add Mask at pointer**. This increases the number of inside points in the mask.

To delete sections from the mask use the tool **X-AUTOFIT:X-BUILD Map Mask.../Del Mask at pointer**. To remove voids from within the mask select **X-AUTOFIT:X-BUILD/Map Mask.../Check for voids**.

On completion of editing the mask, you should have a mask that forms the monomer molecular outline. You should save the mask by picking **X-AUTOFIT:X-BUILD/Map Mask.../Save Mask to file** and giving the file a name.

5. Using the mask as a bounding region

This mask can now be used in subsequent sessions as a bounding volume for all calculations.

> The tool **X-AUTOFIT:X-BUILD/ Bones.../Mask bones by mask** will set up the bones to always be truncated so that they lie inside the mask.

If this tool is used now, then the textport message indicates that there are no points outside the mask. This is because this mask has been generated by the bones, and so surrounds the bones anyway.

> To see the action, turn off the bones using the **X-AUTOFIT:X-BUILD/ Bones.../Bones on-off** tool, then pick the tool again to turn the bones back on.

The program indicates that some points lie outside the mask, and have been deleted. Therefore in any subsequent session, read in a mask, then turn on the bones, then mask these using the mask.

> You should turn off the visibility of the mask (using the Object Management table) in the subsequent analysis.

6. Identifying secondary structure

For the identification of the secondary structure, you should have:

- The map open, and displayed and centered in the middle of the monomer/mask and covering the molecule/mask. A map radius of 18 Å should be suitable if centered at the monomer.
- The bones turned on and the start level set to 1.3 using the tool **X-AUTOFIT:X-BUILD Bones.../Change start value**.
- You should have a map mask read in and the bones masked by the mask (i.e. the bones should fill only the mask volume).

> Now turn off the display of the mask using the Object Management table.

It is not necessary to view the mask when the **Mask bones by mask** tool is active, as the program takes care to prevent the user building outside the mask. The hiding of the mask increases the speed of the next part of the calculation, as the image display will update during the analysis.

To calculate the secondary structure of the molecule, pick the tool **X-AUTOFIT-X-BUILD/X-POWERFIT/Find sec. struct**.

You should see the pattern recognition algorithm marking possible strand and helix positions. Then the next stage of the calculation carries out a cluster analysis, followed by directed refinement and finally an overlap analysis. You should end up with six or seven structure elements, depending on the mask quality.

7. Conversion of the secondary structure to Cα atoms

You should check that the vectors are correct by looking at the bones (and map) closely when converting the vectors to Cα trace.

To do this, pick the tool **X-AUTOFIT-X-BUILD/X-POWERFIT/ Vector to CA Trace**, then select a vector.

The algorithm tends to work better for larger maps where solvent flattening does not affect the ends of the strands.

8. Extending the Cα Trace

Once the secondary structure has been placed, it is now possible to extend this with the tool **X-POWERFIT/Auto extend CA**. The extension is carried out from the current Cα atom in the current Cα trace. You should note:

- The starting Cα atom from which the building is to be continued should be well fitted.
- The building process is very critical of its progress and often stops with errors - these should be checked.
- It is possible to interrupt the auto building process by a click with the left mouse button.

You should experiment with using the tool from each end of the secondary structure placed, using the pie chart in the bottom left hand corner of the display to view the progress.

> **Note:** You can also use any of the manual Cα tracing tools that were utilized in the previous lesson.

9. Cα refinement

Segments of placed Cα trace can be refined with the tool **X-POWERFIT/CA refinement**, particularly where secondary structure elements have been added in places where the density indicates deformation within the secondary structure. When using this tool make sure that the map completely covers the current segment you wish to refine. This is a very powerful tool that can improve the Cα trace significantly.

10. Comparing your results to the known structure

You may want to compare your build Cα trace with the final refined coordinates for the RNase.

Display the molecule by setting the Visible column to yes in the Maps Management table. Compare the complete structure to the segments produced by X-AUTOFIT.

11. End session

To end your session click on **Finish** on the X-AUTOFIT:X-BUILD palette and go to the File pulldown menu. Select **Exit QUANTA** from the menu

References

1. Oldfield TJ. Pattern recognition methods to identify secondary structure within X-Ray crystallographic electron density maps. *Acta Crystallogr D*. **D58**, 487-493 (2002).

2. Sevcik J, Dodson E, Dodson GG, Zelinka J. 1987. The X-Ray analysis of ribonuclease Sa. In Metabolism of Nucleic Acids, including Gene Manipulation, pp. 33. Bratislava: Slovak Academy of Science.

3. Sevcik J, Dodson EJ, Dodson GG. 1991. Determination and restrained least-squares refinement of the structures of ribonuclease Sa and its complex with 3'-guanylic acid at 1.8 A resolution. *Acta Crystallogr B* **47**: 240.

4. Oldfield TJ. Automated tracing of electrom-density maps of proteins *Acta Cryst.* **D59**, 483 (2003).

Lesson 5: Sequence assignment

In this tutorial, the following material will be covered:

- Sequence assignment, and
- Converting a Cα trace to an all atom representation.

The tutorial involves the use of the electron density map of the protein RNase Sa (1,2) that you have used in the previous two lessons. During this session, you will be converting the Cα trace into an all atom representation.

1. Start X-Autofit

Move into the directory lesson7/ and start QUANTA.

Select **Applications/X-AUTOFIT**.

The program opens the following palettes:

- The main X-AUTOFIT:X-BUILD palette
- The Object Management palette
- The Molecule Management palette
- The Pointer palette
- The Sequence palette
- A palette containing all of the standard amino acids labeled Specific.
- The Cα angle/torsion window

X-AUTOFIT has a sequence alignment algorithm that allows you to generate alignment information that matches molecular sequence information of

the structure you are studying with alpha-carbon segments you have generated. The algorithm can be applied after you have labeled at least one residue either specifically or using a fuzzy residue type. When you start the application a section of Cα trace will appear as a red object and the sequence for the monomer is displayed at the top of the molecule window. Each Cα has the text string UNK to indicate that no sequence information is known at the moment.

2. Map management table

From the **Draw** pullright menu, pull across to **Show Map Table**. Then, select **Map Table**. The Map Management table appears as the front window. The map rnase.mbk is already loaded and has been contoured at 0.23 e Å-3. To toggle the map on and off click on the level box in the Map Management table that currently reads 0.23.

3. Assignment of the first residue

Click **Text…** on the X-AUTOFIT:X-BUILD palette to open the 3D Text palette. Text strings have already been added in this tutorial. You can get to the first text string by picking the tool **Goto defined text**. Since the first text string (Tyrosine) is already highlighted click the **Goto** button. This takes you to a yellow text mark.

Make sure that the map is now turned on. On the Sequence palette, click **Current res. and seq**. You are prompted to pick a Cα atom. Select the Cα closest to the text item labeled Tyrosine.

The atom you have just selected is now colored yellow to indicate that it is the current CA.

Now select **Tyrosine** from the Specific palette.

Now that you have assigned a sequence definition to an alpha carbon, X-AUTOFIT shows all forward and backward sequence alignments from that residue for the Cα trace. The alignments are displayed as arrows (blue for forward alignment and red for reverse alignment) underneath the aligned residues in the residue sequence table at the top of the molecule window.

The current residue is marked as blue boxes for forward fitting and red boxes for reversed fitting.

4. Assignment of the second residue

Go to the second text string by picking the tool **Goto defined** text. Highlight the second text string (Arginine) and highlight click **Goto**. This takes you to the second yellow text mark. Make sure that the map is now turned on. On the Sequence palette, click **Current res. and seq**. You are prompted to pick a Cα atom. Select the Cα closest to the text item labeled Arginine. Select **Arginine** from the Specific palette.

At this point, X-AUTOFIT has identified a unique sequence for a segment and that sequence is shown in uppercase letters in the sequence table to indicate this.

5. Checking the polarity of the chain

To determine the correct direction of the chain X-AUTOFIT attempts to fit polyglycine to the trace in both directions.

The map must cover the entire current Cα segment since the atoms are built into the electron density. If it does not, then you should move the pointer into the center of the Cα trace and recalculate the map. To do this, first of all activate the pointer dials by clicking **Pointer dials** on the Pointer palette. You can then move the pointer around using the virtual track ball option. When you are satisfied with the pointer position click **Go to pointer** on the Pointer palette to re-calculate the map. The volume of map displayed can be changed using the **Options…** palette selected from the main X-AUTOFIT:X-BUILD palette. To get polarity information for the current segment, select **Check CA direction** on the CA Build palette.

The following information is then reported in the text port:

- The percent fit.
- A statement that the chain is the correct/wrong way around.
- A fit ratio that gives a probability on the correctness of orientation.

The ratio compares the fit for the current chain direction (left number) with the reversed chain direction (right number).

To change segment polarity select **Reverse chain** on the CA Build palette. The colors of any arrows indicating sequence alignments for the segment are reversed (that is, red becomes blue and blue becomes red).

6. Building all-atom representation from Cα trace

Once a Cα trace has been generated and the sequence has been assigned then the production of an all-atom model is a trivial process. There are three methods of generating all atom models from CA-traces: just using real space refinement (RSR), fitting the main chain with database fragment fitting (and the side chains with RSR), and fitting the main chain atoms by direct correlation of the Cα conformation with all atom geometry (and the side chains with RSR).

> Make sure the chain is the correct way round and then build the entire chain using **Fit seg by CA corr** tool. If you have contoured the map over the entire chain you should find that the forward direction (blue arrows) is significantly better fit.

On completion, the coordinates are colored by fit to the electron density. The goodness of the fit is color coded as follows:

- Green atoms fit density well.
- Yellow atoms fit density moderately well.
- Red atoms fit badly.
- Blue atoms had negative density or no density.

7. Opening a molecular structure file

You may want to compare your build section with the final refined coordinates for the RNase.

> Click on **Finish** on the X-AUTOFIT:X-BUILD palette. When prompted save the coordinates to a new file name. To open the coordinates rnase.msf use **File/Open**. Click on the file rnase.msf (it will become highlighted) and then click on **Open**. The RNase appears in the main model window. If you go back into X-AUTOFIT you will be able to compare the positions. To do this effectively, you may wish to reduce the volume of map displayed to 10-12 Å3 and turn off the Cα trace by toggling the display to **No** in the Object Management window.

8. End session

To end your session click on **Finish** on the X-AUTOFIT:X-BUILD palette and go to the **File** pulldown menu. Select **Exit QUANTA** from the menu

References

1. Sevcik, J.; Dodson, E.; Dodson, G. G.; Zelinka, J. “The X-Ray analysis of ribonuclease Sa.”, *Metabolism of Nucleic Acids, including Gene Manipulation. Slovak Academy of Science, Bratislava.*, 33 (1987).
2. Sevcik, J.; Dodson, E.; Dodson, G. G. “Determination and restrained least-squares refinement of the crystal structures of ribonuclease Sa and its complex with 3'guanylic acid at 1.8 Å”, *Acta Cryst*, **B47** 240 (1991).

Lesson 6: Model Building and Real Space Refinement using X-BUILD

The model building functionality in X-BUILD allows real space refinement in torsion angle space, rigid body refinement, regularization and general manual editing of macromolecular coordinates.

For this tutorial, you will be carrying out some model building on the protein apolipoprotein KIV-10 from lesson 2. In lesson 2, the coordinates of human plasminogen kringle (1) were used as an initial model for a Molecular Replacement calculation using X-Ray reflection data collected from apolipoprotein a KIV-10 (2). For this tutorial, the best molecular replacement solution (rigid_k4.pdb) has been used to calculate a 2Fo-Fc electron density map (1.9 Å) (2Fo-Fc_k4.mbk). Recall that the sequence identity between the two proteins is approximately 80%.

You will work through the list of tools on the main model building palettes to demonstrate the building functionality within X-BUILD. The text editor text markers are used to indicate a residue for the action, but after working through the tutorial any part of the apolipoprotein structure can be modeled.

This tutorial covers the following material:

- Opening a molecular structure file (MSF),
- Opening an electron density map,

- Torsion angle real space refinement,
- Regularization,
- Alternate amino acid conformations,
- Addition/deletion of terminal residues,
- Mutation of residues,
- Loop fitting, and
- Validation.

1. Opening a molecular structure file

Move into the directory lesson8/which contains the necessary files for this tutorial. Start QUANTA.

When QUANTA has finished loading open the coordinates rigid_k4.msf using **File/Open**. Click on the file rigid_k4.msf (it will become highlighted) and then click on **Open**. The molecule appears in the main model window.

2. Setting up the map display

Select **Map Table** from the **Draw** pulldown menu, while keeping the mouse button depressed pull across to **Show Map Table**. The Map Management table appears as the front window. Select Maps/Add a Map in the Map Management table. Choose the map called 2Fo-Fc_k4.mbk by highlighting it and click **Open**.

The Define Contour Levels and Characteristics dialog box appears. It is recommended that the maps be contoured at 1.0 and 2.0σ (which is 0.95 and 1.9 e $Å^{-3}$). To do this, enter the values 0.95 and 1.9 in the Level boxes. Click **OK**.

3. Starting the application

Select **Applications/X-AUTOFIT**. The main X-BUILD:X-AUTOFIT palette opens, along with the Pointer palette, an Object Management table, a Ramachandran Plot, and the display of some of the map. You will also need the Text palette and Building palette. Open these by picking the tools **X-BUILD/Text…** and **X-BUILD/Build atoms…**.

For each tool, you will be asked to use the 3D text editor tool **Goto defined text** to go to a specific part of the structure (labeled with a text string) to try the tool. To complete the tutorial as designed, please follow the instructions.

4. Refine 1 residue tool

The **Refine 1 residue** tool fits single residues or atoms using gradient refinement. You will use this tool to refine a ligand into some density as it is approximately 0.5 Å from the center of the density.

Go to the {Refine 1 residue} text string using the **3D text tool/Goto defined text**. Pick the tool **Refine 1 residue** from the Build atoms palette. The message line at the bottom of the graphics window asks you to Pick a residue/water. Move the mouse arrow to the ligand and select an atom with a click with the left mouse button. An Accept:Quit palette appears and the ligand begins to refine.

The refinement is graphics-limited and will proceed much faster without the map turned on. The message line indicates the progress of the refinement and indicates when the refinement has completed. (If we wanted to stop the refinement, clicking with the left mouse button would do so.)

When the message line says Finished, you can either Accept or Quit the solution using the Accept:Quit palette.

Note: Note: The X-Ray applications do not allow the use of any other tool while the Accept:Quit palette is active.

5. Geometric conformation tool

This tool is used to place side chains of amino acids in conformations normally found in proteins when there is no experimental information to define the conformation experimentally. In this case, a glutamic acid residue requires attention.

Go to the {Geometric conformation} text string using the 3D text tool Goto defined text.

Pick the tool **Build atoms/Geometric conformation**. The message line at the bottom of the graphics window asks you to Pick atom to set geometry. Pick any atom in the glutamate residue. A Geometry palette appears on the right-hand side of the screen, containing seven rotamer options, all trans conformation, and an **Accept:Quit** pair. You cannot pick any other tool while this palette is active. The first option (**Rota 1 -> 27.2%**) is highlighted, and this conformation appears in white on the graphics view of the molecule. Pick different rotamers from the Geometry palette and look at the different conformations, their energy and bumps.

You can either Accept or Quit the solution using the **Accept:Quit** pair.

6. Side chain fitting using grid real space refinement

Go to the {Fit side chain by RSR} text string using the 3D text tool **Goto defined text**.

There you will find an arginine residue on the surface of the protein that does not fit the density. In fact, it is a very long way from the correct position and too far for gradient-type refinement to work.

To fit this residue automatically into the density, pick the tool **Fit side chain by RSR**. The message line at the bottom of the graphics window indicates that you should pick an atom at which to begin the refinement. This means that the program will **ONLY REFINE OUTWARD FROM THE ATOM SELECTED**, so you **MUST** pick the Cα atom of this residue to fit the whole side chain.

The arginine immediately jumps into the electron density.

The **Move atom + RSR** option (the fifth tool down the Build atoms palette) continuously refines all the side chain chi angles that follow a side chain atom that you are moving. Make sure you remember how to use the virtual track ball option (you can practice moving the pointer first) before doing the following:

Pick the tool **Move atom + RSR** in the Build atoms palette. The program prompts you pick an atom to move, and RSR simultaneously. Pick the Cα of this arginine again. The **Accept:Quit** palette and a white current solution appears. The message line at the bottom of the graphics window indicates the current fit to density.

Move the Cα atom using the virtual trackball XY motion and watch the side chain trying to stick in the density. Eventually, if you move the Cα atom far enough, the side chain will flip to some other density.

The application only applies local non-bonding, so you should also watch the energy value and the non-bond indicators during the refinement.

When you have finished moving the atom, pick either the Accept or Quit option from the **Accept:Quit** palette, and the white solution disappears. If you choose **Accept**, the side chain is updated.

7. Editing the peptide plane

Go to the {Edit peptide plane} text string using the 3D text tool **Goto defined text**.

There are three ways to edit peptide planes. At the labeled peptide bond, you will use all three methods.

First, we will use the **Flip torsion by 180 deg** tool.

Pick the tool **Flip torsion 180 deg**. on the Build atoms palette. You are prompted to pick a rotatable bond, in this case the peptide bond marked. (Note that this is not strictly a rotatable bond because this selection results in the rotation of the peptide plane, and not the bond). The peptide plane flips 180º.

Now, we will use the **Fit main chain by RSR tool**, which is the second method.

To change this back again, use the tool **Fit main chain by RSR**. You are prompted to pick a peptide bond. (If you do not pick a peptide bond, the program ignores the tool.) The application fits the atoms into the density by refining the peptide plane AND the omega angle (within limits). The atoms should fit the density again.

Finally, the third approach uses the **Edit backbone torsion** tool.

To edit the peptide plane and omega manually, you can use the tool **Edit backbone tor**. After selecting a peptide bond, this tool allows you to change the peptide plane with the virtual trackball, but the omega angle can ONLY be changed with the dial box.

Adjusting the omega angle only with the dial box helps ensure that the omega angle is not moved by mistake. If you look at the Ramachandran plot, you find the two phi-psi pairs of torsion that are changed by this tool displayed, and as you more the peptide plane, the dots move. The message line also shows the actual values of phi and psi plus omega.

8. Regularization of coordinates

Go to the {Regularize} text string using the 3D text tool **Goto defined text**.

You find a glutamine residue at this position in the structure with poor geometry.

To correct the geometry of this residue, select the tool **Regularize** from the Build atoms palette. The program prompts for the first residue in the regularize zone and the last residue in the regularize zone. Pick any atom in the glutamine residue twice to indicate that this is the only residue to be regularized.

The residue regularizes to the expected geometry. The regularization proceeds much faster when the map is turned off, as the display refresh is much faster. You can now try the interactive editing while the regularizer is active.

This time pick the tool **Build atoms/Move atoms + reg. res**. and select one of the atoms, Oε1 or Nε2, from this residue. The Accept:Quit palette appears and the message line at the bottom of the QUANTA graphics window shows the residual. A green atom representation of the residue appears in the graphics window. Now use the virtual track ball option to move this picked atom. When moving a single residue with the regularizer active, the refresh of the coordinates is limited by the graphics of the machine. You can now drag the side chain around by this atom, where the N and C atoms of the residue are anchored at their initial positions as shown by the crosses.

The bonds of the residue change color, depending on the deviation of the geometry at each atom at the ends of this bond. To change the moving atom, pick a new atom with the left mouse button. The moving atom is marked with a cross. You can free the fixed ends of the residue (O and N) atoms by picking these as moving atoms.

On completion of the editing, select the **Quit** option from the **Accept:Quit** palette.

To try out **Move atom + reg. zone**, you need to be able to view THR 37, GLN 34 (the residue with the text label {Regularize}, and residue ARG 32.

Use the left mouse button to pick and identify the residues. (Although it does not matter on which residue you actually try this, these residues are used for the tutorial.) Select the tool Move atom + reg. zone. You are prompted for:

* First residue in zone ---------- ARG 32

* Last residue in zone ---------- THR 37

* Pick moving atom ---------- pick the Cα atom of GLN 34

You need not pick the first/last residues in order, and you can pick any atom of a residue to specify the range to regularized.

Initially the zone regularizes to completion. The message line notes that a new moving atom can be picked at any time (with the left mouse button), and a value for the current residual. A green display of the regularized atoms also appears.

If a residue in the zone of regularization contains a cystine residue as part of a disulfide bond, a text port comment, Disulfide link added, appears. This is because the last residue in the zone was a cystine residue, so the link + the other cystine residue is added as part of the regularized residues.

You can now move the current moving atom with the virtual trackball (or dials), and change the moving atom. Experiment with changing the Cα positions, and then either **Accept** or **Quit** the editing.

9. Alternate conformations

Go to the {Alternate conformation} text string using the 3D text tool **Goto defined text**.

An arginine residue is at this text marker with two side chain conformations. You can view which conformation is the B-conformer using a coloring option:

From the Build atoms palette select the **Color atoms...** tool. This opens a new palette, Atom color. There are 6 tools on this palette. Pick **Color B-alt different**. The B-conformer of this residue turns pink.

Find the tool **X-BUILD-X-AUTOFIT/Build atoms/Add-delete/ Delete residue**. When you pick this tool the application prompts you for a residue to delete. Pick the pink B-conformation. The B-conformation disappears.

Now, to place the B-conformation back again, use the tool **Add alternate conformation** on the same palette.

The B-conformation appears, with Cα and Cβ atoms in the same position and the side chain different from the CG atom. This is because the default mode of action for alternate conformation is to place the main chain atoms in the same place, and allow only side chain atoms to have different positions, where CB is the branch point.

To change this action you should change the toggle option on the **X-BUILD/X-AUTOFIT/Options...** dialog box B-conf clamped to A backbone so that it is turned off. Now you can move B-conformation atoms so that the whole can be an alternate conformation, and therefore separate the CB atoms.

10. Adding residues to termini

Go to the {Add LEU then THR} text string using the 3D text tool **Goto defined text**.

You should find that the display has centered on a region of the map where the density continues without any atoms. You will place a leucine residue and a threonine residue in this density.

> You should still have the Add-delete palette open; if not, open this from the Build atoms palette using the **Add/delete...** tool. Select the tool **Add res at termini** and select the carbonyl carbon atom at the 3D text marker. (If you do not pick the terminal residue, the addition aborts.) A palette appears so you can select a residue. Select leucine, which then appears within the density. Select this tool again and place a threonine residue.

Both residues have been placed into the electron density close to a sensible conformation. This is because the application fits the terminal residues using a tree search algorithm for density fitting to the last/previous residue, so that any new atoms are fitted to the electron density in the best possible conformation. Hence, adding a residue to the termini is extremely easy.

You may want to try to improve the fit to the density using other tools after completing the rest of this set of tutorials.

11. Mutate residue

> Go to the {Mutate residue Ser to Tyr} text string using the 3D text tool **Goto defined text**.

You find serine in this position of human plasminogen kringle 4 that should in fact be a tyrosine in apolipoprotein (a) KIV-10

> To correct this residue, select the tool **Mutate residue** from the Build atoms palette. The program prompts for the residue to mutate. Pick any atom in the serine residue. You are then prompted to select a new residue type, select **Tyrosine** from the Specific palette. You may want to try to improve the fit to the density using other tools.

12. Using the Structure palette

Go to the {Loop fit 1} text string using the 3D text tool **Goto defined text**. You can experiment with the Structure palette for modeling zones of residues. Use the tool **Loop fit** to fit the loop Cα 67 to Cα 72 of segment ARIG. Then use the **Refine zone** tool to improve the loop fitting solution.

You should use the tool **X-BUILD:X-AUTOFIT/Structure.../Loop fit** and pick the two labeled atoms ({Loop fit 1} to {Loop fit 2}). The program searches for loop conformations until you click with the left mouse button to abort the search, or the time limit expires (See the Options... dialog bog for the time limit. The default is 10 minutes.). You can then view the best 10 solutions (defined by the fit to electron density), Accept/Quit a solution, or continue the search. If the loop search finds a better solution, use the **Refine zone** tool to improve the fit by gradient refinement.

13. Validation

To validate the structure, select the **Protein validation** tool on the X-Autofit:X-Build palette, and then work through the 3D text list of errors using the **Goto next text** tool to look at each one.

14. 14. End Session

To end your session click on **Finish** on the **X-AUTOFIT:X-BUILD palette** and go to the **File pulldown** menu. Select **Exit QUANTA** from the menu.

References

1. Mulichak, A. M.; Tulinsky, A.; Ravichandran, K. G. “Crystal and molecular structure of human plasminogen kringle 4 refined at 1.9 Å resolution.”, *Biochemistry* **30** 10576 (1991).
2. Mochalkin, I.; Cheng, B.; Klezovitch, O.; Scanu, A. M.; Tulinsky, A. “Recombinant kringle IV-10 modules of human apolipoprotein(a):

structure, ligand binding modes, and biological relevance." *Biochemistry* **38** 1990 (1999).

Lesson 7: Ligand Fitting using X-LIGAND

X-LIGAND (1) is designed to search for unsatisfied electron density (density containing no molecular coordinates), sort these in order of volume, and fit a ligand to these sites automatically. The application is also able to search conformation flexibility of a ligand by varying any rotatable bonds and fit these rotamers to density at a rate of more than a thousand per second. This entire process, including refinement, can be carried out rapidly.

In this tutorial, the following material will be covered:

- Searching for possible ligand sites in an electron density map (1.9 Å)
- Conformational searching of ligand
- Refining the ligand.

For this tutorial, you will be fitting a ligand to the protein apolipoprotein KIV-10, which has already been used in several earlier lessons.

1. Set up for the tutorial

Move to the directory lesson9/ that contains necessary files for this tutorial. Start QUANTA.

When QUANTA has finished loading, open the coordinates kiv-10.msf and e-aminocaproic.msf in that order using **File/Open**. Click on **Open**.

The small molecule is aminocaprioc acid (EACA), a competitive inhibitor for the interaction of kringles with other proteins. The binding mode for this ligand was a major objective in the original crystallographic study for this data (2).

2. Setting up the map display

For this tutorial, you will be displayed two electron density maps. An Fo-Fc map will be used to actually fit the ligand while a 2Fo-Fc map will be displayed for reference.

Select Map Table from the **Draw** menu and pull across to Show Map table. The Map Management Table appears as the front window. From that table, select **Maps/Add a Map**. Choose the map called 2Fo-Fc.mbk by highlighting it and click **Open**. The Define Contour Levels and Characteristics dialog box appears. It is recommended that the maps be contoured at 1.0 and 2.0, to do this make sure that two include levels are selected (indicated by black crosses) and click **Levels from Sigma**. Click **OK**.

Again, from the Map Management Table, choose the Maps/Add a Map command. This time select the file Fo-Fc.mbk. The recommended values for the level and color boxes for this map are shown in Figure 1. Click **OK**.

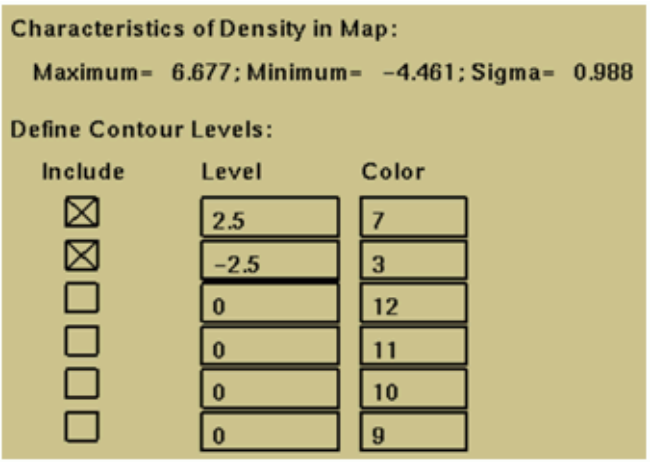


Figure 10. Suitable contour levels and colors for the Fo-Fc map

Select **Applications/X-LIGAND**. First of all, you will be prompted to choose which map you wish to use for the ligand fitting procedure. Select **Fo-Fc.mbk** and click OK.

When you start X-LIGAND, there will be a few seconds delay. This is because the program must generate a sphere of symmetry atoms, and then make sure the map covers these atoms. The X-LIGAND application generates a sphere of symmetry atoms around the working molecule so that symmetry ligand sites are not found as multiple sites. The message line at the bottom of the graphics screen will indicate the progress of the setting up. The X-LIGAND palette will appear with only 4 tools that can be selected.

If you zoom out a little you will see that the graphics display shows the molecule of apolipoprotein KIV-10, the ligand, and a blue sphere of symmetry atoms. The Object Management Window has also popped up to show that symmetry has been created as an object (SYMMETRY).

3. Finding possible ligand sites

Click the palette tool **Change search setting…**.

This sets the map level (in σ) at which the search for possible ligand sites is made in the map. A lower threshold will result in more sites being found, and will also increase the extent of the volume of each site that does not overlap the protein. A lower threshold will often result in the overlap of sites, giving fewer, larger sites. This is the only parameter that needs to be set by you. The rest are automatically determined by X-LIGAND.

Enter 3.0σ as the **Search Threshold**. Click **OK** on the **Ligand Search Parameters** box and click on the palette tool **Search for ligands**.

This searches the map within the displayed sphere for possible sites for the ligand, and sorts these in order of volume. The display centers at the first (largest) possible ligand site, the Fo-Fc map is displayed at 2.5σ along with a search volume that is indicated by yellow crosses. In addition, the 2Fo-Fc map is displayed at 1.0 and 2.0 although it is not being used in the ligand fitting procedure.

At this point, the application has already fitted the ligand at this site as a rigid body, so you will find that it will be in the correct position already, but in an incorrect conformation and orientation.

Note: You can turn the map on and off with the Map on/off tool on the X-LIGAND palette, the display manipulation is much easier when the map is off. You can also toggle the maps on/off from the Maps Management table.

4. A conformational search of the ligand

Click the **X-LIGAND** tool **Search conformations**.

As the ligand is currently a long way from the best solution, there is a delay of a few seconds. Each improved solution is displayed as it is found. The actual computation runs at about 1000 conformations/second.

This example has five internal degrees of freedom and the program uses the Monte Carlo search algorithm. The maximum time that the search will run is by default one minute although you can interrupt it at any time by clicking with the left mouse button. Leave it running for at least 30 seconds.

After a few minutes, the program returns the best 20 conformations.

Use the tool **...all 20 conformations** to view all 20. You will see for this ligand that they are very similar. You can toggle through the different conformations using the tools **...previous conformation** and **...next conformation** tools.

5. Refinement

You can now refine the ligand conformation, orientation, and position using the **Refine** tool on the X-LIGAND palette. You will need to select just one conformation before using this tool.

The refinement method is torsion angle real space refinement using both linear and quadratic interpolation of the electron density. The final coordinates will be within 0.01 Å of the refinement carried out with standard refinement procedures, but the radius of convergence is much higher with this routine.

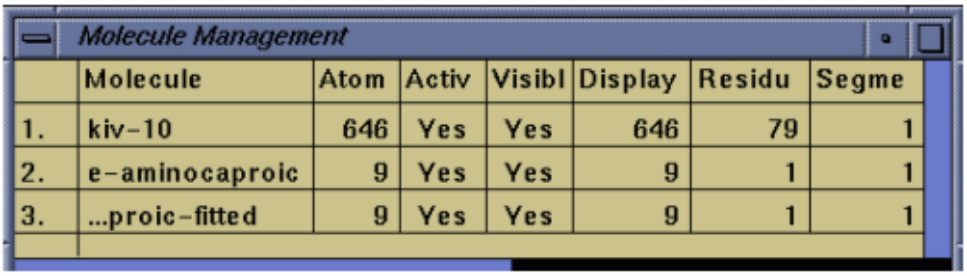
6. Saving the result

In the X-LIGAND palette, use the tool **Save ligand to MSF**. Enter e-aminocaproic-fitted and click **Save**. This creates a new coordinate file containing the refined coordinates.

7. Exit the X-Ligand application

To end your session click on **Exit** on the X-LIGAND palette

At this stage, you should have three molecules displayed in the main QUANTA window and listed in the Molecule Management window (see Figure 11.).


Molecule Management

	Molecule	Atom	Activ	Visibl	Display	Residu	Segme
1.	kiv-10	646	Yes	Yes	646	79	1
2.	e-aminocaproic	9	Yes	Yes	9	1	1
3.	...proic-fitted	9	Yes	Yes	9	1	1

Figure 11. The Molecule Management window

You should retain the protein (kiv-10) and the fitted conformation of the ligand (e-aminocaproic-fitted). To close the coordinates **e-aminocaproic** use the **File/Close** command. Click on the file e-aminocaproic.msf (it will become highlighted and appear in the Close panel) and then click on **OK**.

You will now save the protein and fitted ligand to a single msf file. Click **File/Save As...** and accept the default writing options. Click **OK**. Call this new filename kiv-10-ligand and click **Save**.

Leave QUANTA running in preparation for the next lesson.

References

1. Oldfield, T. “X-LIGAND: An application for the automated addition of flexible ligands into electron density.” *Acta Cryst* **D57** 696 (2001).
2. Mochalkin, I.; Cheng, B.; Klezovitch, O.; Scanu, A. M.; Tulinsky, A. “Recombinant kringle IV-10 modules of human apolipoprotein(a): structure, ligand binding modes, and biological relevance.” *Biochemistry* **38** 1990 (1999).

Lesson 8: Water Fitting Using X-SOLVATE

X-SOLVATE (1) is a tool that allows you to search an electron density map for water peaks, to adjust the peak positions, and to save them as water coordinates. In this tutorial, we will fit water molecules to the electron density map of apolipoprotein KIV-10 used in the previous lesson.

1. Set up for the tutorial

You should have completed lesson 9 before working through this tutorial.

If QUANTA is not running then move into the directory lesson9/ and at the UNIX prompt, type:

```
> quanta
```

followed by enter.

Before proceeding, ensure that kiv-10-ligand is the only molecule listed in the Molecule Management table.

Select **Applications/X-SOLVATE**. As in the previous tutorial, you will be prompted to choose which map you wish to use for the ligand fitting procedure. Select Fo-Fc.mbk and click **OK**.

QUANTA saves the current selection, color, and display masks. It then generates a selection mask that contains all atoms plus their symmetry equivalents for the map currently loaded. The message line at the bottom of the graphics screen will indicate the progress of the setting up. The graphics display will show the molecule of apolipoprotein KIV-10 with a ligand (fitted in the previous lesson) and a blue sphere of symmetry atoms. You may wish to zoom out a little to see this properly. A three-dial set is displayed containing dials for x/y/z positioning of current water molecules. Additionally, the Search for peaks palette appears with only 4 tools that can be selected.

8. Finding water sites

Click the palette tool **Change search setting…**.

This sets the map level (in σ) at which the search for possible ligand sites is made in the map.

Since we are using an Fo-Fc map enter 2.5σ as the **Min density level**. Click **OK** on the **Peak Search Parameters** box and click on the palette tool **Search for waters**.

This searches the map within the displayed sphere for possible sites for the ligand, and sorts these in order of volume. The display centers at the first (largest) possible water site, the Fo-Fc map is displayed at 2.5σ, and a putative water site is indicated by a rhombohedral cursor, including non-bonds (yellow) and hydrogen bonds (white). In addition, the 2Fo-Fc map is displayed at 1.0 and 2.0σ although it is not being used in the water fitting procedure.

You can move the water site using the dials.

Note: The water would move more quickly with the map off, but this would prevent you from finding the best position for the water.

For a single spherical water site, the program will normally find the best position (by quadratic interpolation) without any need to move the site position.

You can tell the program that this is a water site using the **Save as water** tool. A red cross then appears at the pointer position. To go to the next site use the **Next peak** tool. You can continue for all the water sites, saving each probable site as you go.

After you have saved several sites, you may continue with the exercise.

9. Saving the results

Clicking the **Exit** tool saves the new water positions as a new file, while the **Quit** option leaves and make no changes. Click the **Exit** tool.

At this stage, you should have two molecules displayed in the main QUANTA window and listed in the Molecule Management window. These are the protein itself (kiv-10-ligand) and any saved water molecules (searchwaters).

> You will now save the protein and fitted water molecules to a single msf file. Click **File/Save** As and select all the writing options except **Apply rotational transformations**. Click on the **OK** button.
>
> Enter as the new filename kiv-10-lig-wat and click **Save**.

Leave QUANTA running in preparation for the next lesson.

10. Other variations to try

You might like to investigate the effects of changing various parameters if you have extra time.

> You can adjust the parameters for a search by selecting **Change search setting...** in the Search for peaks palette.

A dialog box is displayed and contains the following parameters with defaults shown on the screen. You can investigate changing various parameters as you wish.

- **Min density level**: This is the minimum value of electron density in sigma where a water can be placed.
- **Min distance to protein**: This is the minimum distance possible for a peak to be from the protein.
- **Max distance to protein**: This is the maximum distance from the protein that a water peak is found. The setting prevents X-SOLVATE from finding peaks in a density region where no other data justifies placing one. The aim of the routine is to find the first solvation shell, then to repeat the peak search with the first solvent shell present, thus allowing the determination of further solvent shells by iteration.
- **Min sep between peaks**: This is the minimum distance allowed between peaks. Peaks that are closer together than the min-sep value are replaced by a single peak in the average position. The value used here depends on your working practice.

- **Calculation radius**: This is the radius that describes the region to be searched by X-SOLVATE.
- **Display radius** of map: This changes the map display radius about each peak.
- **Temperature factor**: This is the BVALUE of the temperature factor saved with the water position coordinates.

References

1. Oldfield, T. “X-LIGAND: An application for the automated addition of flexible ligands into electron density.” *Acta Cryst* **D57** 696 (2001).

A. References

Brünger, A. T., "Crystallographic Refinement by Simulated Annealing: Application to a 2.8 Å Resolution Structure of Aspartate Aminotransferase", *J. Mol. Biol.*, **203** 803 (1988).

Brünger, A.T., X-*PLOR Version 3.1: A System for X-Ray Crystallography and NMR*, Yale University Press, New Haven and London (1992).

Brünger, A. T. and Nilges, M., "Computational Challenges for Macromolecular Structure Determination by X-Ray Crystallography and Solution NMR Spectroscopy," *Quarterly Rev. Biophys.*, **26** 49 (1993).

Dickerson, R. E. and Guis, I., *The Structure and Action of Proteins*. Benjamin/Cummings, ISBN 0-8053-2391-0.

Engh, R.A. and Huber, R., Acta Cryst. Sect. A., 1991.

Goodford, P. L., *J. Med. Chem*., **28** 849-857 (1985).

Greer, J., *J. Mol. Biol.*, **82,** 279-301 (1974).

Lattman, E. E., "Use of the Rotation and Translation Functions," *Methods Enzymol.*, **115** 55 (1985).

Oldfield, T.J. and Hubbard, R. H., *Protein: Structure, Function and Genetics*, **18,** 324-337 (1994).

Ramachandran, G.N. and Sasisekharan, V., "Conformation of polypeptides and proteins", *Adv. Prot. Chem.* **23,** 283–437 (1968).

B. Creating a Fragment Database

Use the following procedure to create a database to be used by the Search Fragment Database utility.

1. From the Brookhaven database, select a set of protein coordinates files that have good resolution and include different structure types.
2. Construct a file (dmlist) that contains a list of these protein coordinate files. Use the following format in constructing the file:

- Number of proteins to be used.
- Name of coordinate file 1.
- Name of coordinate file 2.

  .

  .

  .
- Name of coordinate file *n.*

3. Run the program $HYD_MSF/dmprep. The program prompts for the name of the file (dmlist) containing the list of proteins and asks for a name for the distance matrix file (dmfile.new) to be created. The program then reads each protein coordinate file and constructs a distance matrix file. It also creates a QUANTA input command file. The command file is used from within QUANTA to generate an MSF for each of the protein coordinate files. You are prompted to name this file.

   The dmprep executable distributed with QUANTA can handle up to 2,000 proteins with limits of 2,000 residues and 100,000 Cα distances per protein. The FORTRAN sources for dmprep (dmprep.f and dmsubs.f) are also distributed. This gives you flexibility to increase the dimensions as you need them.

4. Move the distance matrix file to the $QNT_ROOT/dmatrix directory and rename it to dmfile. Because the variable $HYD_DMF is already defined in the QUANTA environment as $QNT_ROOT/dmatrix/*dmfile*, you can do this easily by typing:

```
cp dmfile.new $HYD_DMF
```

   where *dmfile.new* is the filename of the distance matrix file created in step 3.

5. To create required MSFs, start QUANTA and type **`@command_file`**, where *command_file* is the name given to the QUANTA command file. Respond appropriately to the dialog boxes. Treat the sixth character in

the atom field as a disorder using the no-hydrogen dictionary file, and exclude symmetry in the molecular structure file.

6. Move the newly created MSFs to the directory $MSF_LIB.

C. The cnx.bat File

The cnx.bat file is used by QUANTA to execute CNX jobs using the Accelrys version of CNX. The main function of this file is to pass the input and output filenames to CNX and allow CNX to be started from QUANTA. The last line of the file locates the CNX executable. If appropriate, replace the executable with a local location of CNX on your system. The location of cnx.bat is specified in the file $HY_LIB/applcomm.db under the classification CNX.

The cnx.bat file

```
#!/bin/csh
echo “cnx.bat started...”
#
#       Change the following line if CNX is
#       installed elsewhere

setenv XROOT $QNT_ROOT/cnx

$XROOT/exec/cnx.exe < $1.inv $1.out
echo “cnx.bat finished”
```

Academic sites may point to an executable for CNX created from a separate Yale distribution of CNX.

D. Searching for Fragments

QUANTA's X-Ray crystallography applications include a function for searching a specialized database to locate a suitable set of fragments for modeling undefined regions of a protein.

This chapter describes:

- Searching for fragments.
- Search Fragment Database palette.
- Example: Using fragment searching to complete a model of renin.

Searching for fragments

Coordinates for undefined residues can be defined using the fragment search process. To search for fragments, you use a database for a set of proteins that matches certain specifications. Matched fragments are retrieved, displayed in the viewing area, and listed in the textport. When you accept a fragment, its conformation is copied to the undefined region of your structure.

Fragment database

The fragment database consists of two parts: a distance matrix file and a library directory of complete structures. The distance matrix file contains all the inter-Cα distances for a representative set of protein structures. The library directory contains MSFs with complete sets of atomic coordinates for the structures used in the distance matrix file.

The standard QUANTA installation contains a database of twenty structures derived from the Brookhaven PDB. By default., $HYD_DMF points to the file $QNT_ROOT/dmatrix/dmfile containing the inter-Cα distances for the protein structures. $HYD_LIB points to the library directory $QNT_ROOT/msflib.

If you use the fragment searching facility frequently, you can create a more complete set of structures to be searched. For information on this process, see Creating a Fragment Database (page 331)

Fragment selection criteria

A fragment is chosen using the following criteria:

- The fragment forms a good fit with residues on either side of the undetermined region
- The results of total distance and least squares fit are low compared to other retrieved fragments
- Minimal close contacts of the fragments exist with neighboring regions of the protein
- The residues in the fragment are similar to those in the known structure sequence

Running a search

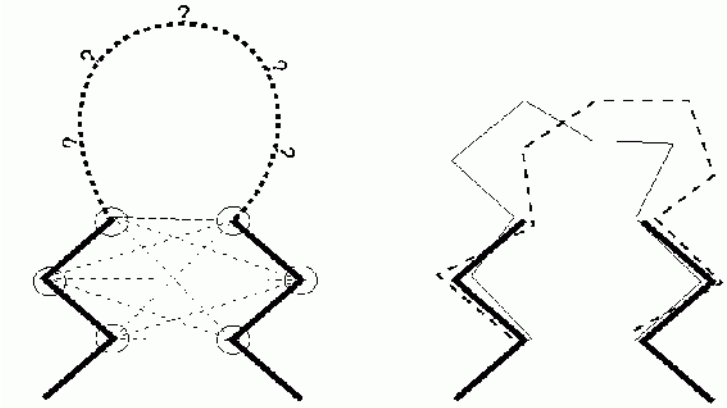
To run a fragment search, you first must specify the anchor residues on either side of the region to be filled by the fragment. Anchor residues are specified by using one of the tools for picking residues on the Search Fragment Database palette to choose the anchor residues in the modeling window or the Sequence table. When you have completed residue picking, you initiate the search by selecting **Search Database** from the Search Fragment Database palette.

Fragment searching also can use the bump checking tool, Bump. This tool takes retrieved fragments and fits them over the search template. The interatomic distances between the main chain, and the beta carbon atoms of the fragment and neighboring residues are calculated. The fragments with close contacts are rejected. Because this procedure reduces the number of fragments finally selected, the initial database search retrieves extra fragments by relaxing search criteria.

When a search is complete, the structures of matched fragments are read from MSFs and displayed superposed over the template residues. Each fragment is color coded and a color-coded legend is displayed in the lower-right corner of the screen. The legend gives the name of the protein from which each fragment is taken, its distance fit, and the RMS difference in Cα atom position when the fragment is superposed on the template.

Accepting a fragment defines coordinates for unknown residues and redefines coordinates of residues in the anchor regions. In some cases, it is advisable to select adjacent residues having structural homology. These residues act as an anchor, holding the fragment in place.

The following figure illustrates the setup and results of a fragment search:



Cα trace of model structure with search template residue Cα atoms marked with * and residues of unknown conformation marked with ?. The calculated Cα- Cα distances are shown as a dashed line.

Cα trace of two possible hit fragments from the database. The Cα atoms marked * are roughly equivalent to the search template Cαs with similar Cα-Cα distances. The residues marked with · are equivalent to the residues of unknown conformation, and represent two possible conformations for the unknown region.

Search Fragment Database palette

This section describes the tools on the Search Fragment Database palette.

To access the Search Fragment Database palette, select **Fragment Database …** from the X-Structure palette.

> **Note:** Many of the tools on this palette are not activated until you have selected the residue range to be searched. Additional tools are activated when a search is initiated.

List Proteins

Lists the proteins in the currently active Cα distance matrix file in the textport.

Pick Alpha Carbon Range

Selects template residues by picking the first and last residue in a range. This picking process is additive. Previously selected ranges are not cleared

even if a search has been performed. Use either **Undo Last** or **Undo All** to clear selections.

Pick Alpha Carbon

Selects template residues by picking each individual residue. This picking process is additive. Previously selected atoms are not cleared even if a search has been performed. Use either **Undo Last** or **Undo All** to clear selections:

Undo Last

Deletes the last Cα anchor selection.

Undo All

Deletes all Cα anchor selections.

Search Database

Searches the fragment database by Cα–Cα distance for matches to the currently selected anchor residues.

With Bumps

When this tool is active, any database search is followed by bumps checking before the optimal retrieved fragments are displayed. This selection is active by default when **Search Database** is active.

Display All Fragments

Displays all the retrieved fragments.

Display Next

Displays the next fragment on the fragment list and removes all others from the viewing area. The initial fragment to be displayed is the first on the fragment list.

Display Previous

Displays the previous fragment on the list and removes all others from the viewing area.

Select Display...

Opens the **Display selected fragments** dialog box with all fragments listed. Use the dialog box to select one or more fragments for display.

List Residues

When only one fragment is displayed, this tool lists names and IDs for the residues in the fragment, and for corresponding residues in the active molecule.

Accept Fragment

Copies the coordinates of a fragment onto the corresponding residues of the active molecule including the original anchor residues. This tool is grayed unless only one fragment is displayed

> **Note:** If the structure being built includes N- or C-terminus atoms, polar hydrogens, or all hydrogens, certain atoms will remain undefined even after accepting a fragment. Undefined atoms occur when no equivalent atoms are present in the incoming fragment. This happens because PDB files do not include hydrogen positions, and C-terminus oxygens have a different naming convention. Correct this situation by regularizing the terminal region.

Anneal Fragment

Averages the coordinates of the fragment with the corresponding residues of the active molecule. This tool is inactive unless only one fragment is displayed. The tool should not be used if the target sequence has undefined atoms.

Reject Fragment

Clears all fragments from the display

Options...

Opens the Fragment Modeling Options dialog box as illustrated:

From this dialog box, you can specify the number of fragments to be displayed after a search, indicate how fragments are displayed and labeled, change the fragment file specifying the database matrix, and specify the number of residues for annealing.

Finish

Exits the Search Fragment Database palette and returns to the X-Structure palette.

Example: Using fragment searching to complete a model of renin

This section is an exercise taken from part of the tutorial in the *Protein Homology Modeling Tutorial*. If you have this book, use it to complete the full exercise.

Before you begin

To do this exercise, you must have the file renin_regularize.msf in your current working directory. This file contains a partially built model of human renin with undefined loops. If you do not have the file, check with your system administrator. The file is also available from the Accelrys Scientific Support group.

The distance matrices (dmfile) for this example are located in the file $HYD_DMF.

To begin this exercise, you make selections from the QUANTA **Draw** menu to display and focus on only those residues of interest in the renin file (renin_regularize.msf). The undefined regions that are appropriate for fragment searching are residues 49 through 62, residues 103 through 112, and residues 117 through 127. Each of these regions spans a reasonable distance and has residues that can be used as anchors.

Complete the following steps:

1. Open renin_regularize.msf to display the renin structure.

2. From the QUANTA Draw menu, select **Display Atoms** to open the pull-right menu. From the pull-right menu, choose **Selection Tools**. The Display Atoms and Display Utilities palettes are displayed.
3. From the Display Atoms palette, choose **Type** in a selection. The Enter Display Selection Commands dialog box appears.
4. In the text entry fields, enter the text strings:

zone 49 to 62
zone 103 to 112
zone 117 to 127

5. Click **Done**. Each of the areas of interest are displayed in color 2 (red) in the structure.
6. From the Display Utilities palette, choose **Save Selection-Commands**. A File Librarian is displayed. In the text entry field enter: Display_frag_area_renin, then click **Save**.
7. Select **Finish** from the Display Atoms palette to exit and save the changes.
8. From the **Draw** menu, select **Color Atoms** to display the pull-right menu. From the pull-right menu, choose **Selection Tools**. The Color Atoms and Color Schemes and Utilities palettes are displayed.
9. From the Color Atoms palette, select **All Atoms** and type in a **Selection**. The Color Selection Commands dialog box appears.
10. In the text entry fields, enter the text strings

zone 49 to 62 = col 2
zone 103 to 112 = col 3
zone 117 to 127 = col 4

11. Click **Done**. Each of the areas of interest is colored and displayed uniquely on the structure.
12. From the Color Schemes and Utilities palette, select **Save Selection-Commands**. A File Librarian is displayed. In the text entry field, enter Col_frag_area_renin. Click **Save**.
13. Choose **Finish** from the Color Atoms palette to exit and save the changes.
14. Display the **Draw** menu. Choose **Label Atoms** to display the pull-right menu.
15. From the **Label Atoms** pull-right menu, choose **Residue ID** and **Selection Tools**. All labels for the structure are visible, and the Label Atoms palette is displayed.

16. From the Label Atoms palette, select **Include**, **Alpha-Carbon Atoms**, and **Finish**. The Cα atoms of the residues in the structure are labeled by their residue IDs. The changes are saved and the utility is exited.

17. From the **Edit** menu, choose **Bond options**. The Specify Bonding Mode dialog box is displayed:

18. Set the options to the following:

    Bonding Algorithm: Intra-residue + Named Link Atoms
    First link atom: C
    Second link atom: N
    Local Search
    Disable Inter-Segment Bonding
    Minimum Distance Criterion: 0.05
    Apply to what? Both

19. Click **OK**. These options result in a display of side chains as they are constructed and prevents spurious bonds from being drawn.

20. From the **Applications** menu, select **X-Structure**. The X-Structure palette is displayed.

21. From the Structure palette, select **Fragment database...** The Search Fragment Database palette is opened.

22. From the Search Fragment Database palette, select **Options…** The Fragment Modeling Options dialog box is displayed.

23. Set the options in the dialog box to the following values:

    Number of Fragments: 5
    Fragment Drawing: Main Chain
    Fragment Labelling: CA Labels
    Anneal over Residues: 1

24. Click **OK**. The dialog box is removed from the display.

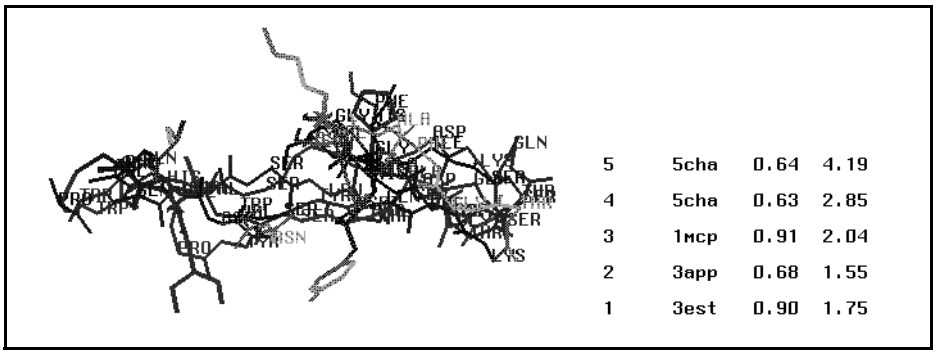
25. From the Search Fragment Database palette, choose **Pick Alpha Carbon Range**.

26. Rotate the molecule so that the first group of atoms in residues 49 through 62 is visible. Select residues 50 (K) and 51 (C), to mark the beginning of the fragment placement area.

27. Select residues 61 (H) and 62 (K) to mark the end of the fragment placement area. The selected residues are indicated by red crosses on the Cα atoms.

28. From the Search Fragment Database palette, select **with Bumps**, then choose **Search Database**.

29. When the search is completed, several best-fit fragments are displayed as illustrated:



30. Rms values for the fragments are listed both in the textport and in the modeling window legend. The fit (dist) column shows the least squares fit between the Cα atoms picked in the current model and the matching Cα atoms for each superposed fragment. The rms (lsq) column shows the least squares fit of the whole peptide backbone of the same selected residues.

Protein	Residue	rms(lsq)	fit(dist)	No bumps	Sequence
3est	79	1.75269	0.90024	0	VHPYWNTDDVAAG
3app	194	1.55386	0.67538	0	NVDSYTAGSQSGD
1mcp	290	2.04288	0.90797	0	FIVSRDTSQSILY
5cha	23	2.84700	0.62772	1	WQVSLQDKTGFHF
5cha	259	4.18792	0.64162	0	PWQVSLQDKTGFH

31. From the Search Fragment Database palette, choose **Display Next** and **Select Display…** to begin the process of studying fragments. The Display selected fragments dialog box is displayed as illustrated, listing all the fragments found in the search:

32. Select fragments 2 and 4 for viewing. Both fragments appear to fit well, but 2 (3app) appears to join better.

33. This fragment is a good candidate for filling the sequence gap because it has one of the lowest values for the least squares fit, it joins well with the surrounding residues, and it has fewer close contacts than other fragments.

34. Again choose **Select Display**. Select fragment 2, 3app. Only this fragment is displayed superposed on the structure.

35. From the Search Fragment Database palette, choose **List Residues**. Residues are listed in the textport for both the fragment and the selected Cα range.

36. From the Search Fragment Database palette, choose **Accept Fragment**. The Copy Fragment dialog box is displayed.

    This dialog box lists the fragment and range of coordinates to be copied to renin_regularize.msf.

37. Click **Yes**. The coordinates from fragment 3app, starting at residue 194, replace this region in renin_regularize.msf.

38. From the Search Fragment Database palette, select **Pick Alpha Carbon Range**. Select residues 103 (T) and 104 (V), to mark the beginning of the fragment placement area. Then select residues 110 (E) and 111 (V) to mark the end of the fragment placement. The four selected residues are indicated by red crosses on the Cα atoms.

39. Repeat steps 27 through 34 for this residue area.

40. Repeat the searching process for residues 117 through 127.

41. From the Search Fragment Database palette, select **Finish**. The palette is replaced with the Edit Protein palette, Protein Utilities palette, and the Protein dials emulator.

42. To regularize the fragment regions, choose **Select Active Range** from the Protein Utilities palette. The Pick Range palette is displayed. Select the residue 50 (K) to mark the beginning and residue 60 (Y) to mark the end. Red bars indicate the selected range.

43. From the Edit Protein palette, select **Regularize….** The specified region is regularized using previously selected parameters for CHARMm.

44. Accept the regularization and repeat the process for the other two residue regions (106 through 108 and 117 through 127). Observe energy levels and relaxing of the regions.

45. From the Edit Protein palette, choose **Save to MSF….** The Save Options dialog box is displayed.

46. Choose the option **Save to a new filename** then click **OK**. A File Librarian is displayed.

47. Enter the filename renin_fragment and then click **Save**. A new MSF is created and the structure is displayed. The name renin_fragment.msf replaces renin_regularize.msf in the **Sequence** table.

You have completed this example.

E. Using RTF and PSF Modes

To minimize operational problems, it is useful to have a structure that includes polar hydrogens and appropriate bonding before you do CNX calculations or perform regularization. Accelrys recommends that you assure that the structure is suitably prepared by calculating CHARMm energy using the tools in the QUANTA CHARMm menu. If you are using a structure that is prepared using X-AUTOFIT, this procedure is not necessary.

This appendix describes the two different modes for calculating CHARMm energy: PSF and RTF. While these modes can be used interchangeably, each has certain strengths that make it best for specific circumstances.

RTF mode uses a set of dictionaries in which each residue is defined. CHARMm uses RTFs to define the connectivity of a sequence of residues. This mode works well for common biological systems where there are multiple repeating residues. RTF mode is less flexible for ligands and general organic molecules.

In PSF mode, connectivity is defined by a model on the screen. This works well for proteins containing nonstandard amino acids or ligands. QUANTA communicates with CNX in PSF mode only.

This chapter describes

- RTF and PSF files and modes
- Setting up a protein in PSF mode

RTF and PSF files and modes

Several fundamental differences exist between RTF mode and PSF mode in QUANTA.

RTF mode

In RTF mode, QUANTA uses a dictionary of residues to assign connectivity, charges, and atom types. The RTF provides a template for each residue, but has no coordinates directly associated with it.

CHARMm can use RTFs to create a PSF which is then associated with coordinates. But RTFs alone do not provide sufficient information to create a PSF. A sequence is also required. In RTF mode, QUANTA sends the

sequence, RTFs, and coordinates to CHARMm. The program reads each residue, then looks up the residue in an RTF file. If an RTF exists for that residue, the residue is built into the PSF based on the RTF and regardless of the connectivity displayed in QUANTA. The process is repeated for all residues in the sequence.

RTF mode works especially well when you import files from the Brookhaven PDB or other sources, use structures that have many bad contacts, or add polar hydrogens.

A major disadvantage of RTF mode is that definitions are needed for all residues. No calculations can be performed if an RTF is missing or does not exist. If a protein contains nonstandard amino acids, you must construct an RTF. This problem appears most often with ligands. QUANTA provides tools for building an RTF in the Molecular Editor.

*s*In PSF mode, QUANTA uses stored and displayed connectivity of the molecule structure. Additionally, information on bond angles, torsion and improper torsion angles, and atomic specifications such as atom name, atom type and charge, stored charges, and bond order are included. QUANTA generates a PSF directly, then sends the PSF coordinates and parameters to CHARMm or CNX for further calculations. QUANTA also provides a parameter chooser that finds all the parameters needed and provides estimates for any missing parameters.

The PSF is the central data model in CHARMm and CNX. In either RTF or PSF mode, a PSF is created, but the source of the connectivity and other atomic information differs. A PSF defines the relationship between atoms but does not contain coordinates. When a valid PSF is generated, coordinates can be provided for the atoms in the PSF. The PSF coordinates and parameters are used by the energy equation to calculate the energy of the structure and all other derivative data such as velocity and forces.

A single PSF can be used with many sets of coordinates as it is used in a minimization or molecular dynamics run. Cartesian coordinates change, but connectivity and molecular specifications do not.

PSF mode works well for unusual ligands for which no RTF exists. The difficulty of PSF mode is that with large numbers of atoms, such as in proteins, it can be difficult to locate and identify a problem in the PSF.

When you use CNX in QUANTA, QUANTA generates a PSF for CNX. You do not need to communicate RTF information to CNX. If you want to do this, you can use CNX in standalone mode.

PSF files for CNX include additional information that is supported by CNX: alternate location indicator, chain name, and element name.

Setting up a protein in PSF mode

This section describes how to set up a protein file in PSF mode using a PDB file as the source of the starting coordinates. Although importing a PDB into QUANTA is straightforward, successfully completing the import task does not ensure that a file can be used immediately in PSF mode. This example is based on a simple case of a single polypeptide chain of 20 standard amino acids and water as the solvent.

The process includes these steps:

1. Organizing and separating segments.
2. Adding polar hydrogens.
3. Calculating a CHARMm energy.
4. Saving changes.

Organizing and separating segments

A segment name is not particularly important in the PDB file, but in QUANTA, CHARMm, and CNX the segment is a fundamental part of the data structure. This makes it critical that the different parts of the system be partitioned appropriately into segments.

QUANTA contains an Atom Property Editor that you can use to easily manipulate various aspects of the molecular structure file including segments.

For this example, use the Atom Property Editor to assign one segment (PROT) to contain the polypeptide and a second segment to contain all the water molecules (SOLV). Segment names can be any four-letter combination, but all segments must be uniquely named.

For more information on the Atom Property Editor, see *QUANTA Basic Operations.*

Adding polar hydrogens

You can add polar hydrogens to a structure in QUANTA in several ways: using CHARMm functions in RTF mode, using the tools on the Protein Editor, or using the QUANTA Molecular Editor.

Using CHARMm functions in RTF mode

Use this mode when RTFs are available for all the residues in your system. For a list of RTFs that come with QUANTA, see “Appendix D, Residue Topology Files” in *QUANTA Basic Operations*. The files are located in the $CHM_DATA directory.

After you set CHARMm to RTF mode, you can add polar hydrogens by selecting **Hbuild** from the **CHARMm** menu or by requesting a **CHARMm Energy** from the Modeling palette. One advantage of adding polar hydrogens in RTF mode is that optimization of water hydrogens is done automatically.

Using the Protein Editor

The Protein Editor provides tools for atom typing, bond assignments, and adding and deleting atoms and other tools to modify a molecule. The tool **Use Polar Hydrogens** adds polar hydrogens to a molecule based on the assigned atom types. After you add polar hydrogens, exit the Protein Editor, saving the changes. You may notice that all water hydrogens have the same orientation. This is because orientation has not been optimized.

Calculating a CHARMm energy

If you are not already working in RTF mode, set the CHARMm mode to RTF, request a CHARMm energy, and note the reported energy value. Set the **CHARMm mode** to **PSF**, set **Bonding** to **Special Protein Algorithm**, and pick **Disallow Bonding Between Segments**. Request a CHARMm energy in PSF mode. The energy should be identical to RTF mode.

Saving changes

After an energy is returned, use **Save As** from the QUANTA **File** menu to update the stored connectivity and bond order. This is the only way to get QUANTA to update without going into the Molecular Editor.

Problems

QUANTA detects a problem if some atoms have too many bonds. You can deal with this situation in several ways:

- Make sure you are using the special protein bonding algorithm.

- Delete offending bonds using **Delete Bond** from the QUANTA Molecular Editor palette.
- Optimize hydrogen positions in RTF mode.
- Minimize the protein in RTF mode.

Experiment to find the simplest solution that solves your problem. Then save the MSF and perform a calculation in PSF mode.

F. X-Ray - X-BUILD Command Conversions

This appendix correlates equivalent commands in the newer X-BUILD applications and the old X-Ray tools palette.

The style of the X-BUILD interface is more advanced than the X-Ray interface, so many of the tools do not map directly to an equivalent tool.

Map management

All map management is now global to QUANTA, so that general changes such as contouring can be made directly from the Maps Management table, and the complete list of options is available from the Map dialog box.

The Maps Management table is accessed using the **Map table** command on the **DRAW** menu. The Maps Management table can also be opened using the keyboard command **map table show**.

The Map dialog box is opened with the keyboard command **maps**.

Specific command conversions

X-Ray	X-BUILD
	Options... dialog box
Display map on atoms Display map in Volume Display map at pointer	These three commands are replaced by the display of the map (of radius defined by the user) at the current display position. The position can be defined by any of atom name, atom coordinate, bones pointer position, or text label. It is possible to set the map to be displayed only on the atoms, using the map management table tool **Contour/options**, although this is *not* recommended within the X-BUILD functionality, as it can result in false bones and related problems. You should refer to Chapter Managing Maps (page 25) for information on the limitations on the use of map display options and the X-AUTOFIT:X-BUILD functionality.
Neighbor cutoff [*value*]	X-BUILD/Options.../Show bumps less than [*value*]
Bump Cutoff [*value*]	X-BUILD/Options.../Show bumps less than [*value*]
VDW Cutoff [*value*]	*Cannot be manually altered*
Electrostatic Cutoff [*value*]	*Cannot be manually altered*
Spin Increment [*value*]	*Cannot be manually altered*
Extra map radius	X-BUILD/Options.../Map radius [*value*]
Sphere radius	*Defined by above*
[X] Mask map to cover only atoms	Map management table/Contour/Options
Cover radius [*value*]	Map management table/Contour/Options
[X] Display torsion values	*Automatically displayed— Cannot be manually altered*
[X] Pick residues for regularization	BUILD/Build atoms.../Regularize
[X] Prompt for Regularization Protocol	X-BUILD/Options.../<*regularization options*>
[X] Edit Torsion definitions	X-BUILD/Build atoms/Add-delete/Define torsions
[X] Hydrogen bond display	X-BUILD/Build atoms/Hydrogen bonds
Symmetry coloring	X-BUILD/Color table...
Rotamer Library	X-BUILD/Options....
Display Sphere	X-BUILD is based on the use of a display sphere of map, (plus atom display and bones) at a point in space. The use of a display sphere is implicit within the X-AUTOFIT:X-BUILD functionality
Automatically display map	X-BUILD *always* automatically displays the map. For general map control, you should use the map management table, where the individual map levels can be turned on and off.
Display symmetry atoms	X-BUILD always displays the symmetry atoms. For general control, refer to the object management table, where the visibility of the symmetry atoms can be turned on and off.

X-Ray	X-BUILD
Display all	Control the number of atoms displayed with the option **X-BUILD/ Options.../Atom radius**. If the radius of atom display is 1000 or greater, then all atoms are displayed, that is, no test is made on their position. Any radius less than 1000 results in a test on the radius of atoms to display about the current center point.
	Display Options...
Label atoms	X-BUILD/Text... (the text editor)
Label Residues	X-BUILD/Text... (the text editor)
Cα packing diagram	X-BUILD/Symmetry/CA packing diagram
Pack current display	X-BUILD/Symmetry/packing diagram
Display cell symmetry	X-BUILD/Symmetry/Filled unit cell
Cell bounds	X-BUILD/Symmetry/Unit cell
Display map table	EDIT/Map table/Show
Manage Maps...	Use the maps palette, accessible using the keyboard command **maps**
Move atom	X-BUILD/Build atoms/Move atom
Move fragment	X-BUILD/Build atoms/Move zone
	X-BUILD/Build atoms/Model first-last 4 res
	X-BUILD/Build atoms/Move atom + reg zone
Torsions	X-BUILD/Build atoms/Edit chi angles
	X-BUILD/Build atoms/Edit backbone tor
	X-BUUID/Build atoms/Model first-last 4 res.
	X-BUILD/Build atoms/Flip torsion 180 deg.
Bumps	X-BUILD/Build atoms/Show residue bumps
Cut residue	X-BUILD automatically cuts the chain on any edit. This tool therefore has no equivalent.
Model residue Side chain Rotamer	X-BUILD/Build atoms/Geometric conformation
Break bond	X-BUILD/Build atoms/add-delete/Delete bond
Make bond	X-BUILD/Build atoms/add-delete/Create bond
Recalculate bonds	X-BUILD has no equivalent: the bonds are always calculated, as necessary, after any editing. Otherwise it is possible to use the **EDIT** menu option **Bond options**
Regularize	X-BUILD/Build atoms/Regularize
	X-BUILD/Build atoms/Move atom + reg res.
	X-BUILD/Build atoms/Move atom + reg zone
Constrain atoms	X-BUILD/Build atoms/...Fix atoms
	X-BUILD/Build atoms/...Clear fixed atoms
	X-BUILD/Build atoms/...Fit in stages

X-Ray	X-BUILD
Fragment database	X-BUILD/structure/Fragment fitting
Set up CNX	Main **Applications** menu, **CNX Interface**
Undo changes	X-BUILD/Undo last fit The undo in X-BUILD is slightly different from the **Undo** under X-RAY. The undo in X-BUILD re-reads the packed data file and checks for changes from the most recently saved changes. Hence, the X-BUILD equivalent will undo multiple changes up to the last saved change.
Save changes	X-BUILD/Save changes Saves to the local packed datafile that is used for the undo option and a recover option. X-BUILD/Save build atoms to MSF Writes the data to file as a MSF file and can therefore be re-read at any time, outside of the X-BUILD application.
Reject changes	X-BUILD/undo last fit

G. Extend

Where to find the program

The extend program can be found in the directory $HYD_MAP.

Introduction

The extend program is a new program based on the existing mbkall program but with the addition of a set of command line arguments. It is designed for the conversion of maps in various formats to map brick format, but it also allows for the extension of maps to a different volume and the generation of compressed O format files for map masks.

The original functionality of mbkall required the following command line arguments:

```
mbkall ccp4 filein.map fileout.mbk b
```

Where:

1. ccp4 is the original map format.
2. filein.map is the input file name.
3. fileout.mbk is the output file name, i.e., mbk file (the output file type binary/integer/real — binary is the most usual).

The additional functionality uses additional command line arguments (beginning with a "-"), some with parameters. Only the first letter is required for the arguments, since there are no conflicts in the naming convention. If no default is show in square brackets ([ ]), then none is provided.

-extend

Extend the map to the limits defined using the -l,-r,-p,-m arguments. No extension of the map occurs unless this argument is provided, and if no limits are defined by other parameters, then the limits are a full cell. If this argument is used, then the -g argument *must* be specified as well.

-group (space group)

Used to define the space group. You should provide a valid space group defined in the symmetry file.

-symops (file) [$HYD_LIB/symlib.sym]

Used to define the full filename for the symops file. The default should be correct if QUANTA has been set up to run.

-limits (xmin xmax ymin ymax zmin zmax)

Defines the limits of extension using a real space box from “xmin” to “xmax”, “ymin” to “ymax”, and “zmin” to “zmax”. Six numbers must be given in the order defined above.

-radial (x y z r)

Defines the limits of extension using a real space sphere, from a point and radius. Four numbers must be given.

-pdb (pdb-file)

The limits of extension are defined by the minimum and maximum bounds of all atoms in the given PDB file.

-msf (msf-file)

The limits of extension are defined by minimum and maximum bounds of all atoms in the given MSF file.

-input

“Extend” to the same limits as input map

-boundary (value [0.0])

This defines a boundary beyond the limits generated from a PDB/MSF file. It would be normal to use a boundary of 4 Å about a molecule for extension.

-oformat (filename)

Generate an O compressed output file type. This action is independent of the extend functionality and can be used in conjunction with any other argument. This argument should be used where the input file is a map mask and allows the conversion of mask from various formats into the O compressed format which will allow it to be read into **QUANTA/X-AUTOFIT/map masks/**.

-cutoff (cutoff [0.5])

Valid only with the -o option, this defines the cutoff that defines the inside/outside of a mask. Default = 0.5

Examples

To extend a map round a set of MSF coordinates.

```
extend ccp4 mapin.map mapout.mbk b -e -m myfile.msf -b 5 -g r3
```

To generate a map mask from a CCP4 map file

```
extend ccp4 mask.map mask.msk b -o
```

General comments on the program

The program when carrying out map extension dynamically allocates memory on run time. The memory required is defined by the volume of the input map and the volume of the output map. If more memory is requested than is available, the program aborts with a memory error.

The program actually runs faster in extend mode (to small volumes — i.e., about a single molecule) than the standard map conversion. This is because in extend mode, the input and output maps are in memory.

There is no requirement about the starting map asymmetric unit and the final position of the extended map; the former and latter need not overlap and need not be in the principle unit cell.

Index

A

Accept Fragment tool 339
Accept Point 58
Add alternative conformation, Add/delete… 176
Add atom at pointer, Add/delete… 177
Add helix/strand tool 144
Add mask at pointer tool 138
Add menu space tool 200
Add res. at termini, Add/delete… 176
Add water at pointer, Add/delete… 177
Add/delete… palette 169
…all 20 conformations tool 239
all atom model
    CA trace 39
all-atom representation 70
alpha carbon
    next 62
alpha-carbon building 58
alpha-carbon geometry
    probability map 61
amino acid
    refining coordinates 94
Anneal Fragment tool 339
Applications menu
    X-SOLVATE 246
arrows
    indicating polarity 63
Atom Selection tool 192
atoms
    missing or incorrect in .GSD file 95

B

B- conf clamped to A backbone option 205
bones
    overview 52
    symmetry 56
Bones on/off tool 132
Bones palette 52, 132
    Bones on/off 132
    Calc bones symmetry 136
    Change start value 135
    Change trim/analysis 136
    Create bones atoms 135
    Delete 1 section 134
    Delete fragments 134
    Find nice area of map 133
    Main side 133
    Map quality from bones 133
    Set bones/RSR map 132
    Side chains on/off 133
    Smooth bones 133
    Symmetry off 136
    Undo last delete 135
bones points
    deleting 56
Bones setup palette
    Change bones start value 54
    Change trim parameters 55
bones strand
    changing type 56
bones trees 53
branch trimming parameters 55
brick map
    maximum size 41
Build atoms palette 162
    …fix atoms 173
    Add/delete… 169
    Color atoms… 173
    Edit atom info 173
    Edit backbone tor 165
    Edit chi angles 166
    Fit main chain by RSR 164
    Fit side chain by RSR 163
    Geometric conformation 162
    Model first/last 4 res. 168
    Move atom 167
    Move atom + reg. res. 170
    Move atom + reg. zone 171

Move atom + RSR 164
Move zone 167
Mutate residue 169
Refine 1 residue 162
Regularize 172
Save changes 174
Show residue bumps 174
Undo last fit 174
By Protocol…, Do all 185

C

Cα angle/torsion graph 61
CA BUILD palette
Reverse chain 143
CA Build palette 142
Add helix/strand 144
Check CA direction 145
Current res/seg 143
Delete current CA 143
Delete fitted seg 148
Fit seg by CA correlation 147
Fit seg by D.E.E. 148
Fit seg by database 147
Fit seg by RSR 146
Join 2 segments 144
Load CA coordinates 142
Move current seg 145
Next bones box 142
Next CA 142
Refine current seg 145
Save changes 149
Show points ~3.8 Å 146
Undo last build 148
Unjoin 2 CA 144
CA Build, X-AUTOFIT
X-BUILD palette 63
Cα trace
all atom model 39
building an all-atom representation 70
Cα tracing
alpha-carbon
building 39
Calc bones symmetry tool 136
Calc. mask from bones tool 137
Calc. mask from coord tool 137
Calculation radius tool 251
Calculations tool 193
carbon building
assisted 58
Cα-tracing
overview 57
CCP4 file
importing 28
Center at Ramachandran point option 206
Change all, Do all 184
Change bones start value 54
Change search parameters… , X-LIGAND 234
Change search setting tool 247, 249
Change start value tool 135
Change trim parameters 55
Change trim/analysis tool 136
CHARMm energy
calculating 348
calculation modes 345
Check CA direction 63
Check CA direction tool 145
Check for voids tool 139
clamping 97
Clear all tools tool 200
Clear sequence tool 160
CNX 253
Color atoms… palette 179
Color B-alt difference 180
Color by (ρ) 180
Color by atom 179
Color by B-value 180
Color by occupancy 180
Color by progress 179
Color atoms… tool 173
Color B-alt different tool 180
Color by (ρ) 180
Color by atom tool 179
Color by B-value tool 180
Color by occupancy tool 180
Color by progress tool 179
Color table tool 203, 211
Column function tool 192
Conf search mode, Ligand serach parameters
dialog box 235

Conformation search tool 238
conformations
    clamping 97
Coordinate radius option 210
Create bones atoms tool 135
Create peptide link, Add/delete palette 175
C-terminal oxygen 98
Current res/seg 143
Current res/seg tool 160

D

Define NCS symmetry tool 124
Define new tools tool 199
Define symmetry tool 124
Define-Edit group tool 199
Del mask at pointer tool 138
Del. Current col. tool 191
Del. Current row tool 191
Del. Table tools 191
Delete 1 section tool 134
Delete current CA 59
Delete current CA tool 143
Delete current segment tool 143
Delete fitted seg 148
Delete fragments tool 134
Delete Map from List tool 29
Delete peptide link, Add/delete… 174
Delete range, Add/delete… 176
Delete residue, Add/delete… 176
Delete sequence tool 159
Delete text tool 140
Delete water tool 250
density skeletonization 38
dial box 47
Differences tool 195
Dipeptide restraint option 209
disorder 97
display
    updating center 79
Display All Fragments tool 338
Display Next tool (Search Fragment Database) 338
Display Previous tool (Search Fragment Database) 339
Display radius of map tool 251
dmprep program 331
Do all dialog box 184
DSN6 file
    importing 28

E

Edit atom info tool 173
Edit backbone tor tool 165
Edit chi angles tool 166
Edit position tool 242
Edit torsions tool 241
electron density maps
    determining quality 53
    displaying 52
Exit
    from X-LIGAND 243
External program palette 217
    keywords and other details 217
external programs 85, 217

F

Find Negative density tool 202
Find nice area of map tool 133
Finish tool, X-AUTOFIT 204
Fit by D.E.E. tool 187
Fit main chain by RSR tool 164
fit ratio 63, 146
Fit seg by CA correlation 147
Fit seg by D.E.E. 148
Fit seg by database tool 147
Fit seg by RSR 146
Fit side chain by RSR tool 163
Fitting/Continue search tool 183
…fix atoms tool 173
Fix Validate error tool 141
fragment
    database 335
    search process 336
    searching example 340
    selection criteria 336
fragment database
    creating 331
fragments

searching 2, 335
Frodo map file
importing 28
fuzzy residue types 18

G
Geometric conformation tool 162
Go to pointer tool 128
Goto defined text tool 141
Goto next text tool 140
Goto previous text tool 140
gradient body refinement 94
Graph window 47
GRID file
importing 28
Grid search parameters / Max conf to search,
Ligand search parameters 236
.GSD file
incorrect atoms 95
Guess next CA 59
Guess next CA tool 146

H
Hide this menu tool 190
hydrogen
representations 97
Hydrogen representation option 207

I
importing
map files 28
Increase resolution/ Decrease resolution tool 139

J
Join 2 segments tool 64, 144

L
Label graph tool 198
Last commands tool 204
List Proteins tool 337
List Residues tool 339
Load CA coordinates tool 142
Load Property tool 141
Load sequence tool 158
Loop fit & Terminal fit tool 183

M
Main side tool 133
Main ↔ side 56
Map
quality 133
map
change current 44
contouring 37
delete 29
determining quality 53
displayed upon startup 122
improving quality 53
optimizing 53
search for water peaks 245
map files
importing 28
Map mask palette 137
137
Add mask at pointer 138
Calc. mask from bones 137
Check for voids 139
Del mask at pointer 138
Increase resolution/Decrease resolution 139
Mask delete radius 138
Mask dials 138
Mask on 137
Read mask from file 139
Save mask to file 139
Solvent content 138
Map on/off tool 237, 249
Map Quality 53
map quality
adjusting branch trimming parameters 55
getting information 53
improving 53
modifying the skeletonization initial cut-off parameter 54

Map quality from bones tool 133
map radius 55
    Cα-bulding 57
Map radius option 210
mask
    resolution 51
    size 49
Mask delete radius tool 138
Mask dials tool 138
Mask off tool 137
Max distance to protein tool 251
memory allocation
    errors 42
memory requirements 41
memory usage 42
Min density level tool 250
Min distance to protein tool 251
Min sep between peaks tool 251
model building
    general tips 99
Model first/last 4 res. tool 168
Molecular coordinates
    used by X-SOLVATE 245
molecule
    current editable 44
Move atom + reg. res. tool 170
Move atom + reg. zone tool 171
Move atom + RSR tool 164
Move atom tool 167
Move current seg tool 145
Move zone tool 167
MSF
    saving 204
Mutate residue tool 169

N

New segment 58
New segment tool 143
New text at pointer tool 140
…next conformation tool 239
Next bones box 59
Next bones box tool 142
Next CA 58
Next CA tool 142
Next CA, X-AUTOFIT
    X-BUILD palette 62
Next peak 247
Next peak tool 249
Next site tool 238
Non-bonding in regularization option 209
center, screen
    See display 79

O

open angle 58
Options… tool (Search Fragment Database) 340
oxygen
    C-terminal 98

P

palette
    behavior 45
Peak search parameters dialog box
    accessing 250
    Calculation radius 251
    Display radius of map 251
    Max distance to protein 251
    Min density level 250
    Min distance to protein 251
    Min sep between peaks 251
    Temperature factor 251
Phi/Psi restraint option 208
Pick Alpha Carbon Range tool 337
Pick Alpha Carbon tool 338
Pick Density palette 58
    Accept Point 58
Place at next residue tool 128
Place by atom tool 129
Place using bones tool 129
Place using coord tool 129
Plot data tool 196
Pointer dials tool 128
Pointer palette 127
    Go to pointer 128
    Place at next residue 128
    Place by atom 129
    Place using bones 129

Place using coord 129
Pointer dials 128
Show pointer 127
Suggestions 130
polarity
evaluating 63
reverse 143
polarityreversing 63
Position search center tool 237, 249
Postscript settings tool 197
…previous conformation tool 239
Previous peak 247
Previous peak tool 249
Previous site tool 238
Protein property tool 193
Protein validate 200
pseudo-Ramachandran plot 58, 59
using 60
PSF mode
CHARMm 345
setting up protein 347

Q

QUANTA graph facilities 61

R

Radii / Calculation radius, Ligand search parameters 237
Radii / Map display radius, Ligand search parameters 237
Ramachandran plot 82
identify point 82
Read file tool 190
Read mask from file tool 139
Read molecule tool 191
real space refinement 71
recover
last building session 44
Refine 1 residue tool 162
Refine all water tool 186
Refine current seg tool 145
Refine tool 242
Refine zone tool 182
refinement
RSR and gradient body 94
regularization 95
compared to RSR 96
disulfides 96
Regularize across disulfide bonds option 210
Regularize tool 172
Reject Fragment tool 339
renin
model example 340
Renumber sequence ID, Add/delete… 179
Re-patch terminal, Add/delete… 178
residues
defining coordinates 335
Resolution factor option 211
Return fuzzy sequence tool 160
Reverse chain tool 143
Rigid body fit tool 181
Rotamer library option 204
RSR 94
compared to regularization 96
RSR side chain
improving results 94
RTF mode
CHARMm 345
Run external program tool 204

S

Save all peaks tool 248, 250
Save as water tool 248, 250
Save built atoms to MSF tool 203
Save built atoms to PDB tool 203
Save built CA to MSFtool 203
Save changes 149
Save changes tool, Build atoms palette 174
Save ligand to MSF tool 242
Save mask to file tool 139
Saving MSFs 204
search
for fragments 336
Search Database tool 338
Search for ligands tool 234
Search for peaks palette 247
accessing 248

Change search setting 247, 249
Delete water 250
Exit 250
Map on/off 249
Next peak 247, 249
Position search center 249
Previous peak 247, 249
Quit 250
Save all peaks 248, 250
Save as water 248, 250
Search for waters 247, 248
Search for waters 247
Search for waters tool 248
Search Fragment Database palette 337
List Proteins 337
Search threshold, Ligand search parameters dialog box 234
segment
changing polarity 63
editing
Cα atoms
editing 59
evaluating polarity 63
fitting a sequence 65
generating 58
segments
joining 64
Select Display… tool (Search Fragments Database) 339
Select new ligand tool 242
Select Torsions palette
Accept Fragment 339
Anneal Fragment 339
Display Next 338
Display Previous 339
List Residues 339
Options… 340
Pick Alpha Carbon 338
Pick Alpha Carbon Range 337
Reject Fragment 339
Search Database 338
Select Display 339
Undo All 338
Undo Last 338
sequence
aligning 65
sequence alignment
finding unique 69
generating information 66
sequence alignment algorithm 66
sequence coordinates
building 71
Sequence palette 158
Clear sequence 160
Current res/seg 160
Delete sequence 159
Load sequence 158
Return fuzzy sequence 160
Show/Hide Amino acids 161
Show/Hide fuzzy 161
Show/Hide sequence 160
Undo last change 161
Unique sequence 160
Set active torsions tool 241
Set bones/RSR map tool 132, 202
Set up bones parameters dialog box 54, 135
Show bumps less than 3.00 option 206
Show bumps to inactive molecules option 206
Show pointer tool 127
Show points ~ 3.8 Å tool 146
Show residue bumps tool 174
Show/Hide Amino acids tool 161
Show/Hide fuzzy tool 161
Show/Hide sequence tool 160
Side chains on/off tool 133
skeletonization 38
adjusting 54
adjusting branch trimming parameters 55
algorithm 52
changing the type of a bones strand 56
deleting bones points 56
modifying the initial cut-off parameter 54
Smooth bones tool 133
Solvent content tool 138
Start parameter
adjusting 54
Start Value dial 54
Structure palette 181
Do all 184
Fit by D.E.E. 187

Fitting/Continue search 183
Hide this menu 190
Loop fit & Terminal fit 183
Refine all water 186
Refine zone 182
Rigid body fit 181
Undo last fit 188
Suggestions tool 130
swap memory
low 43
swap space
recommended 41
symmetry information 84
Symmetry off tool 136
Symmetry palette 123
124, 125
Define NCS symmetry 124
Symmetry radius option 210

T

tables and graphs 85
fixed atom/residue size 89
general table 88
graphs and plotting 90
inconsistent data 89
picking table data 87
scratch table 89
table 1 87
table 2 88
table contents 87
table data 86
Tables and Graphs palette 190
Atom Selection 192
Calculations 193
Column function 192
Del. Current col. 191
Del. Current row 191
Del. Table 191
Differences 195
Label graph 198
Plot data 196
Postscript settings 197
Protein property 193
Read file 190
Read molecule 191
Temperature factor tool 251
Ten-Eyck files
importing 28
Text palette 140
Delete text 140
Fix Validate error 141
Goto defined text 141
Goto next text 140
Goto previous text 140
Load Property 141
New text at pointer 140
torsion angle 58
torsion angles
undefined residues 95
Trim parameters dialog box 55
typographical conventions 2

U

Undo All tool 338
Undo last 56
Undo last build tool 148
Undo last change tool (Sequence palette) 161
Undo last delete tool 135
Undo last fit tool 174, 188
Undo Last tool 338
Unique sequence tool 160
Unit cell tool 125
Unjoin 2 CA tool 144
User Defined tools palette
Add menu space 200
Clear all tools 200
Define new tools 199
Define-Edit group 199

V

Validate tool 200
validation techniques 85
VSN6
importing 28

W
water molecules
searching for 247

X
X-AUTOFIT
accessing 43
disorder 97
exiting 46
file format conversion 49
graph windows 47
overview 49, 105
preparing to use 43
running external programs 85
starting 43, 121
tables 48
X-BUILD main palette
Color table 203
Find Negative density 202
Last commands 204
Run external program 204
Save built atoms to MSF 203
Save built atoms to PDB 203
Save built CA to MSF 203
Set bones/RSR map 202
Validate 200
X-AUTOFIT Options dialog box 202
X-BUILD palette
Join 2 segments 64
X-AUTOFIT dials 54
X-AUTOFIT Options dialog box 204
B-conf clamped to A backbone 205
Center at Ramachandran point 206
Coordinate radius 210
Dipeptide restraint 209
DNA build mode 205
Hydrogen representation 207
Map radius 210
Next residue step 207
Non-bonding in regularization 209
Number of DP 207
Number of steps for regularize 208
Phi/Psi restraint 208
Protein build mode 205
Regularize across disulfide bonds 210
Resolution factor 211
Return main dials after edit 205
Rotamer library 204
Search time limit 208
Show bumps less than 3.00 206
Show bumps to inactive molecules 206
Symmetry picking mode 207
Symmetry radius 210
X-AUTOFIT sessions
information retained between 122
X-AUTOFIT•X-BUILD palette
Delete current segment 143
Main side 133
New segment 143
Side chains on/off 133
X-BUILD
command logs 91
controlling display 78
last commands 91
modality 128
running external programs 85
tables 48
tables and graphs
see also tables and graphs
validation 141
X-Build
features overview 94
X-BUILD validation techniques 40, 85
X-FIT palette
Delete current CA 59
Delete section 56
Guess next CA 59
Main side 56
Map Quality 53
New segment 58
Next bones box 59
Next CA 58, 62
X-LIGAND
crystallographic protocol 229

description of process 232
exiting 243
MSF file requirements 231
requirements 230
tools, masked and unmasked 233
X-LIGAND palette 234
…all 20 conformations 239
…next conformation 239
…previous conformation 239
Change search parameters 234
Conformation search 238
Edit position 242
Edit torsions 241
Map on/off 237
Next site 238
Position search centre 237
Previous site 238
Refine 242
Save ligand to MSF 242
Search for ligands 234
Select new ligand 242
Set active torsions 241
X-PLOR
troubleshooting 264
X-POWERFIT
file format conversion 49
Xsight file
importing 28
X-SOLVATE
accessing 246
adjusting water molecules 248
preparing to use 246
saving water molecules 248
searching for water molecules 247
start-up 246
use of molecular coordinates 245