Cryo-EM Model-Building with Modern Coot

Paul Emsley

Diamond Icknield, April 2019
Model-building with Modern Coot

• Overview:
  • Refinement & Restraints
    • Local distance restraints (ProSMART)
  • Backrub Rotamers
  • cis-peptides
  • carbohydrate
  • “Pink stick, green pea”

• Recent updates:
  • CURLEW
  • GLSL, Ribbons, VR, AR.
A New Branch for 0.9-pre

- from which a new binary is built
- ~1200 revisions ahead of the master branch
- This new branch is actively developed
- Uses C++11
- Less stable
- More interesting
- Not always clear how to use the new functions
- (or that the new functions are there)
- Limited binary distribution from the Coot web site
  - in the “experimental” directory
A New Branch

- New multi-threaded sections:
  - Jiggle-Fit
  - Refinement
    - Target function and derivative evaluation, model and map all happen simultaneously now
    - Which means: more atoms, smoother updates and/or closer to the minimum
- All-atom contact dots
- Ramachandran Score
  - $\phi, \psi$ hypothesis scoring
- Rotamer Score
- Add Terminal Residue
- Crankshaft Peptide Optimisation
  - simultaneous evaluation of $\kappa_1, \kappa_2, \ldots$ solutions
ProSMART/Local Distance Restraints

- Use previous-solved “template” structures to inform the refinement of the (low resolution) target protein
- by local-distance restraint generation
“ProSMART-like” Restraints

• Instead of using a reference model, more often I use “Self” restraints
  – which can be calculated internally
  – the starting model is the “reference” from which the ideal distances are calculated
  – message to the refinement:
  – “keep the local environments similar to how they were when you started”

• The minimizer in Coot is 1st a order (derivative) based method
  – “Jelly body” stabilizer cannot work
Local Distance Restraints: Prior Structure

Model for High Resolution data
Local Distance Restraints: Today's Structure
Local Distance Restraints:
Structure Comparison

Model for High Resolution data
Model for Low Resolution data
Local Distance Restraints:
ProSMART/Geman McClure Restraints
Modified Target Function

Penalty

Distance \( (r_o - r_{\text{target}}) \)
Additional Restraints in 0.9-pre

- Auto “on the fly” Helical restraints
- Strands not yet added
  - They will be
Rotamers:
Peptide Backbone Geometry
Rotamers

• Rotamers are preferred configurations of a side-chains rotatable bonds
  – where “preferred” means these configurations occur more frequently in a set of reference protein structures
  – “preferred” because they are low-energy conformations
• Several Rotamer “databases” exist
  – (Son of) Penultimate Rotamer Library
4 PHE Rotamers
New Low Resolution Rotamer Search

After Fitting Tools in KING/Molprobity
Traditional

Backrub
Map Properties

Cell and Symmetry:
Cell: 64.90 78.32 38.79
90.00 90.00 90.00
Spacegroup: P 21 21 21 [P 2ac 2ab]

Displayed Map Style:
- Standard Lines
- Solid/Transparent
- "Cut-Glass"
Opacity (%): 50.0

Contouring:
Contour Level:
Set Level: 0.56
- absolute
- rmsd

Map Histogram:

Contour Level Step Size:
- r.m.s.d. step

Map Colour

Skeleton:
- On
- Off
A number of papers have been published recently highlighting the unusually large number of cis-peptides in some structures:

- Croll: The rate of cis-trans conformation errors is increasing in low-resolution crystal structures Acta Cryst. (2015). D71, 706-709
cis-Peptides

trans-peptide

cis-peptide

PRO trans-peptide

PRO cis-peptide
cis-Peptides

\[ \text{trans-peptide with plane restraints} \]

\[ \text{cis-peptide with plane restraints} \]

\[ \text{trans-peptide with plane and trans restraints} \]
cis-peptide Representation
The New Ramachandran Plot

In Preferred Regions: 172 (93.99%)
In Allowed Regions: 9 (4.92%)
Outliers: 2 (1.09%)
CURLEW: Coot Utilites and Refinement Library Extention Wrangler

- Easy access to ”interesting” Coot scripts
N-linked Carbohydrates

- Improved algorithm and re-worked GUI
Problematic Glycoproteins

- Crispin, Stuart & Jones (2007)
  - NSB Correspondence
  - “one third of entries contain significant errors in carbohydrate stereochemistry...”
  - “carbohydrate-specific building and validation tools capable of guiding and construction of biologically relevant stereochemically accurate models should be integrated into popular crystallographic software. Rigorous treatment of the structural biology of glycosylation can only enhance the analysis of glycoproteins and our understanding of their function”
  - PDB curators concur
- More recently Joosten & Lütteke (2017), Agirre et al. (2017)
Problematic Glycosylation

- In the case of carbohydrates, their inherent complexity [and] conformational flexibility [] are causing massive experimental problems which hinder the determination of the exact tertiary structures of these biomolecules
  - Engelsen et al. (2014) “Biopolymers”
Carbohydrate Links

Thomas Lütteke (2007)
Linking Oligosaccharides/Carbohydrates: LO/Carb

- One can fully define carbohydrate structure by the primary structure and a set of torsion angles
- Build complex carbohydrate structure
  - from a dictionary of standard links
  - and monomers
  - torsion-angle refinement
    - by simulated annealing
α 1,6 Link
α 1,6 Link
Refinement Progress
(NAG-ASN example)
N-linked Glycan Modelling in Coot

- Two Modes:
  - Residue by Residue
    - User Control
  - Whole Tree Addition
    - Automated
    - Embedded decision-making
Refinement Stabilizers

- Using Coot's Real Space Refinement
  - (in default mode)
  - allowed saccharides to result in twisted or boat ring conformations
    - even without user intervention
- Unimodal Torsion Restraints
- Pseudo-plane restraints
- Inter-residue Geman-McClure external distance restraints
Unimodal Torsions and Pseudo Planes

New Unimodal Ring Torsion Restraints
External Distance Restraints

- Using unimodal distances from prior-know glycan crystal structures
  - c.f. ProSMART for protein models
  - here we use intent to use a consensus model rather than a particular model
N-linked Carbohydrate
N-linked Carbohydrate

But building in cryo-EM maps is not so easy
Representation
Modules → Multi-Sharpen...
High Resolution

● Apoferritin from Relion 3 paper
  – Wim Hagen
EMD-0144: 1.65 Å Human Apoferritin Relion-3

Note: No waters, no Hydrogen atoms in the model
EMD-0144:
1.65 Å Human Apoferritin Relion-3

Human sample, model from horse: the differences were easily identified. Note also: non-optimized scattering factors (seen on acid side chains).
1.65 Å Human Apoferritin Orientation for next slide
Red/Green is a difference map calculated without Hydrogen atoms being in the model
Hydrogen Atom Density?

- Little to no consistent density for the test points (these are the controls)
- Substantial density at the HA position
**Coot is (or has historically been) Ugly**

- For the most part, I've spent my time
  - putting atoms into density
  - working with ligands
- The menus needs some re-working
- But most of all the graphics are tired-looking and slow
  - recent transparent objects and animation are an improvement
  - ignoring the developments in computation
Interactive Rotamer Goodness
Multi-Criteria Markup
Rotamer and Ramachandran Markup

- `set_show_intermediate_atoms_rota_markup(1)`
- `set_show_intermediate_atoms_rama_markup(1)`
Coot Futures: GPU Ribbons

with Martin Noble
Coot Futures: Virtual Reality

Hamish Todd

- An Intuitive Interface:
- Stereoscopic Representation
- Greater Field of View
- 2 Hands with Articulation

However:
- current tools are not immediately transferable
- because: nausea
CootVR

- Demonstrated at CCP-EM Meeting in Keele in April
• Augmented-Reality *Coot*?
Acknowledgements

- LMB:
  - Garib Murshudov, Rob Nicholls,
  - Alexey Amunts, Alan Brown

- Kevin Cowan, Bernhard Lohkamp

- Libraries & Dictionaries:
  - Jane & Dave Richardson
  - Alexei Vagin
  - Eugene Krissinel
CootVR Example Video

<switch>