Model-Building with Coot

- Overview of Presentation:
  - New algorithms and features
  - Tips on Usage
- Focus on:
  - Tools for Ligand-handling
  - Tools for Low Resolution X-ray data
  - Tools for EM
  - Future Directions
Feature Integration

Refinement

Validation

Validation, Model Building and Refinement should be used together

e.g. REFMAC

Internal

External

e.g. MolProbity

Internal

Internal
Real Space Refinement

- Major Feature of Coot
  - Gradient-based minimiser (BFGS derivative)
  - Geometry library is the standard CIF-based Refmac dictionary
    - Minimise deviations in bond length, angles, torsions, planes, chiral volume, non-bonded contacts
      - Including links and modifications
- Provides “interactive” refinement
- Subject to substantial extension since the start

The first attempt

Student Reaction:

“Oh, I don't look at that window...”
(I maximise the window immediately)
Representation of Results:

Second attempt...

Student Reaction:

“Oh, box of meaningless numbers.

Go away”
Representation of Results: “Traffic Lights”

“Traffic Lights” represent the RMSd values for each of the refined geometry types.

Good refinement

Bad refinement
Refinement Techniques

- Single-Atom Drag
  - Over-dragging
- Key-bindings:
  - Triple Refine
  - Single Residue Refine with Auto-accept
- Ramachandran Refinement
Ligands in Coot

- Importing and building ligand from scratch
  - PRODRG, LIBCHECK
- Ligand Fitting
- Validation
  - Mogul
- Representation
  - Bond orders
  - Surfaces
- Analysis
  - Molprobity, LIDIA
2D Ligand Builder

- Free sketch
- Sbase search
"Yesterday's" Ligand

- Atom name matching
- Torsion matching
- Ligand overlay
Why Validate?

- 11,000+ chemical structures (Het-groups)

Sulfate ions in 1DW9 (1.65Å resolution)
<table>
<thead>
<tr>
<th>Ligand Type</th>
<th>Ligand Site</th>
<th>Known</th>
<th>Unknown</th>
</tr>
</thead>
<tbody>
<tr>
<td>Known</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Cocktail</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Unknown</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Cocktail Examples:
- ![Chemical Structure](image1)
- ![Chemical Structure](image2)
- ![Chemical Structure](image3)
- ![Chemical Structure](image4)
- ![Chemical Structure](image5)
- ![Chemical Structure](image6)
- ![Chemical Structure](image7)
Conformer Generation
```plaintext
loop_
  _chem_comp_tor.comp_id
  _chem_comp_tor.id
  _chem_comp_tor.atom_id_1
  _chem_comp_tor.atom_id_2
  _chem_comp_tor.atom_id_3
  _chem_comp_tor.atom_id_4
  _chem_comp_tor.value_angle
  _chem_comp_tor.value_angle_esd
  _chem_comp_tor.period

<table>
<thead>
<tr>
<th>Compound</th>
<th>Variable</th>
<th>Atom 1</th>
<th>Atom 2</th>
<th>Atom 3</th>
<th>Atom 4</th>
<th>Value Angle</th>
<th>ESD</th>
<th>Period</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADP</td>
<td>var_1</td>
<td>O2A</td>
<td>PA</td>
<td>O3A</td>
<td>PB</td>
<td>60.005</td>
<td>20.000</td>
<td>1</td>
</tr>
<tr>
<td>ADP</td>
<td>var_2</td>
<td>PA</td>
<td>03A</td>
<td>PB</td>
<td>O1B</td>
<td>59.979</td>
<td>20.000</td>
<td>1</td>
</tr>
<tr>
<td>ADP</td>
<td>var_3</td>
<td>O2A</td>
<td>PA</td>
<td>&quot;05'&quot;</td>
<td>&quot;C5'&quot;</td>
<td>-59.942</td>
<td>20.000</td>
<td>1</td>
</tr>
<tr>
<td>ADP</td>
<td>var_4</td>
<td>PA</td>
<td>&quot;05'&quot;</td>
<td>&quot;C5'&quot;</td>
<td>&quot;C4'&quot;</td>
<td>179.996</td>
<td>20.000</td>
<td>1</td>
</tr>
<tr>
<td>ADP</td>
<td>var_5</td>
<td>&quot;05'&quot;</td>
<td>&quot;C5'&quot;</td>
<td>&quot;C4'&quot;</td>
<td>&quot;C3'&quot;</td>
<td>176.858</td>
<td>20.000</td>
<td>3</td>
</tr>
<tr>
<td>ADP</td>
<td>var_6</td>
<td>&quot;C5'&quot;</td>
<td>&quot;C4'&quot;</td>
<td>&quot;04'&quot;</td>
<td>&quot;C1'&quot;</td>
<td>150.000</td>
<td>20.000</td>
<td>1</td>
</tr>
<tr>
<td>ADP</td>
<td>var_7</td>
<td>&quot;C5'&quot;</td>
<td>&quot;C4'&quot;</td>
<td>&quot;C3'&quot;</td>
<td>&quot;C2'&quot;</td>
<td>-150.000</td>
<td>20.000</td>
<td>3</td>
</tr>
</tbody>
</table>
```
Ligand Torsionable Angle Probability from CIF file
Conformer Generation

Non-Hydrogen
Non-CONST
Non-Ring
Orienting the Ligand

Orientation 1

Orientation 2

Orientation 3

Orientation 4
Orienting the Ligand

Orientation 1

Orientation 2

Orientation 3

Orientation 4
Ligand Validation

- Compare the observed structure to the restraints
  - In *Coot*: Validate → Geometry analysis
Parmatisation issues... (what if they are wrong?)

- Perfect refinement with incorrect parameters → distorted structure

- CSD's Mogul
  - Knowledge-base of geometric parameters based on the CSD
  - Can be run as a “batch job”
  - Mean, Z-scores.
Ligand Validation

- Mogul plugin in Coot
  - Run mogul, graphical display of results
  - Update restraints (target and esds for bonds and angles)
Ligand Representation

- Bond orders (from dictionary)
Ligand Representation

- Surfaces using Partial Charges
Ligand Representation

- Transparent surfaces
Binding mode analysis

- Binding site highlighting,
- Isolated Molprobity dots
Ligand Environment Layout

- 2d Ligand pocket layout (ligplot, poseview)

Can we do better? - Interactivity?
Ligand Environment Layout

- Binding pocket residues
- Interactions
- Substitution contour
- Solvent accessibility halos
- Solvent exclusion by ligand
Solvent Exposure

- Identification of solvent accessible atoms
Ligand Environment Layout

- Considerations
  - 2D placement and distances should reflect 3D metrics (as much as possible)
    - H-bonded residues should be close to the atoms to which they are bonded
  - Residues should not overlap the ligand
  - Residues should not overlap each other
  (work in progress)
Ligand Environment Layout

- Initial residue placement
Layout Energy Terms

\[ E = \sum \sum w_{ij} \left( d_{ij}^2 - D_{ij}^2 \right) + \]
\[ \sum \sum \exp\left(-\frac{1}{2}d_{ij}^2\right) + \]
\[ \sum \sum \left( d_{ik}^2 - D_{ik}^2 \right) + \]
\[ \sum \sum \exp\left(-\frac{1}{2}d_{ik}^2\right) \]

- Residues match 3D Distances
- Residues don't overlay each other
- Residues are close to H-bonding ligand atoms
- Residues don't overlap ligand
Ligand Environment Layout

- Residue position minimisation
Determination of the Substitution Contour

How far can we go (in the direction of the hydrogens) before hitting atoms of the protein?

Determined in 3D, project to 2D surface and contoured

\[ \text{c.f. Clarke & Labute (2007)} \]
Substitution Contour: Extending along Hydrogens

Riding Hydrogens

Torsionable Hydrogens
(test multiple directions)
Chiral Centre Inversion

Inverted chiral centre refinement pathology detection

Hydrogen tunnelling
Alpha Helix Placement

- **Scenario:** Looking at a new map, not built with automatic tools:
  - “I can see that there’s a helix here - build it for me!”
- **From a given point:**
  - Move to local averaged maximum
  - Do a 2D MR-style orientation search on a cylinder of electron density
  - Build a helix (both directions)
  - 1D Rotation search to find best fit
  - Score based on density at CB positions
  - Trim ‘n Grow
Centering the Rotation point
Cylinder Search

- Pick the orientation that encapsulates the most electron density

Using 2 rotation axes
2 x 1-D Helix orientation searches
Top

Bottom

baubles
Handling EM maps
Partitioning Maps: Watershed Algorithm

New region created

Different segments
EM Futures

- Use Fast Fourier Feature recognition
  - Positioning protein domains or fragments in complete or segmented maps
  - Extremely parallelizable(??) and non-graphical
  - Map scoring with protein-protein docking hypotheses?
    - Score complexes with PISA?
Low Resolution Model-Building

- Interactive Map Sharpening
- User-define distance restraints
  - And torsion restraints
- Ramachandran Restraints
- “Backrub” rotamers
- Protein-db loop database (from Kevin Cowtan)
Sharpening the Data

Average Amplitude $|F|$ vs. Resolution

- Original Data
- Sharpened Data
Sharpening the Data

Which B-factor shall I use to get the most interpretable map?

Interactively adjust the structure factor amplitudes and re-generate the map with FFT and recontouring...
Successfully read coordinates file coot-download/pdb2xgi.ent. Molecule number 3 created.
2XGJ: Mtr4, Weir et al. (2010)
"Backrub Rotamers"

- High probability models with low resolution data
Current Low Resolution Rotamer Search

Previous

Rotamer Search

+ Rigid Body Refinement
New Low Resolution Rotamer Search

After Fitting Tools in KING/Molprobity
To turn it on...

- (ROTAMERSEARCHLOWRES)
To turn it on...

- (ROTAMERSEARCHLOWRES)
Networking...

- PDBe interface...
- Drag and drop
  - Also with drugbank
PDBBe Recent Structures

JSON parser, network threaded code
Finding Holes

- An implementation of
  - Smart, Goodfellow & Wallace (1993) Biophysics Journal 65, 2455
  - Atomic radii from AMBER
  - I used
    - radii from CCP4 monomer library
    - sans simulated annealing
Restraints Editing in Coot

- Distance Restraints:
  - Alpha helices, A-form RNA
- Add and delete individual restraints
  - User-selectable sigma
- Select 2 residues for range
- User-defined torsion restraints
- Input from ProSMART
- Output to Refmac
Export as Refmac

Restrains:

- EXT DIST FIRST CHAIN A RESI 55 INS . ATOM CA SECOND
  CHAIN A RESI 55 INS . ATOM C VALUE 1.54 SIGMA 0.05

- (And Nat Echols has been working a the phenix interface)
ProSMART Interface

- <demo>