Tuesday, October 26\textsuperscript{th}, 2006

Low Resolution Refinement (with Refmac5)

Tim Grüne

The Shelx Tuesday Seminar Series

http://shelx.uni-ac.gwdg.de
Electron Density at Atomic Resolution

PDB code 1GWE 0.9Å resolution at 4σ

Atomic resolution \( d < 1.2\text{Å} \)
- peaks in electron density map show single atoms.
Electron Density at High Resolution

PDB code 1AQB 1.7Å resolution at 1.6σ

High resolution \( d < 1.8\text{Å} \)
- mostly contiguous density, no separate atoms any more
- aromatic rings show holes
- multiple conformations can / have to be modelled
Electron Density at Medium Resolution

PDB code 1E7I  2.7Å resolution at 1.2σ

**Medium resolution** $d < 3.0Å$

- secondary structure elements
- major part of side chains still visible and distinguishable
- no or very few multiple conformations
Electron Density at Low Resolution

PDB code 1JL4 4.0Å resolution at 1.2σ

Low but usable resolution $d < 5\text{Å}$
- secondary structure elements detectable (especially $\alpha$–helices)
- smaller features (side chains) “buried” in density
The actual Problem: Data to Parameter ratio

Approximate data/parameter ratio depending on resolution

<table>
<thead>
<tr>
<th>Res.[Å]</th>
<th>parameters</th>
<th>data/parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.0</td>
<td>x,y,z</td>
<td>0.9:1</td>
</tr>
<tr>
<td>2.3</td>
<td>x,y,z; B</td>
<td>1.5:1</td>
</tr>
<tr>
<td>1.8</td>
<td>x,y,z; B</td>
<td>3.1:1</td>
</tr>
<tr>
<td>1.5</td>
<td>x,y,z; B</td>
<td>5.4:1</td>
</tr>
<tr>
<td>1.5</td>
<td>x,y,z; U_{11}U_{12}U_{13}U_{23}U_{22}U_{33}</td>
<td>2.4:1</td>
</tr>
<tr>
<td>1.1</td>
<td>x,y,z; U_{11}U_{12}U_{13}U_{23}U_{22}U_{33}</td>
<td>6.1:1</td>
</tr>
<tr>
<td>0.8</td>
<td>x,y,z; U_{11}U_{12}U_{13}U_{23}U_{22}U_{33}</td>
<td>16:1</td>
</tr>
</tbody>
</table>

G. Sheldrick
Resolution? — A Counter Example

The importance of the data to parameter ratio is nicely demonstrated by


- Resolution: 30–3.0Å
- \( R/R_{\text{free}} \): 13.9%/ 19.9%
- Solvent content: 72%
- Unique reflections: 23882
- Multiplicity: 2.8
- Parameters: 19883
- Restraints: 27798
Example: Data to Parameter ratio

Scenario
ideal model: \( f(x) = x^2 \) (generally never known)

Experiment 1: High resolution, 21 data points with errors
Experiment 2: Low Resolution, 3 data points with errors
Example: Data to Parameter ratio

Two Models

Model 1: $g(x) = g_2x^2 + g_1x + g_0$

Model 2: $h(x) = h_4x^4 + h_1x + h_0$

Both Models require three parameters
Example: Data to Parameter ratio

Fitting High Resolution Data

\[ g_2 = 1.19 \quad g_1 = 0.00 \quad g_0 = -0.51 \quad \chi^2 = 1.14 \]
\[ h_4 = 0.16 \quad h_1 = 0.52 \quad h_0 = 0.47 \quad \chi^2 = 19.3 \]

- Assuming errors are correct, \( \chi^2 \) makes a clear distinction between the two models.
- \( \chi^2 \) far from 1.0 in either case but distinguish well between the two models.
Example: Data to Parameter ratio

Fitting Low Resolution Data

\[ g_2 = 0.73 \quad g_1 = -0.00 \quad g_0 = 1.17 \]
\[ h_4 = 0.09 \quad h_1 = -0.49 \quad h_0 = 1.59 \]

- Both Models fit the data perfectly
- No error estimates because 
  \[#parameters = #data\]
- Additional knowledge is required to decide about the correct model
Example: Data to Parameter ratio

Fitting Low Resolution Data — Constrained

Constraint: data passes through \((0, 0)\)

\[ \Rightarrow \text{Model 1: } g(x) = g_2 x^2 + g_1 x \quad \text{Model 2: } h(x) = h_4 x^4 + h_1 x \]

- Model 1 quite close to reality (but we would never know that)
- \(\chi^2\) favours model 1. This requires realistic error estimates.

\[ g_2 = 0.94, \ g_1 = -0.15, \ \chi^2 = 1.30 \]

\[ h_4 = 0.13, \ h_1 = -0.97, \ \chi^2 = 3.00 \]
Restraints and Constraints

Small molecules at high resolution can be refined **unrestrained**.

Macromolecules are almost always refined by **restrained refinement**, i.e. additional information like ideal bond lengths and angles are taken into account.

**Constraints** *reduce* the number of parameters. These are expression like “Property X *must have* this value” — *e.g.* “temperature factor is isotropic”:

4 parameters per atom instead of 9 parameters per atom

**Restraints** *increase* the number of data. “Should be” or “should be approximately” expressions, *e.g.* distance $\Delta(N, C_\alpha) = 1.458\text{Å} \pm 0.019\text{Å}$. Modify the target function being minimised by the refinement program.
Refmac5 — Getting Started

Refmac5 available through ccp4i interface.
The Interface provides good control over most options.
Documentation can be found at $CCP4/html/refmac5.html.

Sometimes it is necessary to modify the script created by ccp4i.
This way some more options for Refmac5 become available.
Testing Parameters with the UNIX Shell

Create an executable script, `refmac5.sh`, say, by copy-and-paste from the GUI:

```bash
#!/bin/bash
VDW=$1
refmac5 xyzin pdb_input.pdb hklin mtz_input.mtz 
    xyzout pdb_result.pdb hklout mtz_result.mtz << eof
    refi -
        type REST -
        resi MLKF -
        meth CGMAT -
        bref ISOT
    solvent YES 
        VDWProb $VDW
    END
eof
```

And calling it with a list of values for the van–der–Waals probe radius:

```bash
#> for vdw in 1.10 1.15 1.20 1.25 1.30 ; do
  sh refmac5.sh $vdw > refmac5_vdw=${vdw}.log ;
done
```

or, with using several CPUs at once (Careful!):

```bash
#> for vdw in 1.10 1.15 1.20 1.25 1.30 ; do
  sh refmac5.sh $vdw > refmac5_vdw=${vdw}.log ;
done
```
Important Parameters for Refmac

1. The WEIGHT-ing term

- Determines the weight between reflections and restraints.
- A high value leads to distortions in case of a bad or incomplete model.
- Good results can be achieved with WEIGHT auto

Refinement Caveats
Important Parameters for Refmac

1. The WEIGHT–ing term
2. Generation of Hydrogens

- Placed at riding positions — no extra costs
- Improve bumping restraints, overall geometry, and contribute to $F_{\text{calc}}$
- Always use this option, be it with high or with low resolution data.

- Refinement
- Model Building
- Model Validation
Important Parameters for Refmac

1. The **WEIGHT**-ing term
2. Generation of Hydrogens
3. B–factor refinement: Isotropic or overall?

- Isotropic B–factor refinement: 4 parameters per atom, *i.e.* 9,000 parameters for an average 30kDa protein
- Overall B–factor refinement: 3 parameters, *i.e.* 6,500 parameters
- Apply overall B–factor refinement, if
  - number reflections $\approx 4 \times$ number atoms
  - isotropic refinement yields very high B–values

---

**Isotropic refinement of low resolution data**

![Graph showing isotropic refinement of low resolution data](image-url)
Important Parameters for Refmac

1. The \texttt{WEIGHT}–ing term
2. Generation of Hydrogens
3. B–factor refinement: Isotropic or overall?
4. TLS–Refinement (translation, libration, screw)
   - models domain movements within the monomer
   - relatively cheap: 20 parameters per domain
   - should make up for overall B–factor

Format of the tls–input file:

```
TLS dom1
range 'X 722' 'X 750' ALL
range 'X 780' 'X 796' ALL
TLS dom2
range 'X751' 'X 779' ALL
TLS dom3
range 'X 797' 'X 867' ALL
TLS dom4
range 'X 868' 'X 977' ALL
```
Importantly Parameters for Refmac

1. The **WEIGHT**-ing term
2. Generation of Hydrogens
3. B-factor refinement: Isotropic or overall?
4. TLS-Refinement
5. Adjustment of refinement cycles

Refinement programs may “overshoot” — run a certain number of cycles, then re-run and stop at appropriate cycle.
Build as much as possible

Refinement programs flatten the electron density map. At low resolution, this may obscure real features. Always build as much as possible before running a round of refinement. Especially after MR or SAD, where phases are of poor quality.
Model Validation — Ramachandran Plot

At low resolution outliers of the Ramachandran plot are difficult to justify.

Coot’s interactive Ramachandran window

N.B.: The Ramachandran plot of PROCHECK is considered out–dated
Model Validation — $R_{\text{work}}$ and $R_{\text{free}}$

During refinement:
$R_{\text{work}}$ drops with the addition of parameters.
$R_{\text{free}}$ drops only if these are an improvement of the model.

At the end of refinement:
the ratio between $R_{\text{work}}$ and $R_{\text{free}}$ should follow

$$\frac{R_{\text{free}}}{R_{\text{work}}} \approx \sqrt{\frac{\# \text{reflections} + \# \text{parameters}}{\# \text{reflections} - \# \text{parameters}}}$$

References


- Tronrud, D., *The Limits of Interpretation*, from his homepage