electron density maps

When the structure is solved structure factors for the model may be **calculated**:

$$F_{calc}(hkl) = |F_{calc}(hkl)|\, e^{i\alpha_{calc}(hkl)}$$

$$= \sum_{N} f_i(hkl) e^{2\pi i(hx+ky+lz)} e^{-4B_i \sin\ \Theta/\lambda^2}$$

α_{calc} - best guess for the phase,

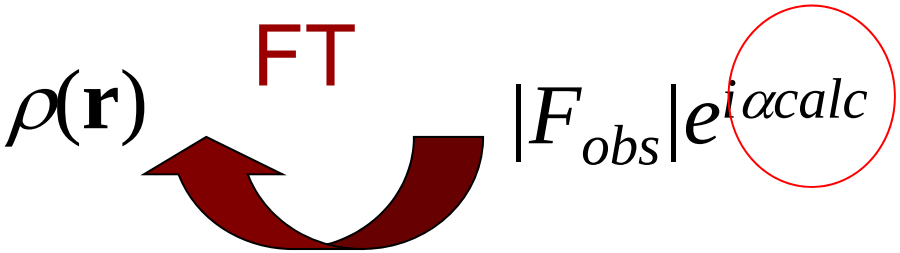
$|F_{obs}|$ – best guess for the amplitude

$$F(hkl) = |F_{obs}(hkl)|\, e^{i\alpha_{calc}(hkl)}$$

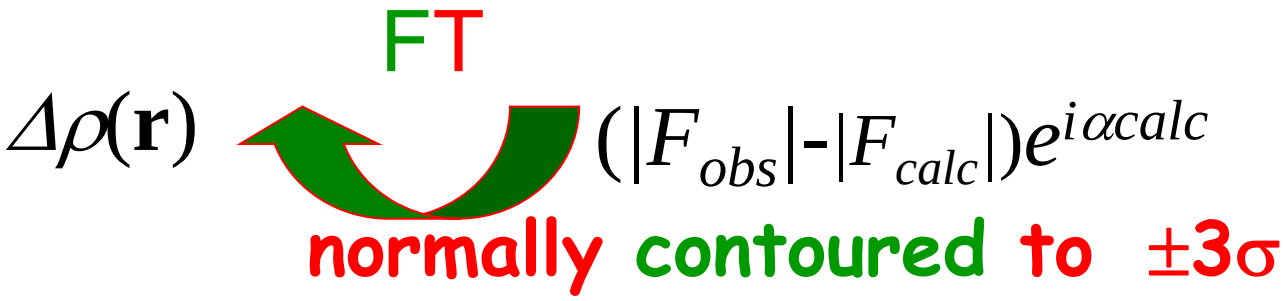
- best guess for a structure factor

the electron density is

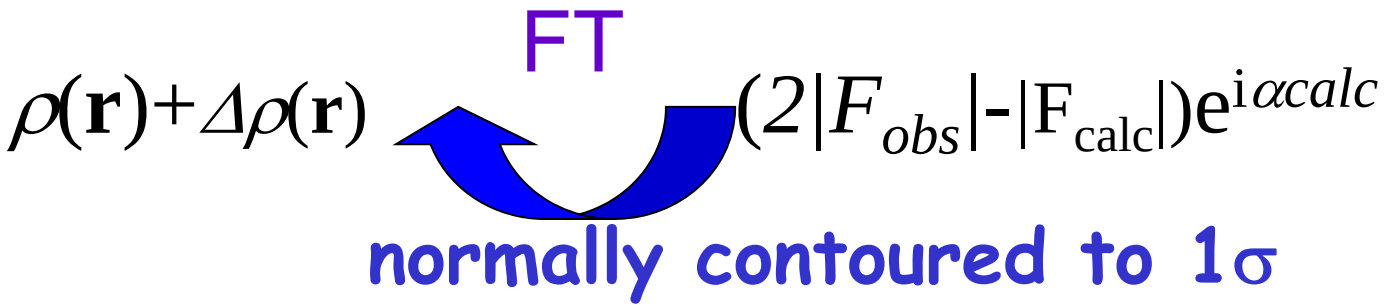
we approximate by



or (green/red in coot) – the difference density



or we use $2F_{obs} - F_{calc}$ density (blue in coot)



refinement

refinement of the model structure so that F_{calc} comes closer to F_{obs}

- Adjusting atomic positions and thermal parameters

- Hydrogen atoms normally omitted

refinement

- *R*-value – a quality parameter (an agreement index)
  - Definition:

$$R = \frac{\sum_{hkl} \| F_{obs} | - k | F_{calc} \|}{\sum_{hkl} | F_{obs} |}$$

- *R* –> 0 the more the observed and the calculated amplitude agree
- *R* is calculated for some *hkl 's* – it can be for all data or for a group of data

starting model

- **Molecular Replacement**
- **MIR**
- **MAD**

bad agreement between F_{calc} and F_{obs}
(R-value: 40 - 50%)

Typical data set to 2.5 Å resolution

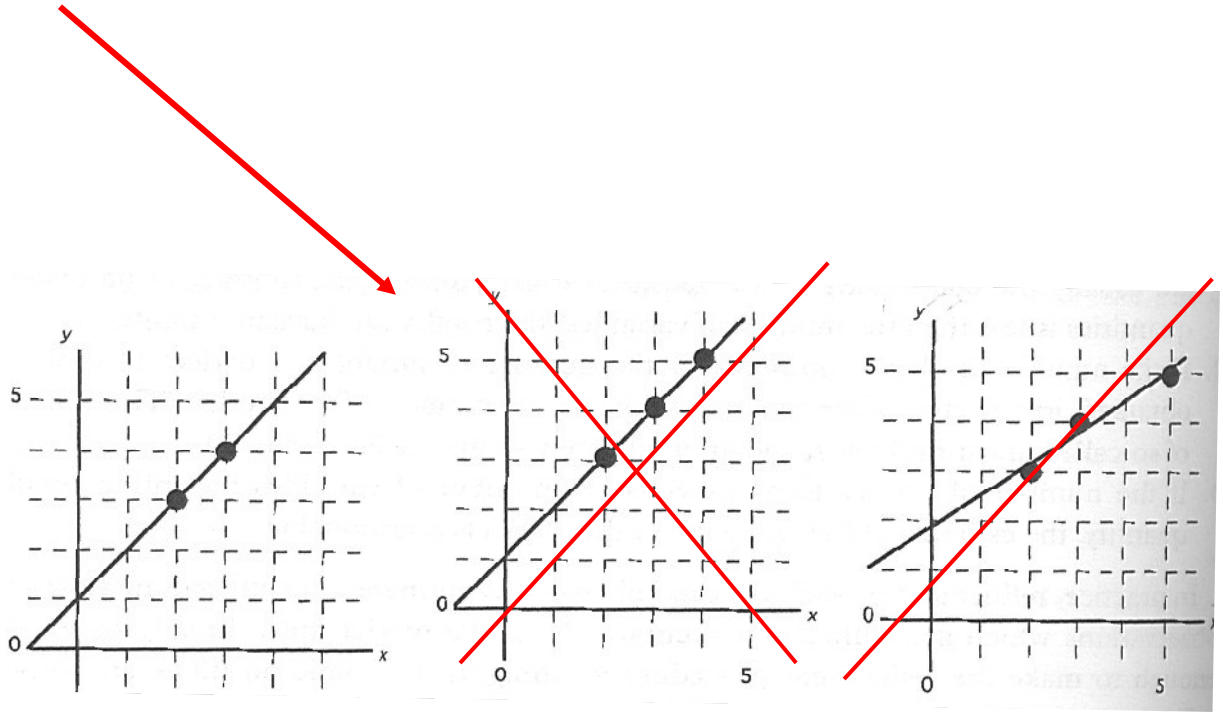
OMP decarboxylase:

- 28000 independent measurements
- 7000 atoms
  - Each have 3 positional and 1 thermal parameter

Refinement

- To adjust the structure, so it fits the measured data in the best way
- the model has four variables: x,y,z coordinates and the B-factor
- least squares method to minimize the differences between model and observations

- A certain number of observations are needed



2 measurements

3 measurements

Problems

- **data/parameter ratio is too small (2-3)**
  - **For OMP decarboxylase 1!**

Solutions

- **Subset of measurements as test set (R_{free})**

- **Restraints or constraints on model**

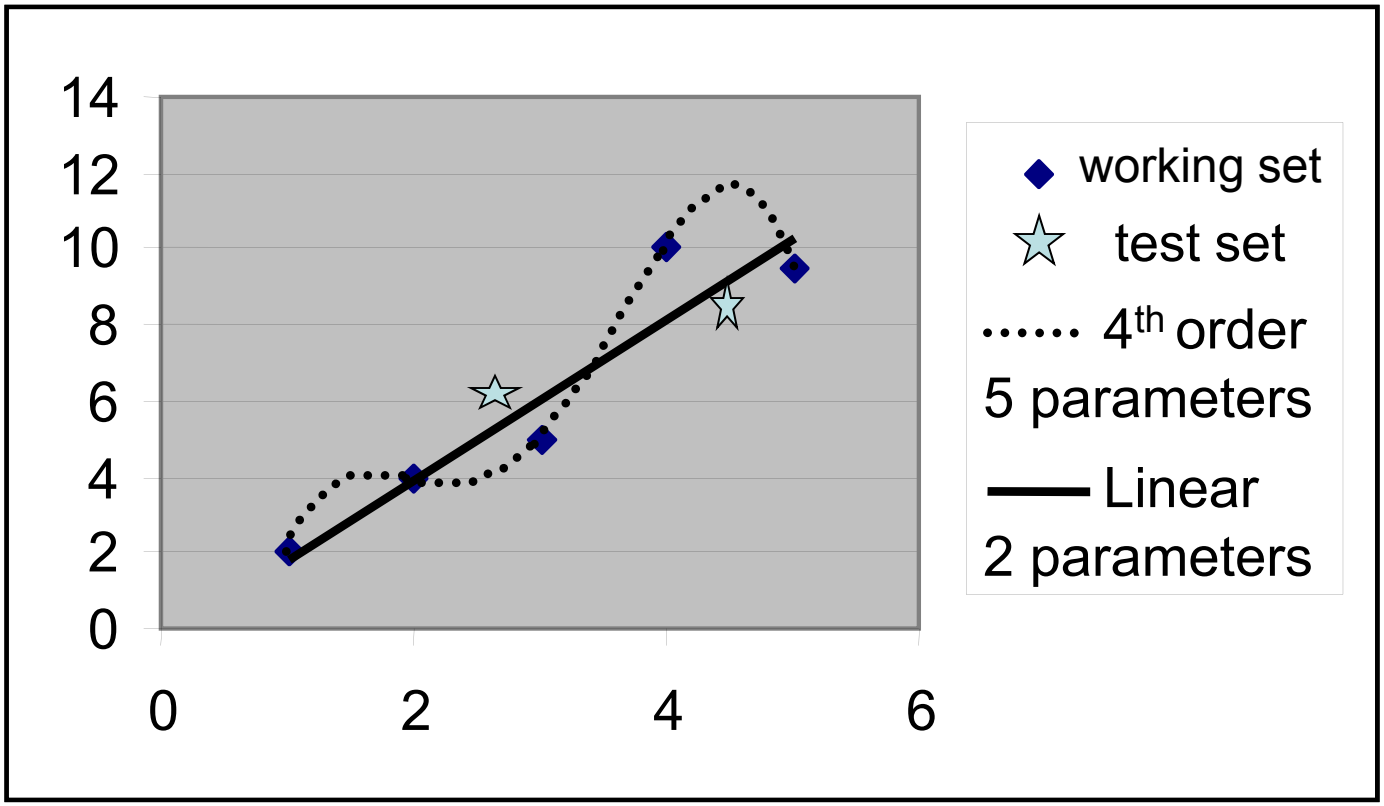
R_{free}

- it is critical to use the same data for optimisation and checking
- refinement procedure may make adjustments that lower R, but doesn't improve the model

- choose a small part of the reflections, that are not used for optimisation. From these data a new R-value is calculated – R_{free}
- R_{free} gives an unbiased validation of the model ($R_{free} \approx 1.2 \cdot R$)

$$R_{free} = \frac{\sum_{hkl,testset} ||F_{obs}| - k|F_{calc}||}{\sum_{hk,testset} |F_{obs}|}$$

1st and 4th order polynomial fit to 5 points illustrates R_{free} subset



Restrains/constraints

- **Restrains - used as additional observations**
- **Constraints – used to reduce the number of parameters**

what restrains?

- **Stereo chemical knowledge from small molecule structures**
  - **bond lengths and angles**
- **solvent correction**
  - **disordered flat solvent in the solvent channels**

non-crystallographic symmetry (NCS)

- **if any!**

delicate refinement using stereo-chemical restrains

Minimizing a function of both a crystallographic part and a stereochemical part

Q=crystallographic part $\Sigma w(|F_o|-|F_c|)^2$

+ *distances*
+ *planes*
+ *chiral centres*
+ *non-bonded distances*
+ *torsion angles*

- Consult all the time with the electron density maps

$2F_o$-F_c maps and F_o-F_c maps

with calculated **or** observed phases

refinement

- Refinement is a cyclic proces
- For each round it is important to investigate the new model for mistakes
  - investigate the (F_{obs} – F_{calc}) difference density map for large deviations
- When to stop?
- If there are still stuff that can be improved, and which leads to a better model, the refinement is not finished (R_{free} should usually come below 0,3)

strategy – good data

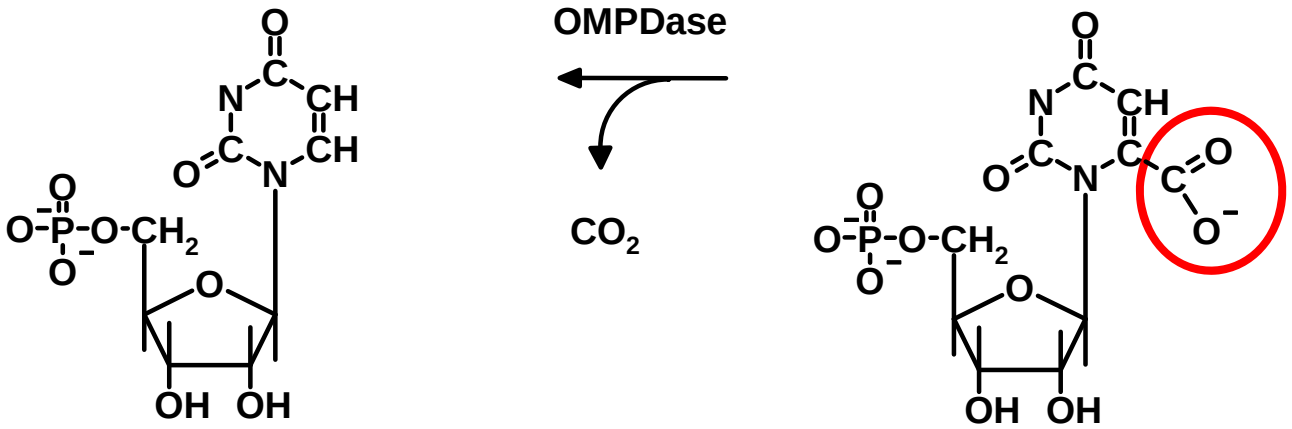
- four variables ok
- Position *(x,y,z)* and B-factor refinement

- Restrains and ncs used

strategy – very good data

- four variables might be to few
  - expand the model using anisotropic displacement parameters (6 instead of 1)
  - hydrogen atoms at ideal positions
- loosen the restrains

Example: ODCase



Uridine
5'-monophosphate

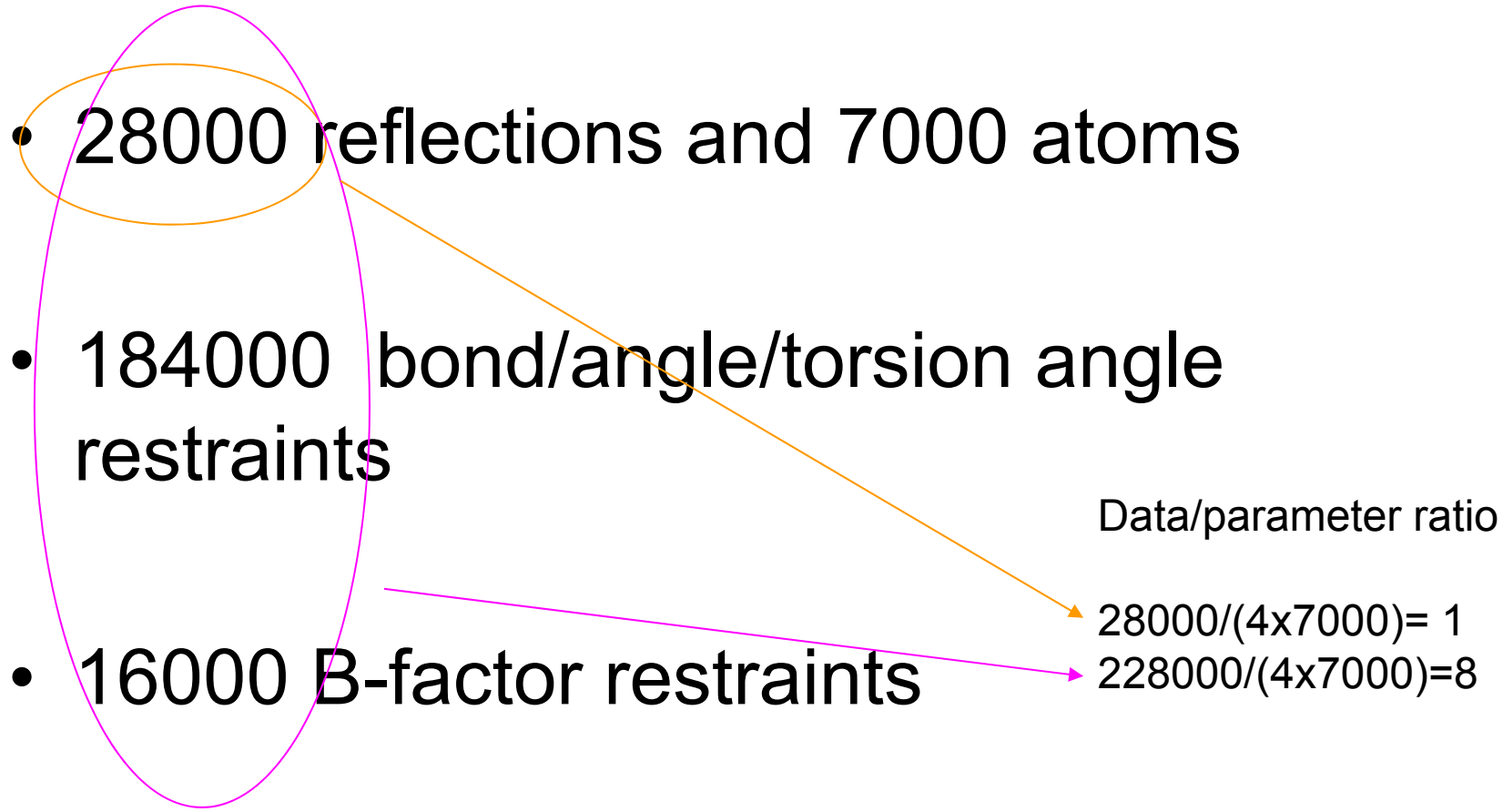
Orotidine
5'-monophosphate

ODCase

ENZYME:BMP COMPLEX

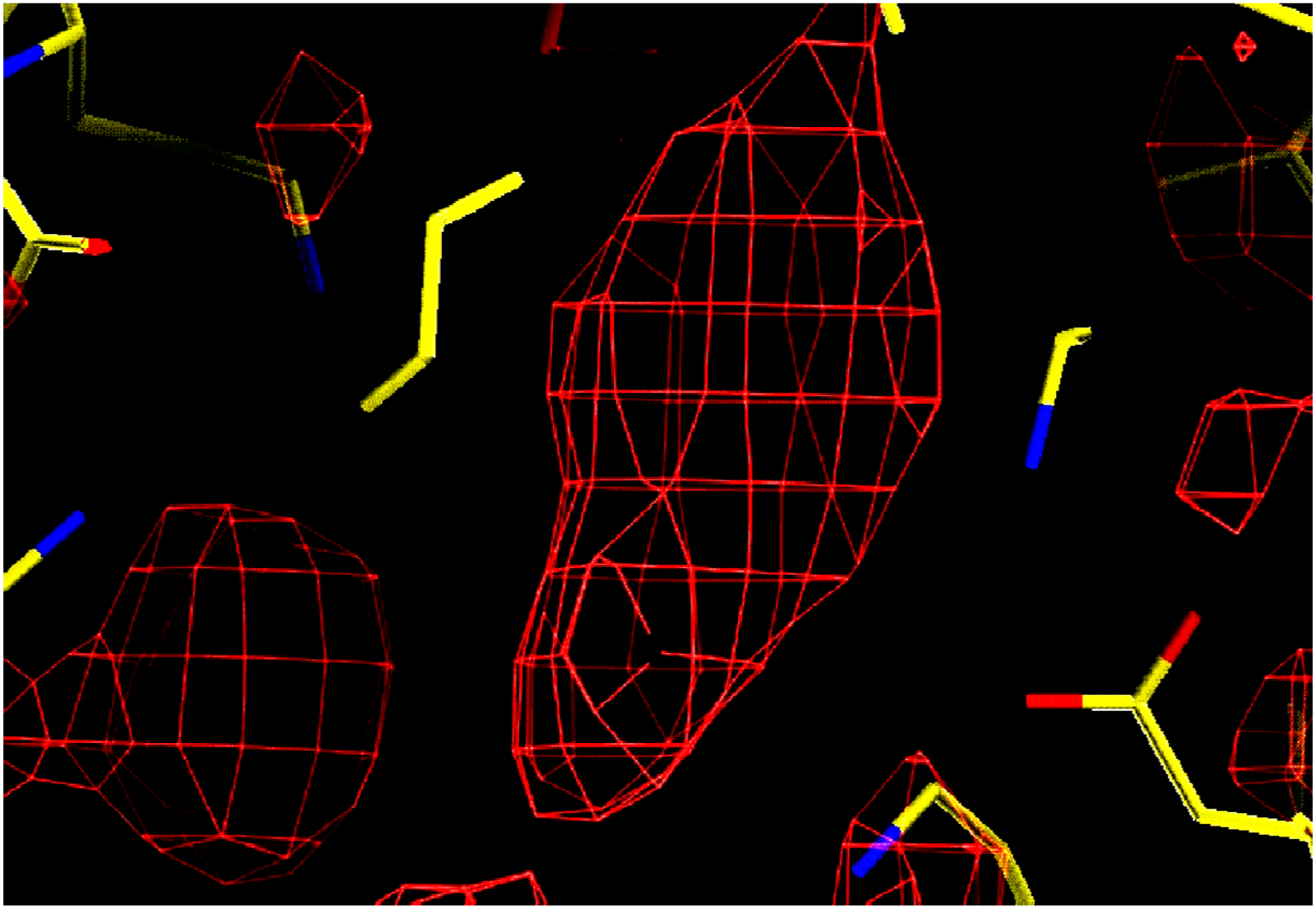
- $P2_12_12_1$
- 61 x 95 x 145 Å
- 2 DIMERS/ASU

- 28000 reflections and 7000 atoms
- 184000 bond/angle/torsion angle restraints
- 16000 B-factor restraints
  - 4 ncs related molecules

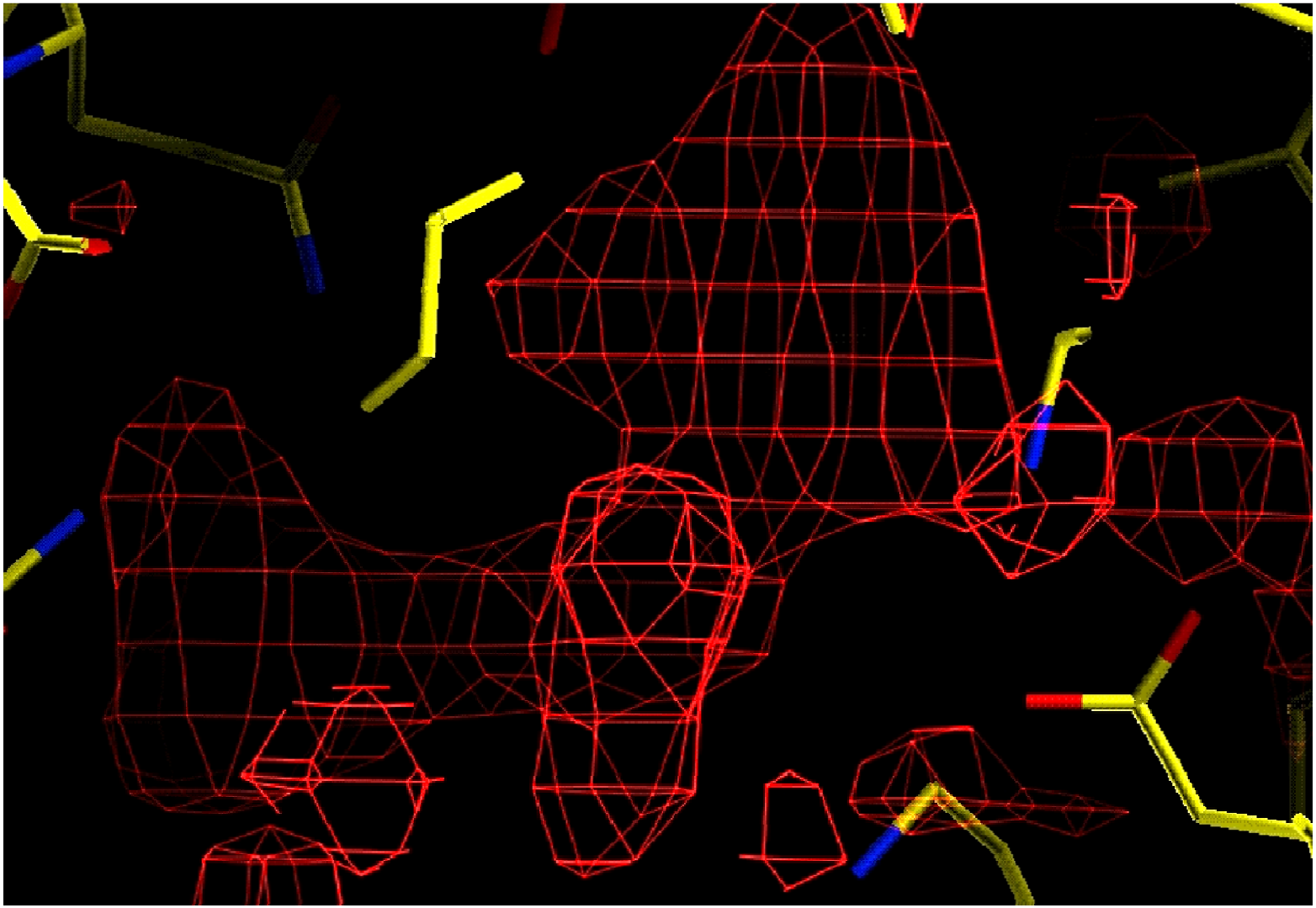


active site - difference density (3.1Å)

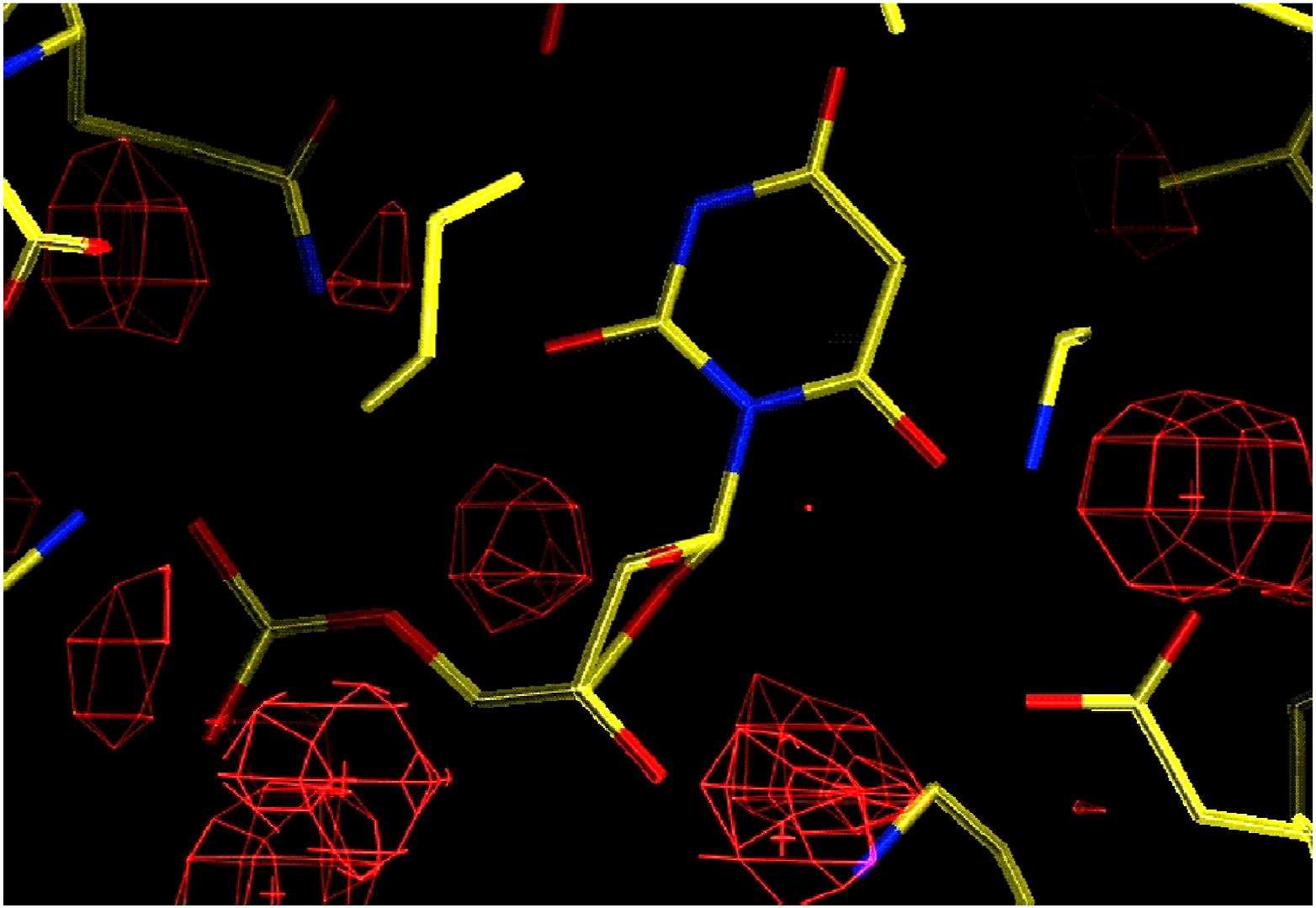
NB: reverse colour coding: red POSITIVE, green NEGATIVE



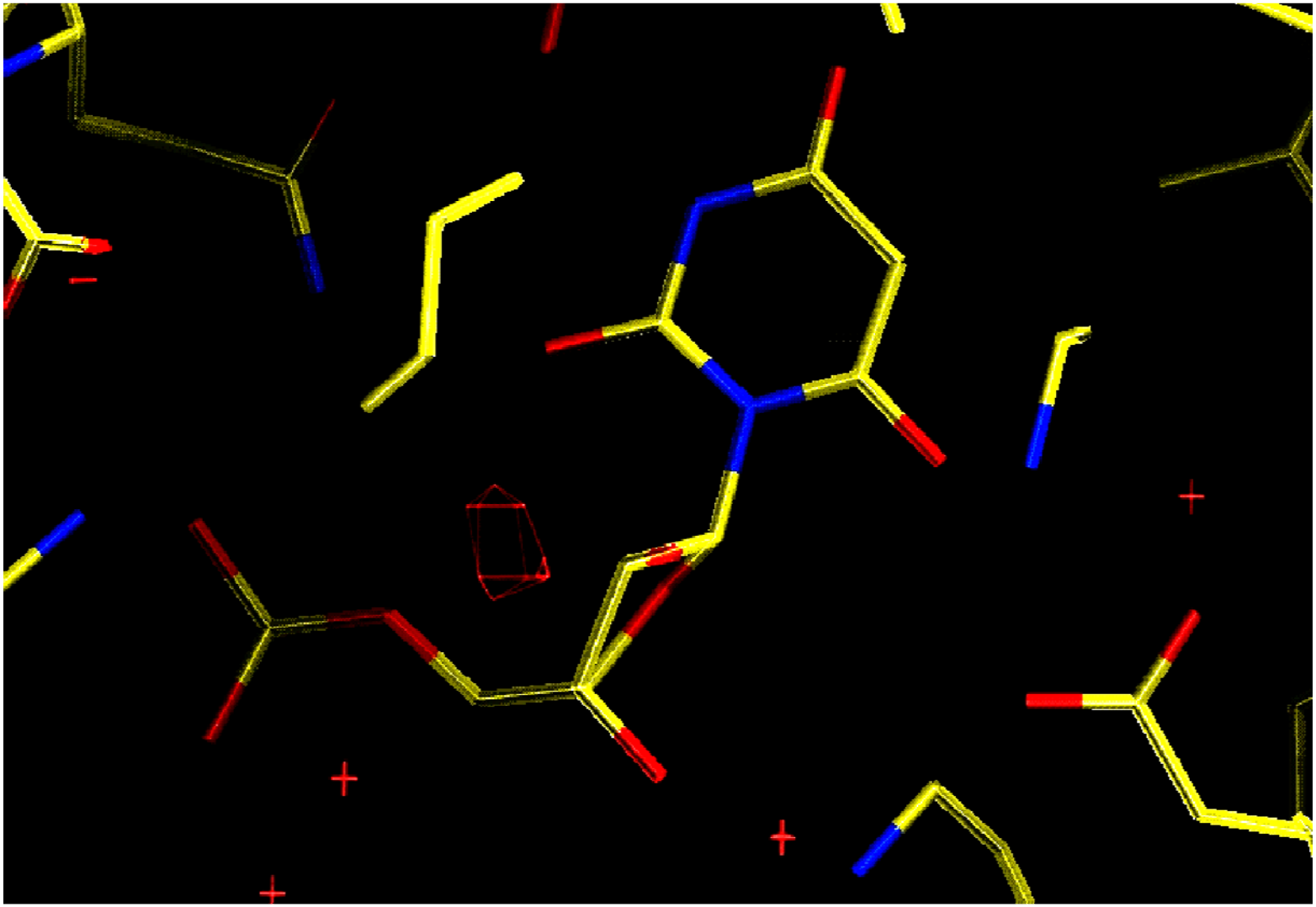
Difference density from native data (2.5Å)



Difference density when BMP is included



Final difference density



Better data set (atomic resolution)

- Continue with stereo-chemically restrained refinement

  (for example using SHELXL)

- Refinement of anisotropic temperature factors

- Hydrogen atoms at ideal positions

  (this is very time consuming)

Cytochrome c_4

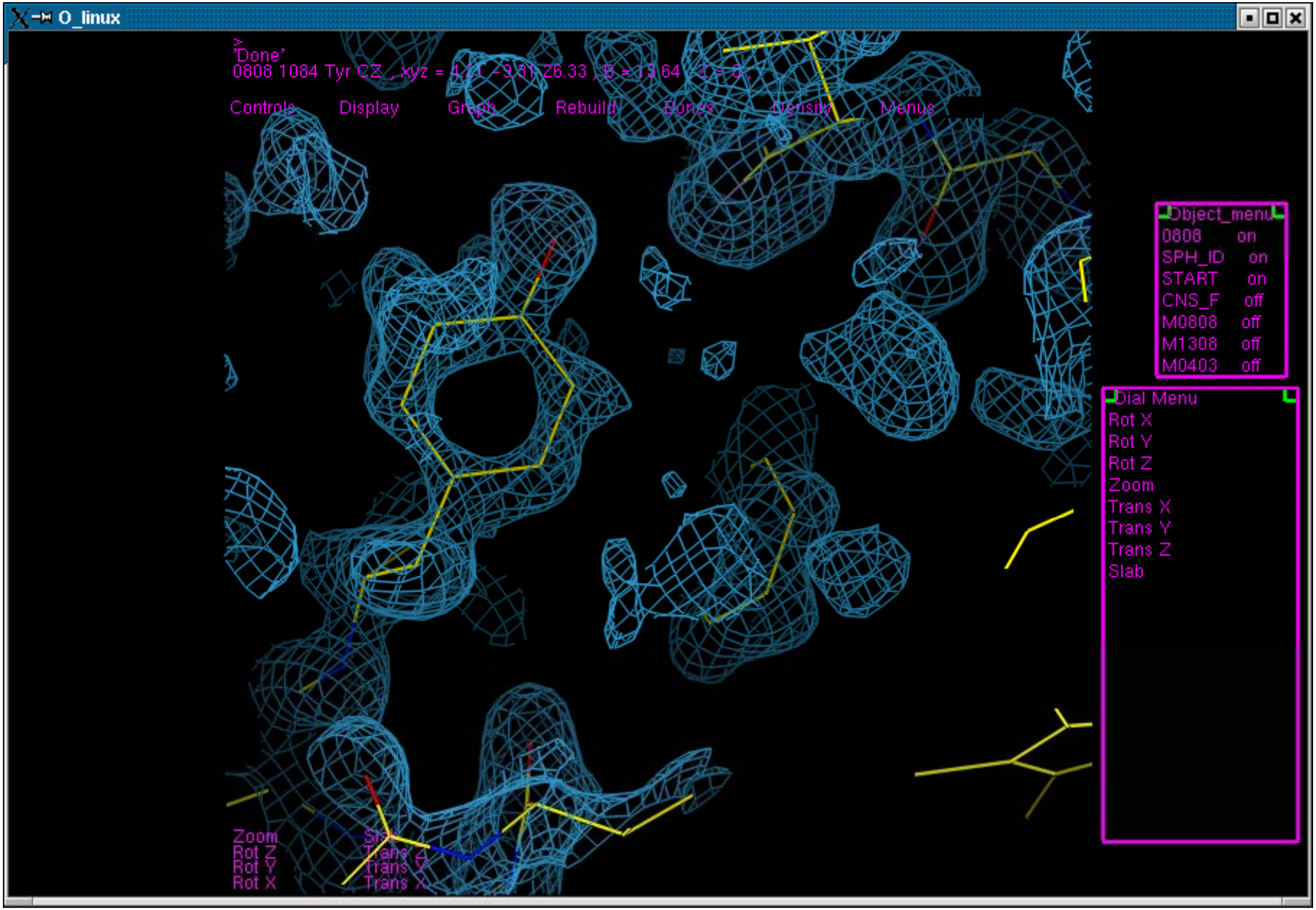
Electron transfer heme containing respiratory protein

- we were interested in the reduced and oxidised crystal forms
- crystal structure to 2.2 Å was already known
- Oxidized form 1.25 Å data (I-7-11 MAXLab)

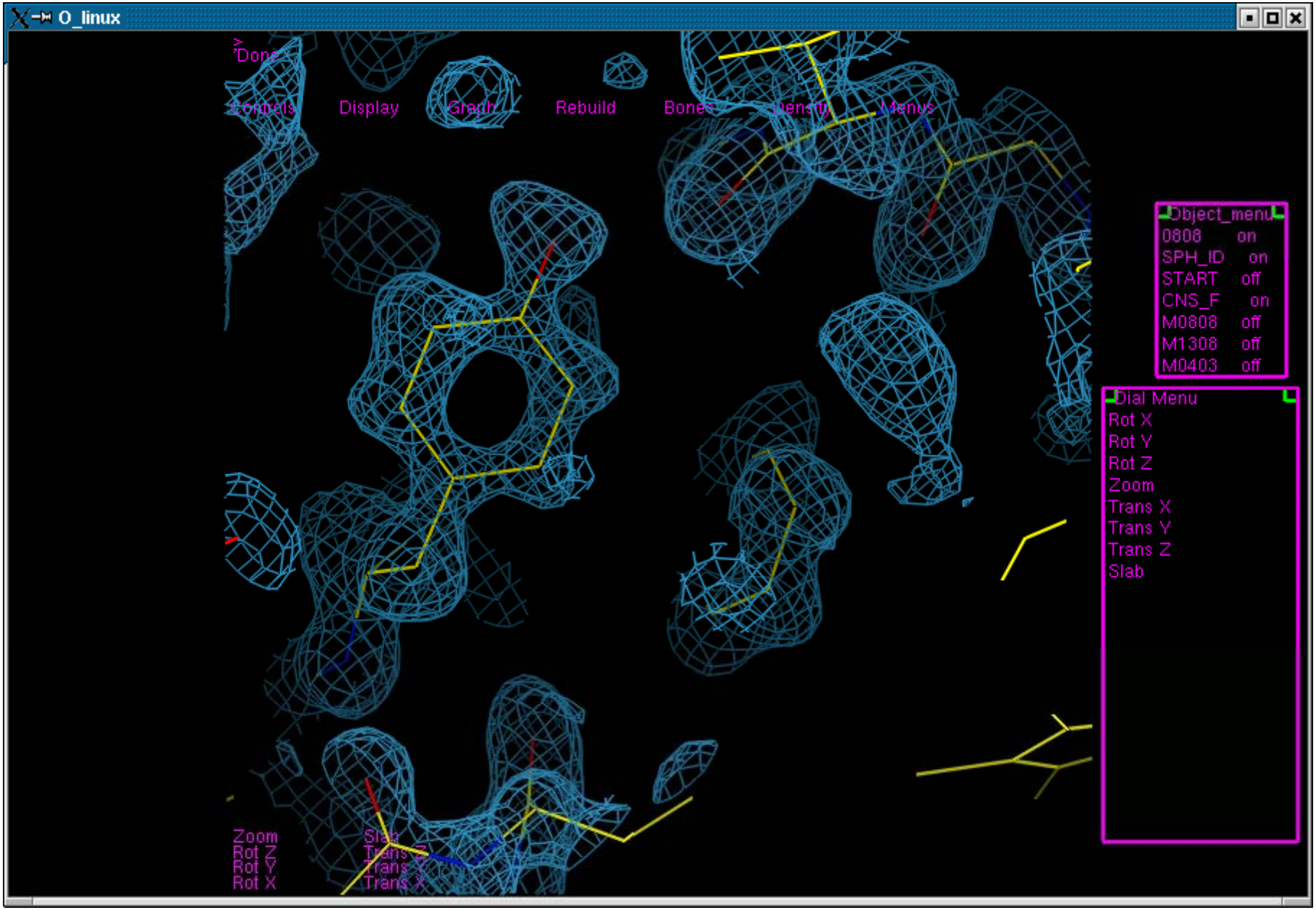
Refinement procedure

- First solve the structure from the previous structure using MR
- Refine isotropically
- Find water molecules R=0.22 R_{free}=0.24
- Refine anisotropically – the displacement is described by an ellipsoid instead of a sphere
- Correct residues in multiple conformations and add hydrogen atoms R=0.15 R_{free}=0.20

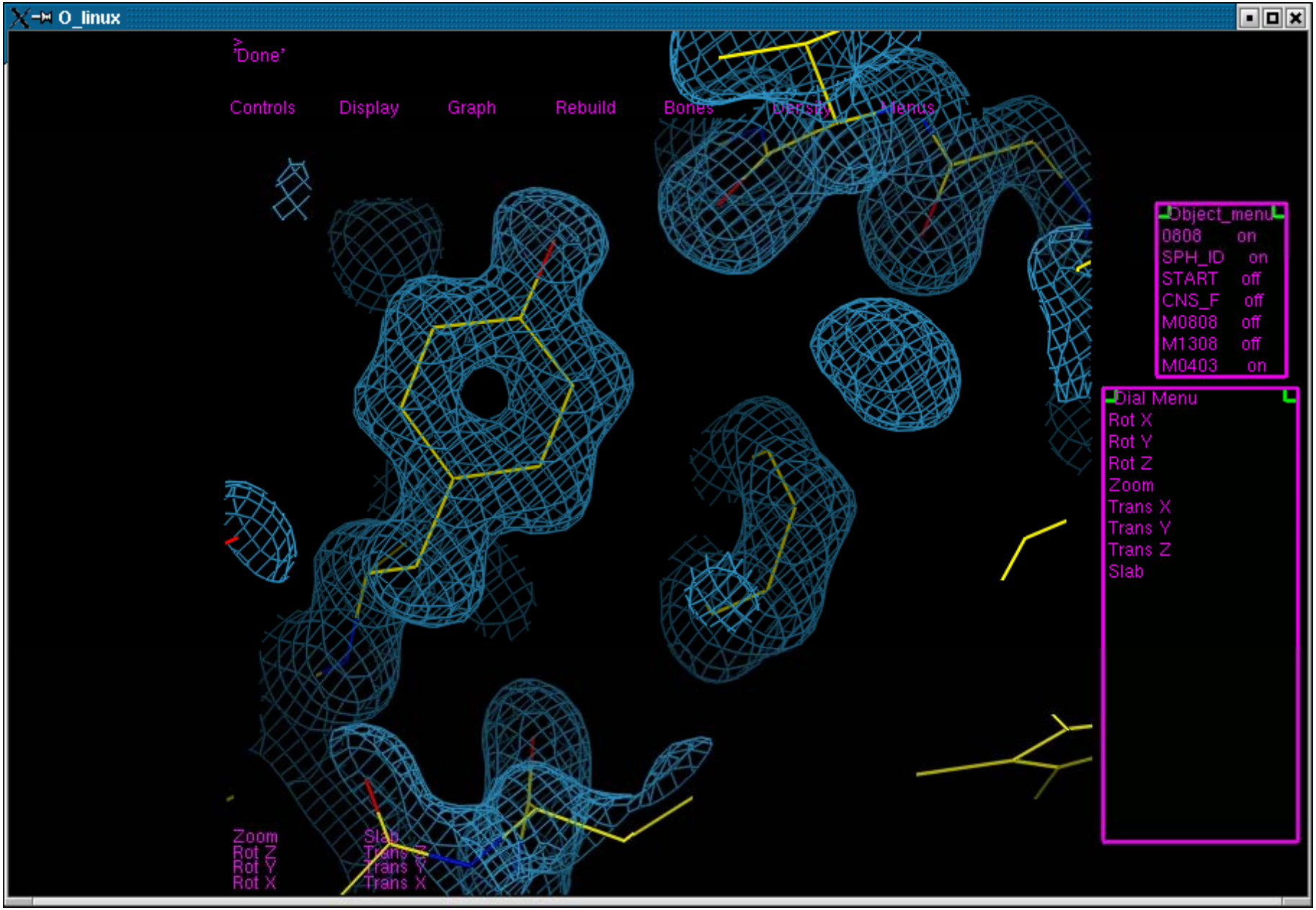
Map after Molecular Replacement



Map after Refinement in CNS



Map after final refinement in SHELX



validation

a structure is a 3D representation of a molecule, containing information of the mutual atomic positions

- low resolution: few details – maybe only the envelope is known
- medium resolution: over all fold is known
- high resolution: most atoms are resolved. Coordinates may be determined within a carbon-carbon bond

be aware of the resolution before you start to draw detailed conclusions

resolution

- a low number → high resolution (e.*g.* 1 Å)
- a high number → low resolution (*e.g.* 4 Å)

A good model...
makes sense in all ways you can think of.
This includes:

- **chemical sense**: normal bond lengths and -angles, correct chirality (no D-amino acids, flat aromatic rings, flat sp2- carbon atoms *etc*.)
- **physical sense**: no clashing, sensible crystal packing, neighbouring atoms have similar thermal parameters, occupancy of alternative conformations adds up to one, *etc*.
- **crystallographical sense**: the model fits the experimental data
- **statistical sense**: the model is the best way to explain the data (no overfitting)
- **protein chemical sense**: the model looks like a protein: Nice Ramachandran plot, not to many unusual conformations, no buried charges, (the amino acid residues ”like” their surroundings.)
- **biological sense**: the model can explain other measurements, like activity, specificity, inhibitors...

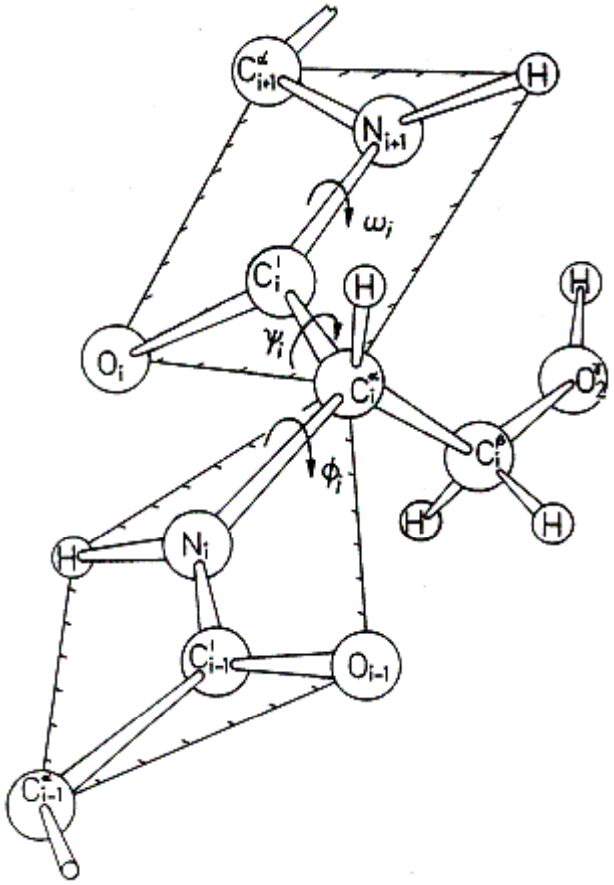
coordinates and temperature factors

- Compare the bond lengths and angles with ”known” values
  - and conclude that a good agreement means a good structure
- Look at the B-factors
  - And conclude that high B-factors means a bad model
- *Validation value: bad*

torsion angle ω

- the conformation of backbone is described by the torsion angles
- ***φ:*** C_{i-1}-N_i-$C\alpha_i$-C_i
- ***ψ:*** N_i-$C\alpha_i$-C_i-N_{i+1}
- ***ω:*** $C\alpha_i$-C_i-N_{i+1}-$C\alpha_{i+1}$

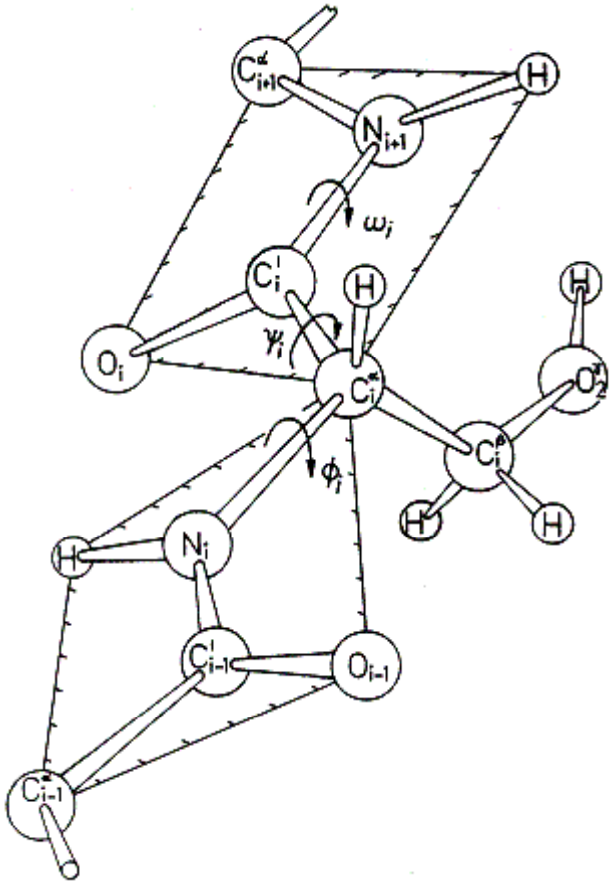
- the peptide bond has partly double-bond character and therefore it is close to either 0° or 180°
- ***Validation value ω : bad***



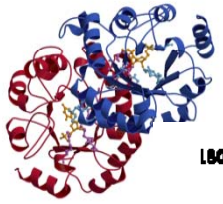
torsion angles φ, ψ

- φ, ψ angles may in principle vary freely. There is steric hindrance, and they will be restricted
- ***φ:*** C_{i-1}-N_i-$C\alpha_i$-C_i
- ***ψ:*** N_i-$C\alpha_i$-C_i-N_{i+1}

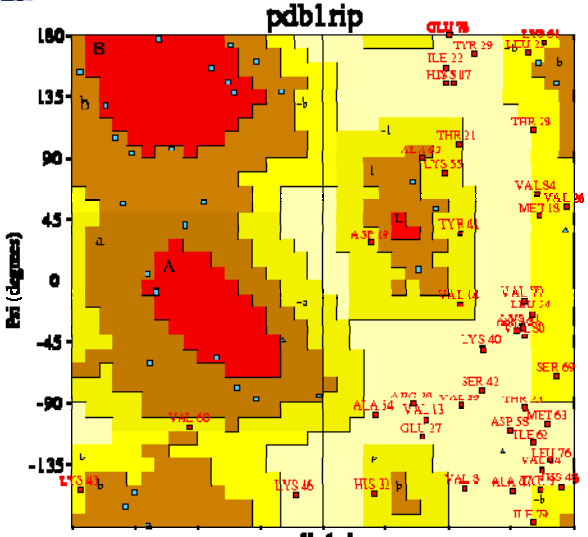
- a φ, ψ plot is a **Ramachandran plot**. If a φ, ψ combination is outside the allowed areas there should be a good reason
- ***Validation value φ,ψ: excellent***



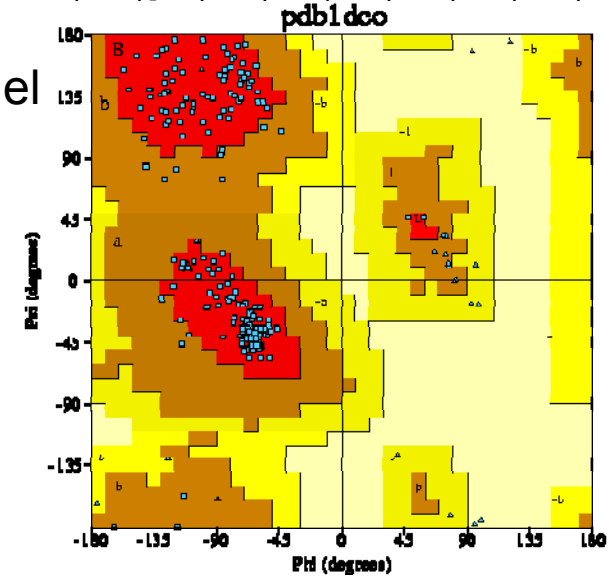
torsion angles φ, ψ

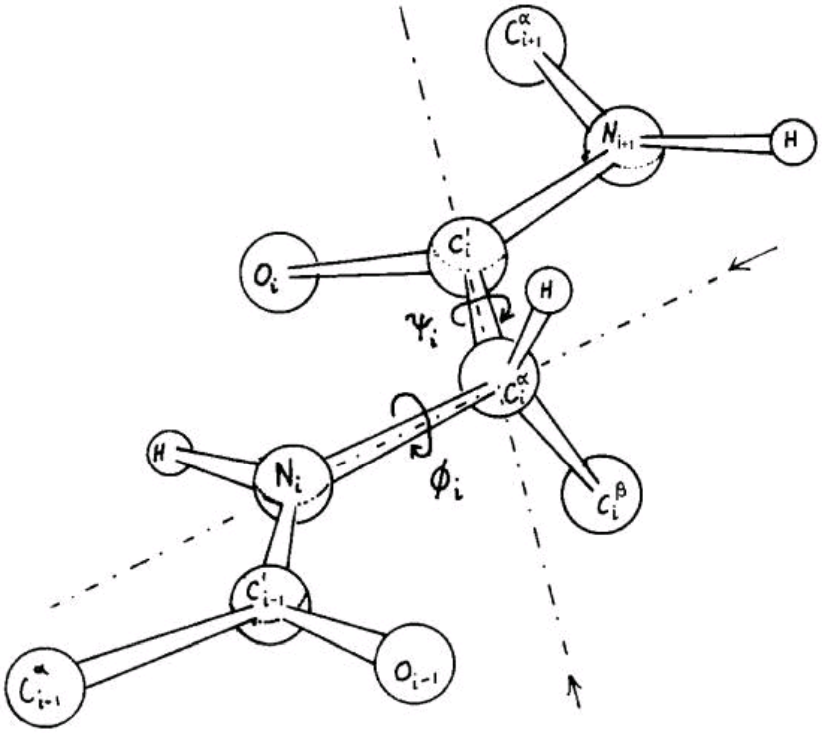


bad model



good model

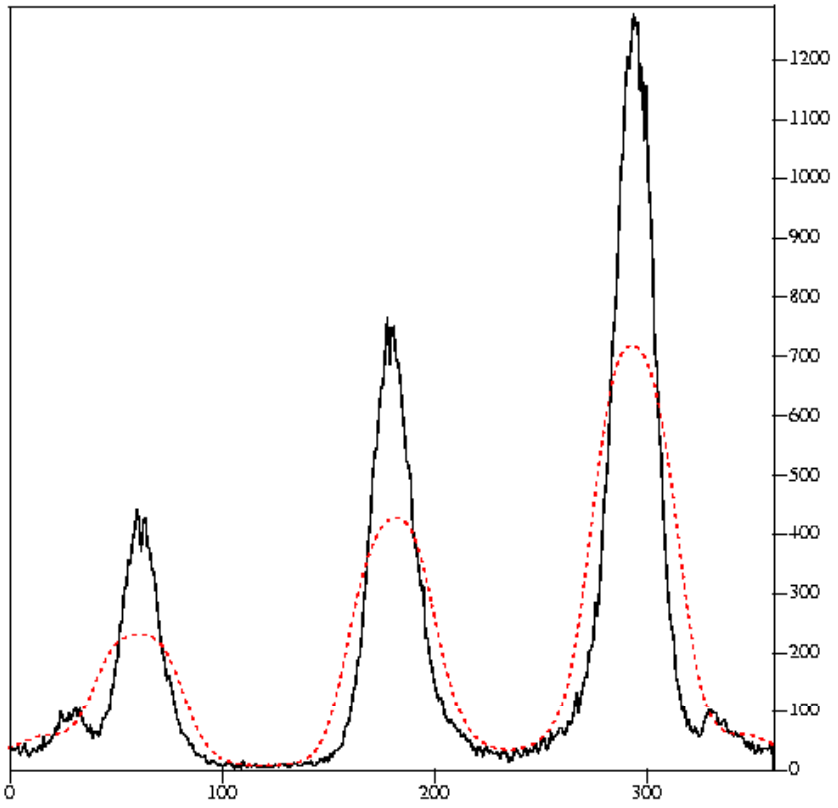




NB: both models are downloaded from the Protein data bank!!

torsion angle χ

- If the side chain is longer than Cβ there is at least one torsion angle,
- χ-1: N-Cα-Cβ-Xγ;
- χ-2 Cα-Cβ -Xγ-Xδ
- ***Validation value χ: moderate***

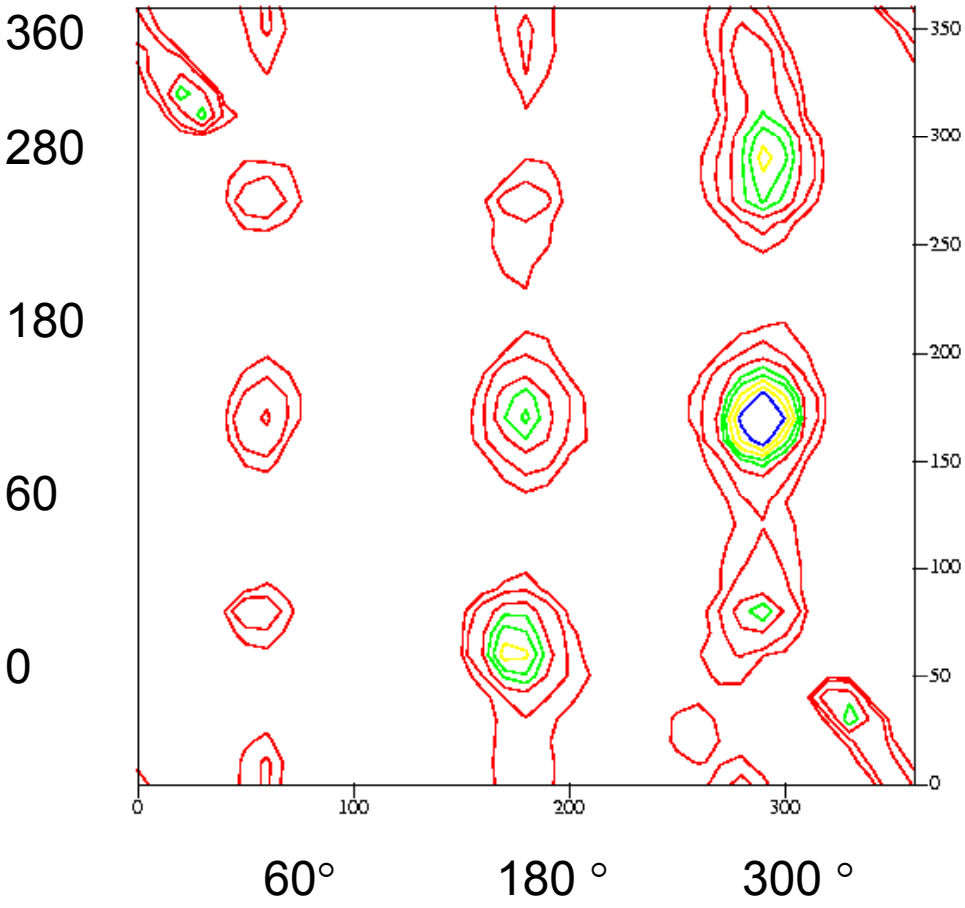


torsion angles χ-1, χ-2

- Some combinations χ-1, χ-2 are preferred rotamers.

  As the Ramachandran plots, a χ-1, χ-2 plot can be very valuable

- *Validation value χ combinations: excellent.*



packing

Protein structurs are stabilised by hydrophobic and hydrophilic interactions. Hydrophobic residues usually stack, charged residues make socalled salt bridges and hydrophilic residues make hydrogen bonds or stick into the solvent. Mistakes in the model could show by unfavorable interactions.

Directional atomic contact analysis (DACA) is used to give a score for each residue (does it approve of its surroundings). An area with a low score is probably wrong.

- ***Validation value DACA analyse: excellent.***

model vs data

a model is somebodys personal interpretation of experimental data.

The resolution tells how good the data are. High resolution means more data and a more detailed model. The resolution is chosen by the crystallographer – it cannot be compared from data set to data set. Generally, a 1.5 Å model should be better than a 3 Å model.

Validation value resolution: moderate

R-value

- The traditionel way is to use the conventional R-value:

$$R = \frac{\sum_{hkl} \| F_{obs} | - k | F_{calc} \|}{\sum_{hkl} | F_{obs} |}$$

- The R-value may be reduced by increasing the number of parameters. It only makes sense when data/parameter is large.
- ***Validation value R-factor: bad***

R_{free} -value

- R_{free} is an independent *R-value calculated on data that are not used to refine the model*

$$R_{free} = \frac{\sum_{hkl,testset} ||F_{obs}| - k|F_{calc}||}{\sum_{hk,testset} |F_{obs}|}$$

- R_{free} is always larger than the *R-value.*
- *Validation value R_{free}: good*

- the local correlation between the calculated and the measured electron density map. A kind of real-space R-*value* RSR *(like the one you get in coot)*
- ***Validation value*** RSR***: good***

poor indicators

- conventional *R-value*
- bond lengths and – angles RMS deviations from ideal values
- Average temperature parameters

good indicators

Global values

- R_{free}
- Packing-score
- Ramachandran plot

- ***local values***
- Real-space fit
- main chain torsion angles
- Side chain torsion angles