IIT Guwahati

Lecture 31

Course BT 631

Protein Structure, Function and Crystallography

Prof. Arun Goyal

Dept. of Biosciences and Bioengineering

Data Processing, Phasing, Model building and Refining

Data processing and reduction

- **Indexing** : Using the position of spots on one or more images to determine the Lattice parameter, Finding the unit cell, Orientation and Space group.
- **Integration** : Determining the intensity of each spot.
- **Merging** : Multiple observations are scaled and averaged for determining the data quality or resolution.
- **Calculating** : the structure factor amplitudes (F_{hkl}) from merged intensities.

Methods for Resolving the Phasing Problems

1.Molecular Replacement

- Requires some prior knowledge of the crystal structure to solve (homologous protein, etc.)

2.Experimental Phasing

- Isomorphous Replacement
- Anomalous Dispersion Methods

Methods for Resolving the Phasing Problem

Is there a closely related structure?

Yes

1) Use Molecular Replacement

No

Are there enough methionine?

Yes

Use suitable expression with Seleno-methionine as methionine source?

Use Anomalous Dispersion

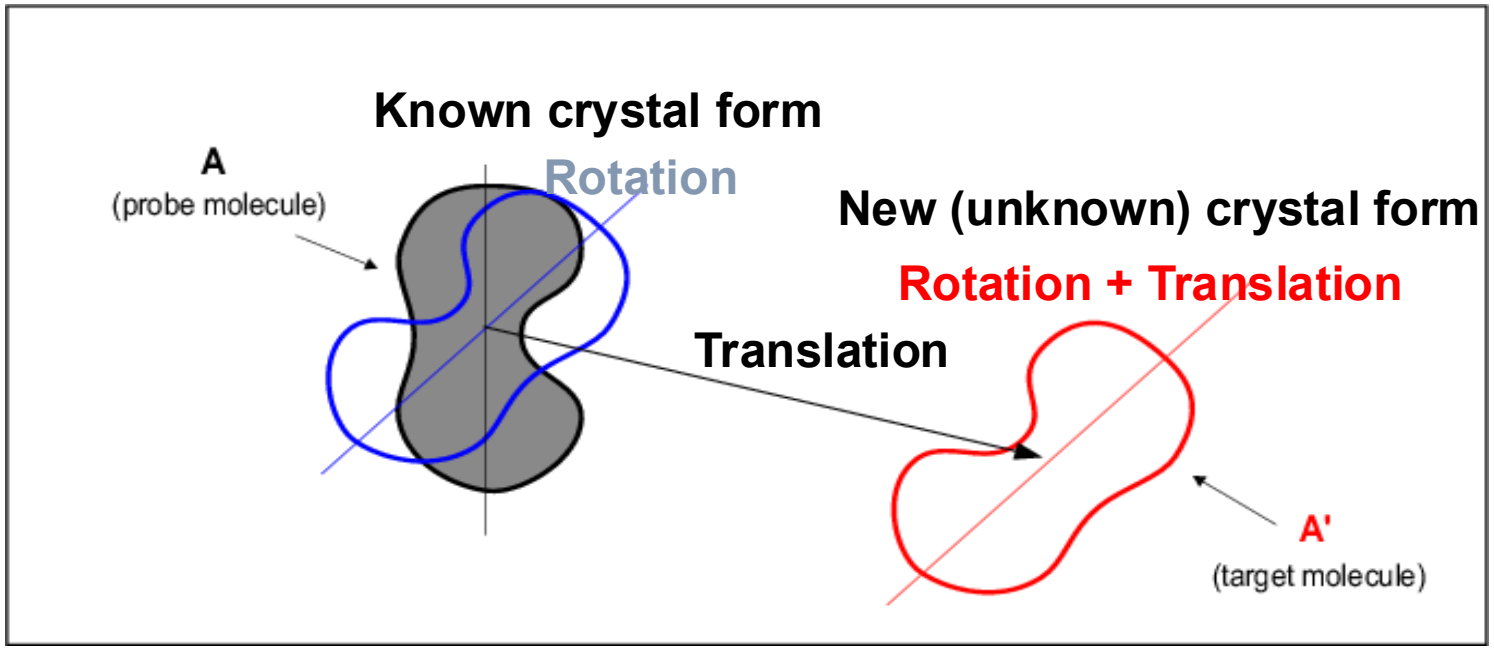
No

Can soak heavy metal atoms (Hg, Pt) into protein crystals?

Use Isomorphous Replacement

What is Molecular Replacement?

- **Molecular replacement (MR)** is generally used to describe the use of a known molecular model to solve the unknown crystal structure of a related molecule.
- **It is most commonly used method and 80% of the total deposited macromolecular structures are solved by Molecular replacement.**
- **Molecular replacement works on Rotation and Translation function.**

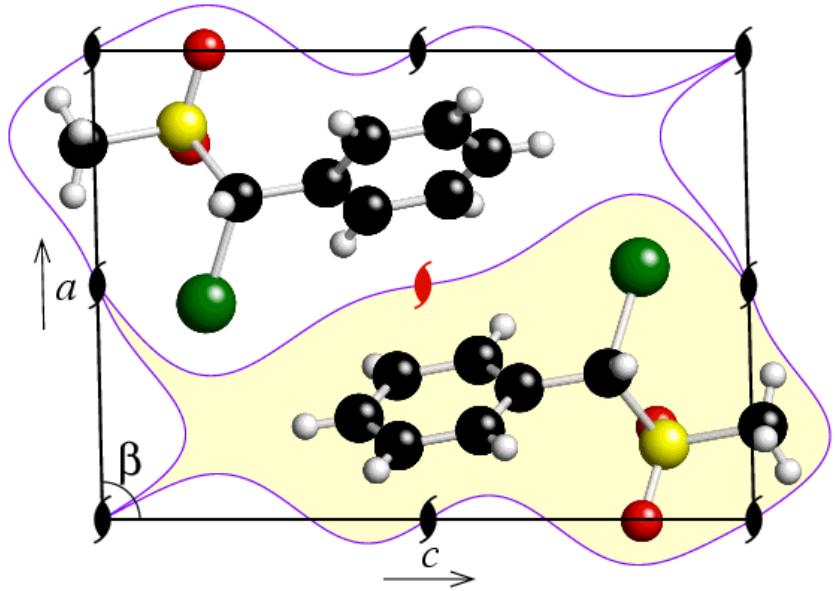


What is Molecular Replacement?

- **MR works best when there is a high similarity (>30% sequence identity) between the target (new) and the probe (known).**
- **MR works best when one is searching for only one molecule of the Asymmetric Unit of the target crystal.**
  (The asymmetric unit is the **smallest fraction of the unit cell that can be rotated and translated** using only the symmetry operators allowed by the crystallographic symmetry to generate one unit cell)

Crystallographers frequently choose asymmetric units that are determined by the "outline of the molecule". This is shown schematically in the figure for S(-)α-Bromobenzylmethylsulfone,

There are two molecules in the unit cell, which are symmetry related by the two-one screw axis shown in Red. More precisely, it is the centre of gravity of the two molecules that lies within the unit cell, since a few atoms lie outside it



α-Bromobenzylmethylsulfone

- **The data from the target crystal are very complete (ideally 100%).**
- **Systematically missing regions (ice rings, detector overloads, blind regions) can cause problems.**

What is Molecular Replacement?

How to Select the Model for Molecular Replacement

- **The higher the sequence identity, easier is the Molecular replacement.**
- **If there are particular regions of low sequence identity, trim side-chains to Ala/Gly.**
- **If a domain in the probe does not exist in the target, delete this from the search model.**
- **If a domain in the probe is flexible, delete this from the search model or (better) use the two domains as separate search models in the same MR protocol.**
- **If several models are available, combine them for use in MR using 'ensembling' (Phaser).**

  (The search models used in **Phaser** are customarily named "**ensembles**")
- **Software used for molecular replacement : Molrep, Phaser, Balbes and Beast.**

How does Molecular Replacement work?

How does molecular replacement work?

Anomalous Dispersion

- Anomalous Dispersion method (AD) is generally used to describe the structure of a protein which does not have homologous crystal structure.
- In this method the methionine amino acid is replaced with Seleno-methionine (a derivative of methionine amino acid) by supplementing it in the growth medium for protein expression and production.
- This method can be used when the protein contains at least 1 methionine/100 amino acid residues.
- For structure determination the spot intensity of seleno-methionine in diffraction pattern of a single crystal will be identified and can be used as starting point for solving the phases.
- Software used for Anomalous Dispersion : SHELXC/D/E.

Isomorphous Replacement

- **Isomorphous Replacement (IR) method is generally used to describe the structure of a protein which does not have homologous crystal structure.**
- **This method is used for the structure determination of protein which does not contain methionine.**
- **This method uses the crystal soaking in heavy metal solutions such as Hg and Pt.**
- **For structure determination the diffraction pattern of two crystals i) native and ii) heavy metal soaked crystal shall be collected.**
- **Substructure present in the heavy metal soaked crystal will be determined and used as starting point for structure determination.**
- **Software used for Isomorphous Replacement: SHELX C/D/E.**

Model Building and Refinement

Steps in making the first trace in Electron Density map

1. Generating Cα chain trace.

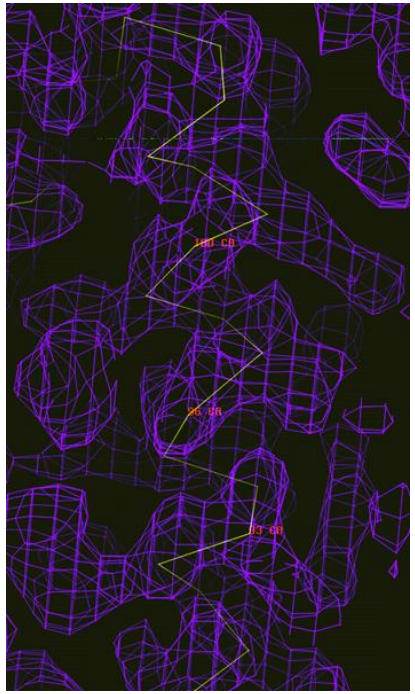
- **The only rule one has to observe is that the distance between Cα atoms of adjacent residues is always approximately, 3.8 Å.**

- **Try to look for large pieces of secondary structure, such as helices and sheets, to start the Cα trace.**

Model Building and Refinement

2. Identifying chain direction

- The side chains on a helix point towards the N-terminal end.
- Another way to put it: the α-helix resembles a Christmas tree, when viewed with the N-terminal end down and the C-terminal end up.

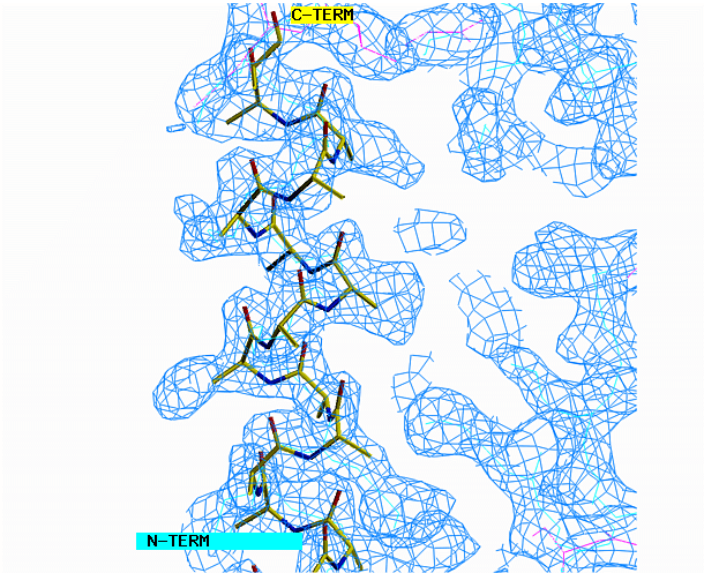


Model Building and Refinement

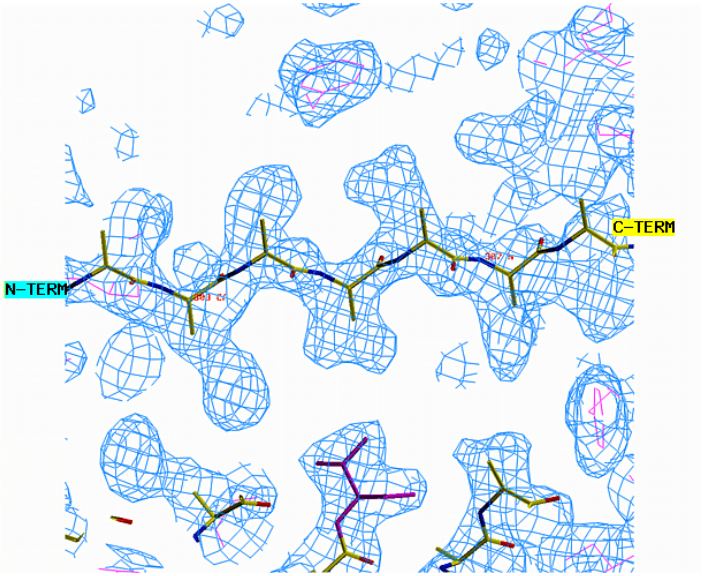
3. Generating main chain trace

- Main chain can be automatically generated from a well traced Cα chain.
- In helices, the side chain positions are so highly constrained that one can accurately predict the main chain and C_β atom positions with a refined α-helix from another protein.

Examples of generated α-helix and β-sheet in electron density map



α-helix



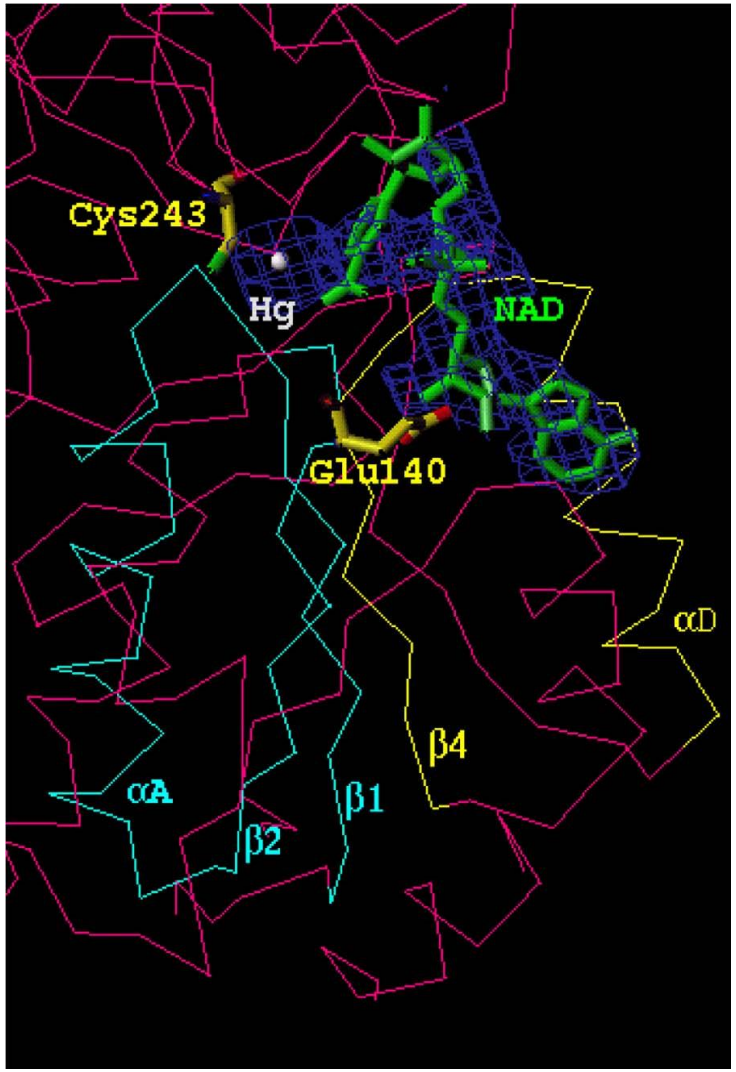
β-sheet

Model Building and Refinement

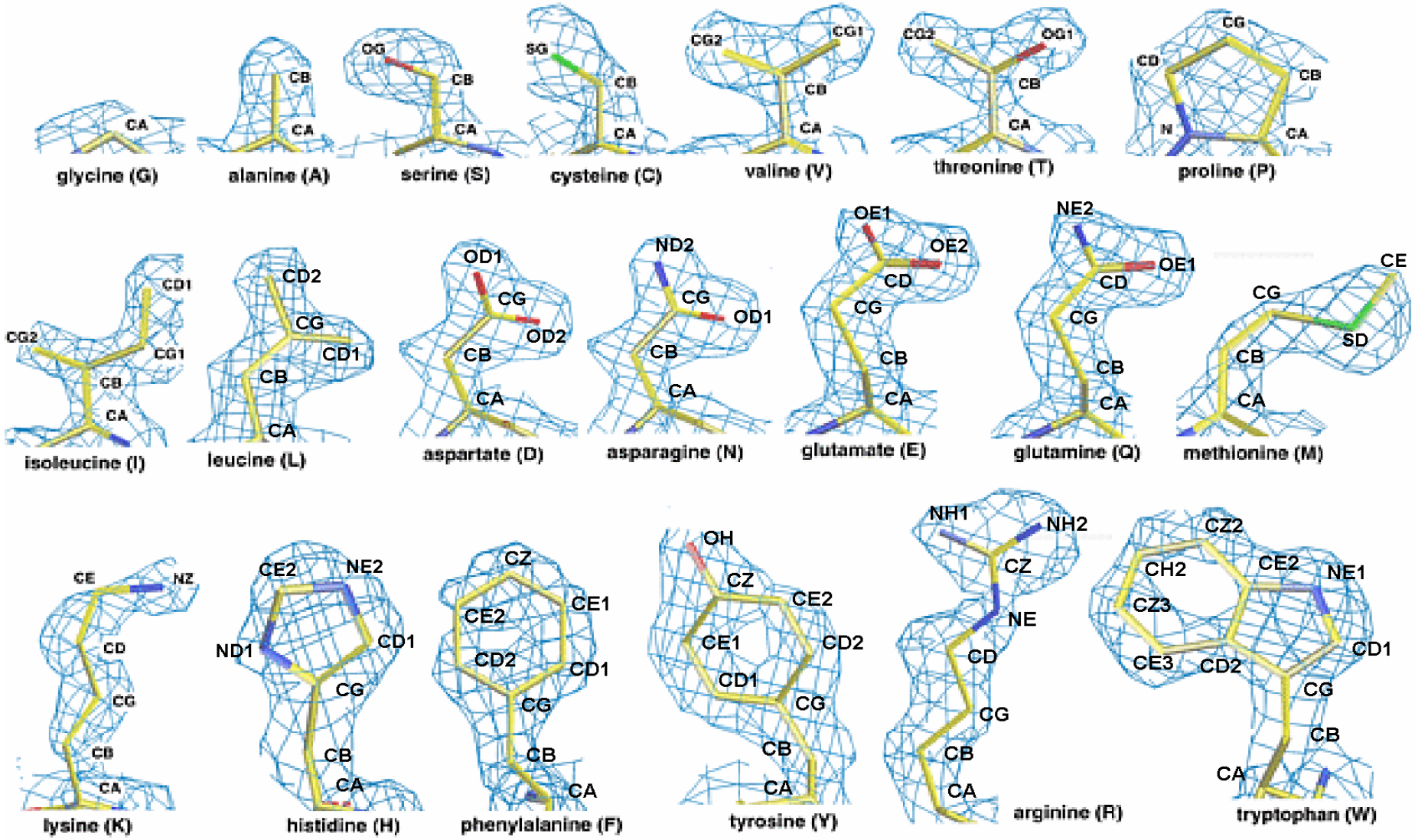
4. Fitting the chemical sequence.

Finding the first match of sequence to the map is a milestone in structure determination. Some tips are listed below:

- Heavy atoms bind to some specific residues. Hg-Cys, Pt-Met
- Start the fitting from a well defined main chain trace, where the density is clear and rich in side-chain information. These regions are often located inside the molecule.



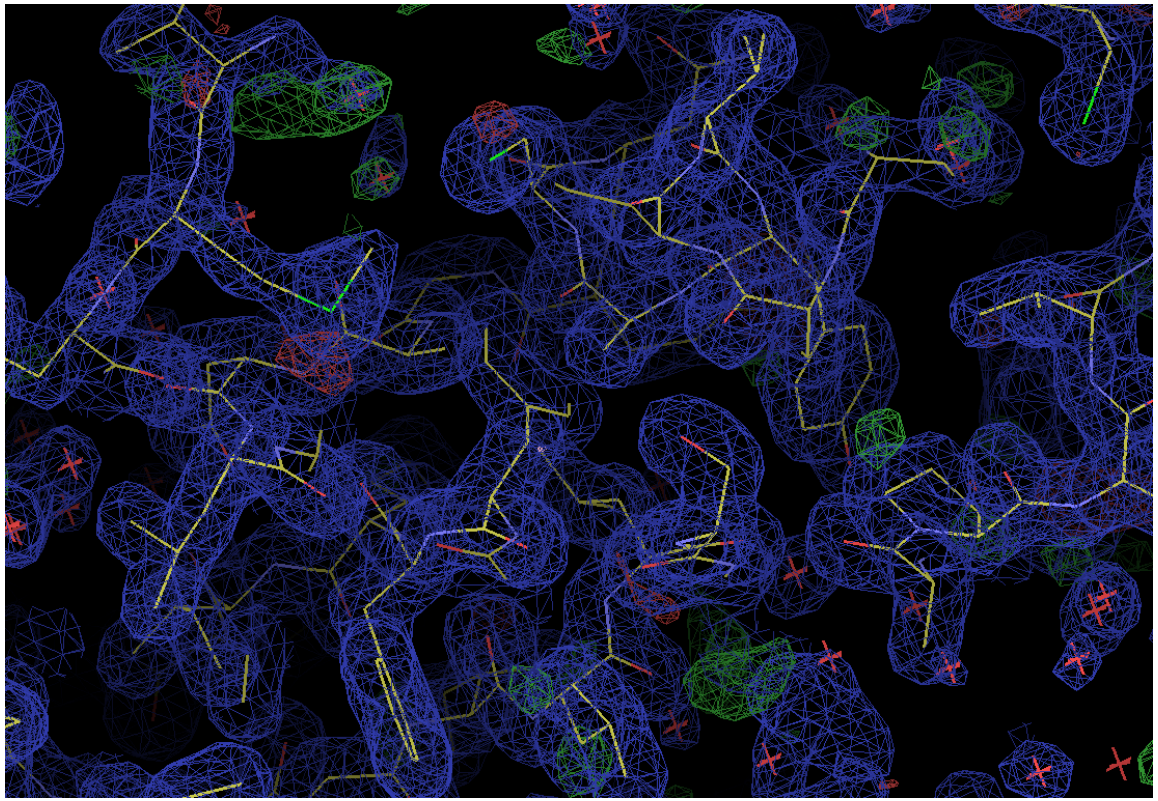
Representative electron density for amino acid side-chains arranged in order of increasing size



Experimental electron density map calculated at 1.5 Angstrom resolution.

Generating and refining the model

- Generate the side chains based on the fitted sequence.
- Sometime the generated side-chain may not point at the correct direction.
- In most cases, the manual adjustments are needed.



Generating and refining the model

- The side chains should fit the electron density map all over the whole molecule. If the fitting suddenly becomes bad in some region, it may indicate that something is wrong with the fitting.
- Missing density is much better than extra density. It is rarely seen that there is a blob (Patches) of extra density for Gly, Ala or Pro residue.
- The model should make chemical sense and satisfy all that is known about the macromolecule.
- It may be useful to evaluate the overall distribution of some residues, such as hydrophobic residues, glycine and proline.
- If certain residues have been identified as being in the active site, are they close together in the model?

Structure Validation and Deposition

- **The stereochemical parameters such as bond length, bond angle etc. should be within the standard deviation from their ideal values.**

- **The target protein should follow the Ramachandran plot rule which can be checked at Molprobity.**
  (**MolProbity** is a structure-validation web service that provides broad-spectrum evaluation of model quality)

- **After the evaluation the determined protein structure can be submitted in the protein data bank, a universal repository.**