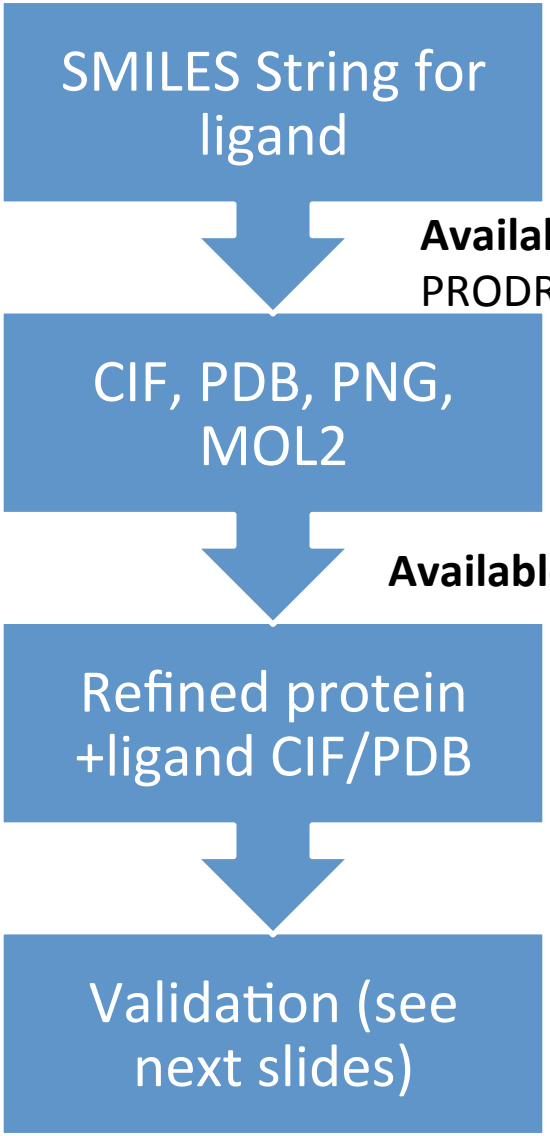
Recommendations and Questions

wwPDB/CCDC/D3R Ligand Validation Workshop
Center for Integrative Proteomics Research, Rutgers
7/30-31/2015

Group D, Academic and Industrial Crystallographers:
Kathleen Aertgeerts (co-chair)
David Brown (co-chair)
Seth Harris
Tobias Krojer
Alan Mark
Guy Montelione
Robert Nolte
John Spurlino
Chenghua Shao
Oliver Smart
Paul Emsley (PM session)

Recommendations

1. Recommended X-Ray Structure Refinement Workflow



Potential ambiguity in chirality, tautomeric state and charge

Available software: CORINA, ELBOW, GRADE, Afitt, ACEDRG, PRODRUG, ATB

Available software: Coot, Phenix, Buster, Refmac

Wiki-style educational recommendation on good practices on ligand chemistry and structural solution, with community cooperation and contribution.

1b. Dictionary/model

- **Dictionary:** Key restraints used in refinement that can be from multiple sources. To incorporate rotation freedom of certain bonds, and certain degree of freedom for conformation flexibility.
- **Model:** Set of 3D coordinates of ligand to start modeling and refinement process. To find low-energy conformation(s) . To combine tools and manual process.

2. Validation of ligand during model building and refinement cycle

- Comparison of B values on protein vs ligand
- Consideration of occupancy in refinement on ligand; consideration of multiple conformations on ligand; Consideration of disordered moiety of ligand
- Restraints in mmcif vs observed geometry in refinement process
- Database methods (e.g. Mogul) or automatic computational scientific software that assesses ligand geometry during refinement (to be developed)
- Issues(breakout session): covalent ligands, unnatural amino acids
- RSCC/RSR/LLDF, and difference density explanation. Alternate modeling, e.g. test hypothesis of the extra density being water
- Include hydrogen atoms to ligand and its binding site residues that facilitates interpretation of protein-ligand interaction

3. Validation during PDB deposition

- Full ligand should be enumerated, and author defines ligand of interest (e.g. LIG vs ATOM/ HETATM) in the PDB/CIF model file
- Restraints dictionary in mmCIF file mmCIF
  - Ligand definition (Recommend to include into mmCIF energy term interpretation, and refinement program to output required files for deposition)
- Slider picture of ligand quality assessment (general and conditional on resolution)
- RSR, RSCC values at atom and ligand level
- Develop simple and clear metrics on ligand quality at atomic level
- Difference electron density figure with fitted ligand
- Additional column of uncertainty measure(TBD, quantitative) per atom in mmcif that can be captured in visualization programs, e.g. well-defined/ill-defined in NMR VTF; no density with color code;
- Automated computational scientific tools available on web; software to predict reasonable geometry. And distributable package for local clients
- Batch deposition process
- Make CAVEAT more obvious and request for authors to fix/explain issues
- Protein-ligand interaction: clash score, interaction fingerprint and energy. To compare a new structure's ligand to the existing validated structures; fragment fitting comparison.

3b. Additional optional information provided by authors during PDB deposition

- Available QC data on ligand (e.g. NMR, MS)
- Binding data. In batch mode deposition, to have access to the experimental binding data for the set.
- Author's processing details/comments in fields specific to individual ligand and its refinement process
- Other info (e.g. source)

4. Ligand Validation during journal submission

- wwPDB validation report including enhanced ligand validation (Buster report as example). Highlight CAVEAT and author's response.
- Initial omit density before ligand is loaded (with the final ligand model overlay); difference electron density figure with fitted ligand.
- Recommend disclosure of fitted ligand and binding pocket. Provide web-access to the coordinates, SF, and map coefficients for reviewers
- Re-refinement on any existing structure should refer to the original structure/publication, as well as new deposition made

5a. Recommendation on existing PDB archived co-crystal structures: what users want

- Flag of bad structures, or bad ligands using validation tools. Display slider bar for ligand(s).
- Alert authors of the entries identified above
- Possible automatic re-calculation on alternate modeling for the co-crystallized ligands identified above, which could motivate CASP-like computation competition and development of new methods.

5. Recommendation on existing PDB archived co-crystal structures

- Update on the model by the same author (or PDB) keeps the same PDB code with versioning, no requirement for obsolete, and requires mandatory description for the reason of update.
- Re-refinement of any structure done by different/same authors, using same data, new PDB code should refer to the original PDB code and data (current practice at wwPDB)
- Capture curated comments from authors/users on the PDB web

6. Recommendation for ligand chemistry description

- Agree to all the recommendations in the doc
- Indicator of the exact charge or tautomer state in the model (author provided, or unknown)
- Standardize atom naming convention, e.g. InChi canonicalization

Questions/
Points of discussion

Questions:

- Refinement vs Validation: Validation can be performed during refinement, after refinement, during validation, during PDB process. What is the best practice?
- Buster's ligand review example (see bottom) and its implication on ligand validation process.
- What is ligand? (e.g. Glycerol, Sodium ion can be relevant ligands but mostly are solvent/buffer). To let author specifically mark what is significant for structure for referee review?

http://www.globalphasing.com/buster/wiki/index.cgi?BusterReport

Questions (cont)

- Occupancy review, e.g. how to deal with zero occupancy?
- B factor review, e.g. how to deal with B factors that are very high?
- Validation components needs to be distributed to the community?
- Accessibility of critical software to diverse academic research groups, so that all users are able to generate files for ligand modeling, validation, and deposition.
- Inconsistent outcome between components, e.g. Mogul vs OpenEye. Leading to direction of cross validation?
- Density fitting restraints at lower resolution may have problems and ambiguity.
- Special cases that are valid can be outlying against reference. How to highlight and deal with it?

Questions (cont)

- Ligand completeness issue. To set artificial occupancy (e.g. zero) can complicate B factor.
- The current problems with refinement programs: covalent, metal, etc.
- Automatic tools at web to assess/predict ligand validity.
- Batch deposition output from in-house sources/databases should be handled, and how?

Questions (cont)

- Explanation for unfitted density? Especially the presence of difference density close to the ligand atoms.
- To include validation components in refinement?