Ligand Validation Workshop
PDB, Rutgers July 30 – 31, 2015
Notes VAF

Blaney suggests that a series of short-term and long-term recommendations be generated.

The group recognizes the need for recommendations around

1. Registration – eg. What is to be included in deposition, what standards are required for deposition
2. Curation/Information associated with a structure – eg. Information that needs to be communicated to the user and opportunities for added value

Discussions/recommendations around item #1. Registration.

- The PDB should adopt a set of "business rules" for small molecule registration and require depositors to adhere to these rules for the deposition to be completed. Any issues that arise during the generation of the Ligand Validation report that violate these rules should be addressed by the depositor, not the PDB curator.
- HELM can provide useful guidelines for business rules. JCIM 2012 52, 2796.
- There should be full valence registration so that substructure and similarity searches can be done later.
- If necessary, deposition of ligand structure known to be added to the crystallization process and the ligand structure observed in the crystal structure; this is especially important for covalent inhibitors; include the original ligand structure in the mmCIF. Authors to indicate/notate on the structure (as a 2D depiction?) the site of covalent modification. Authors required to note the expected tautomers and stereochemistry for the ligand.

Discussions/recommendations around item #2. Respository Information & Added Value.

- The "ligand of interest" should somehow be annotated to distinguish it from crystallization solution additives, co-factors, etc.
- There will be a need to move beyond the current three letter ligand code because they will run out of code options soon.
- Recommend ligands are depicted with hydrogen atoms; hydrogen atom occupancy = 0. Include all "not observed" heavy atoms with an occupancy of zero. This would help with ambiguity of bond order and tautomer issues.
- Provide versioning on dictionaries; components.cif and components_variants.cif
- Output formats need to be better formatted to facilitate their ease of upload to various software programs (e.g. the work of Alex Holden). Adopt a standard for ligand output such as adopting one canonical smiles methodology; ChemAxon or Daylight canonical smiles is recommended as a possibility; try to have agreement between CSD – PUBCHEM – PDB – ChEMBL

in this regard. Discussion proceeded around whether smiles or INCHI was better – INCHI is challenging as one loses tautomer information.

- Revise Validation Report to 1. Move key ligand validation information to the first summary page of the report, just under the Protein summary and make the quality of the ligand more prominent in the report. 2. Provide ligand specific information, perhaps including a "visual" for reviewers, ie. an omit map for the ligand or the ligand model in the electron density. There was discussion of whether an omit map would be useful for the Validation report. Also, discussion on providing RSR values on a per ligand atom basis. 3. Make the location of information on how to interpret the ligand quality and validation parameters more obvious for reviewers.
- Crystal contact information could be more readily visualized.
- Provide representation on the Web page and in the Validation report of the ligand with electron density, 2Fo-Fc (note – Genentech has worked with CCG to automate this for their own internal use).
- There could be a way for visualizing the structures in an overlaid format.
- There could be substructure clustering for interesting ligands.

What to do about the crystal structures already deposited in the PDB?

- Crowd sourcing the re-refinement of structures with known issues was suggested. Perhaps a wiki page for the new structure(s) refinements with a link back to the original structure on the PDB url.
- Colin Groom noted that many companies have already re-refined (fixed) the proteins of interest for drug discovery for their internal use. Could there be a way to have those donated to the PDB for the common good? Perhaps D3R could facilitate this.
- EBI already has such a site?
- Could a re-refinement Jamboree/Hackathon be hosted? "People will do things for a T-shirt". A CASP-like challenge.

What recommendations to the journals?

- Jasmine noted the majority of crystal structures are reported in JBC, then Structure.
- I pointed out that many of the "drug like" ligand co-crystal structures are in journals such as JMC and BMCL. Can we require authors to submit a Ligand Validation Report with their submission of the manuscript to these journals?
- A discussion ensued regarding the ability of the reviewer to evaluate the quality of the structure (esp. if this is not likely to be their area of expertise).
- Colin Groom noted that the CCDC has a set of reviewers for small molecule structures; could something like this be done for ligand-protein structures?