1.b What are current best practices for selecting an initial target ligand atomic model(s) for structure refinement from X-ray diffraction data?

- Visual analysis:
  - Identification of ligand density from appropriately contoured difference density maps (with confirmation in 2Fo-Fc maps)
  - Use of interactive fitting tools that measure fit of ligand to density
- Automated analysis:
  - Use of objective fitting algorithms based on significance of electron density levels and fit of ligand to density
  - Use of methods that screen difference density against libraries of potential ligands
- General:
  - Fitting of ligands should take into account what a priori information such as buffer composition, routes of chemical synthesis and biochemistry of the macromolecule

- Challenges:
  - The actual chemistry of the ligand may not be well determined depending on the physical situation
    - Ligand in isolation may be different from in complex with macromolecule
    - Covalent modifications
    - Radiation damage
    - Ambiguity in chemistry - may be known at the time of synthesis
  - Arriving at the initial conformation for new ligands can be challenging in cases with poor density and conformation variability in ligand
  - A poor starting model will impact the final structure

- Recommendation:
  - Data items are needed to record the ligand(s) added to the crystal (vs what was actually modelled)
  - A record is needed of the method used to introduce the ligand (soak, co-crystal, endogenous)

1.b What are current best practices for generating restraints for modeling and refinement?

- Best practice:
  - Currently the use of information obtained from high resolution small molecule structures (e.g. Mogul/CSD, COD)
  - Alternative approaches make use of semi empirical or higher basis set QM calculations
- There are newer refinement programs that can use MD, QM/MM or other force-field methods instead of using researcher specified restraints (examples include DivCon/Phenix, AFITT/Buster/Phenix, AMBER/Phenix)
  - Information about these kinds of refinements needs to be presented to wwPDB end users
  - Validation programs should take account of the source of the restraints/methods used in refinement

- Challenges:
  - Existing small molecule databases can lead to bias in restraint generation
    - Small numbers of observations in some cases
    - Variable redundancy in the CSD
  - Multiple methods should be considered to generate the restraints
  - What to do when user generated restraints differ from the wwPDB internal information?

- Recommendation:
  - The method used to refine the ligand should be itemized in the deposition
  - Validation metrics for non-traditional methods (QM/MM, FF, CDL) need to investigated
  - Restraints information deposited by the user should be compared to other sources of the same information (e.g. Mogul)
  - Strain energy is not currently a good tool for validation

2. What are current best practices for validating the ligand(s) coming from such a structure refinement?

- Best Practice:
  - Current tools look at relatively crude measures such as bond RMSD(-Z) scores, and local fit to density
  - Ligands are also increasingly validated against information derived from small molecule databases (e.g. Mogul/CSD)
- There is a need for a better, validated, metric for ligand/density fit
  - Reciprocal space CC plots (c.f. BusterReports)
  - The shape of the density and the ligand could be other criteria
- Mogul/CSD analysis needs to account for redundancy or small N effects on sigmas

- Challenges:
  - The source of validation information needs to be considered
    - How to have metrics that are universal given different approaches to ligand refinement?
    - There can be limitations to the experimental databases (ligands may have different chemistry when interacting with a macromolecule)
  - Testing and evaluating alternative solutions may be needed to determine the most likely solution, taking into account:
    - What was added to the crystallization vs endogenous
    - Purity of the compound
  - Industrial access to validation tools?

- General recommendation:
  - Software tools for validation should be available for use by the community
- Recommendation:
  - Internal ligand geometry should be validated with standard approaches (bonds, angles, torsion, planarity, chirality etc)
  - Distributions of library values should be shown (visually) when possible

- Recommendation:
  - Enhance existing, and develop new, tools for assessing interaction with the macromolecule, and/or other ligands
    - MolProbity could be adapted to provide more information about ligand geometry and clashes with macromolecule
    - Simple measures such as clashes, acceptor/donor mismatches are a primary target
    - More complex metrics should be implemented later
      - Group/Chemical type interactions (going beyond atom/atom)
      - Charge/charge, VdW, etc
- Visualization tools should be modified to display this validation information

3. What new information pertaining to X-ray co-crystal structures should be required for PDB depositions going forward?

- The origin of the restraints should be provided (i.e. what methods were used to obtain restraints and/or refine the ligand geometry)
- An “Omit” map should be calculated - giving evidence of ligand
    - A best practice should be defined
- A “Best” density can be provided by the author - that shows the ligand density as interpreted
    - A best practice should be defined

- Spectroscopic data (on crystal, on sample, on ligand) could be provided
- Response to validation reports
  - Should users have to explain the outliers?
  - Can responses be formalized to aid understanding by wwPDB users?
  - Where are the boundaries between borderline and significant outliers/deviations?

- Recommendation:
  - Chemical description and restraints (mandatory for new ligands)
  - In CIF format

- Recommendation
  - A user supplied ligand map (optional)
  - An omit map:
    - Remove the ligand (rather than setting occupancy to zero)
    - Need to determine limits for excluding ligands (i.e. all simultaneously or one-by-one, depends on % of structure)
    - Need a systematic test of strategies for making the maps
    - Should calculated by the wwPDB from deposited information
- Calculate a measure of fit between model and map:
  - Real space or reciprocal space CC
    - Against user supplied map if available, and against omit map

4. What information should accompany journal submissions reporting X-ray co- crystal structure determinations? What supplementary materials should accompany publication of X-ray co-crystal structure determinations?

- Response to validation reports could be provided in reports
  - It would be very helpful if researchers had to justify the outliers in the validation
- In general information should be provided to journal method sections to enable others to reproduce experiments where possible
- A paradigm shift in the review process may be needed:
  - Data (model/maps) provided at review time

- Post deposition:
  - Annotation of changes to ligands (and the deposition in general) should be provided to wwPDB users
  - Deposition authors should be contacted with changes to ligands
    - Let users register to be notified about changes
  - The 3-character limit on ligand names is limiting, can this be increased (i.e. by moving to mmCIF)
  - Information needs to be provided to give provenance of ligand dictionaries and the specific entries
    - By naming the library/file
    - Recommend use of community tools and clear annotation (c.f. Grade)

- Recommendation
  - The validation report should be more comprehensive
    - Including real space and/or reciprocal space fit of ligand to map
  - Images of the electron density and the model (e.g. Animated GIF or orthogonal views)
    - Calculated for omit and user supplied map
- Visual display of Mogul analysis of geometry (c.f BusterReport)
- Depositor need to define the ligand(s) of interest so they can be highlighted in the validation report

5. What do you recommend be done to improve descriptions of ligand chemistry in the PDB archive?

- Ligand restraints (including any links to the macromolecule) need to be provided
  - At the wwPDB these restraints need to be versioned
- Tools/approaches need to be developed to define chemical diversity within a compound
  - A mechanism needs to be available to more completely describe protonation, and tautomeric states.
    - wwPDB has a solution for amino acids, but can this be reasonably extended to ligands?
- Overall we need a better description of ambiguity
  - Guidelines for how to deal with (best practices)
    - Radiation damage - how to best model? (hydrogen vs radical)
      - Use alternate conformations (at the whole ligand level)

- Recommendation
  - mmCIF data items need to be created to identify which atoms are modelled but not by fit to density (data)
  - Visualization tools need to be modified to display this information

6. What do you recommend be done with existing X-ray co-crystal structures in the PDB archive?

- All current and future validation tests should be performed where data availability allow
- Much of the nomenclature has been corrected already, with some specific areas such as carbohydrates and metals remaining - these should be pursued
- The different instances of molecules/structures should be versioned
    - Would help with people cleaning up their own structures
    - But what if the sequence/ligands change with new versions
    - Authentication of depositors may be necessary
    - How to deal with highly related structures (e.g. a series of re-refinements of structures that probe different parameterizations)
- One approach could be Jamboree/Hackathons to remediate structures with community buy in

- Recommendation:
  - All current and future validation tests should be performed where data availability allow
  - Much of the nomenclature has been corrected already, with some specific areas such as carbohydrates and metals remaining - these should be pursued
  - In the future it might be useful to have a community coordinated improvement of models (including ligands)
    - E.g. Jamboree/Hackathons to remediate structures with community buy in