Marc: interesting in ligand conformation and interaction energy

Pamela: Mainly working on structure screening using Astex database. Manually look every map for interesting co-crystal structure. Usually uses maps from refinement program. But initial maps were also stored.

Question: How was the ligand selected? How to handle the ambiguity from map?

The Astex database contains about 1600 ligands. Usually selecting about 100 ligands for co-crystal. The success rate was very low. The ambiguity was very high.

Question: What about resolution?

2.5 A is fine.

Kirk Clark: Developing one stop shop system for ligand-protein interaction and defining global standard of data quality for Novartis.

The minimum requirements:

Better integration of both internal and external data resources.

Provide better computational tools

The key for success is to have high quality structure and data.

Question: What resolution ranges do you have in daily work?

1.3 ~ 3.3 A. Usually

Do you use resolution as quality indicator?

Hard to say, no big difference between 1.6 A vs. 2.0 A structure.

Question: When do you think the structure is done?

Data interpretation is complicate, need cooperation between crystallographer and biologist.

Question: What is the difference between your internal database and PDB ?

In some cases it is not clear how the sample was prepared in PDB archive. Many information about sample were missing in PDB archive.

Very important to capture what is put inside co-crystal.
Question: How to know if there is any radiation damage? How to determine if we did chemistry wrong?

What is missing in PDB?

Quality of ligand, interaction between ligand and protein

Prefer keep as many annotation ad possible

Need tool to generate good quality ligand dictionary

Ensure high quality by validation tool

How to better display the PDB data?

Make user aware of missing data

Make user aware of interpretation of structural data conflicts to biological view

Better software tool to display data

Question 1:

1. If ligand exists, go with existing one.

2. Use different computational tools to get best possible initial structure.

3. The minimum requirements: should have correct bond type so that hydrogen can be predicted by chemical software

4. PDB should provide dictionary:

include multiple models (or assembly of structures?) if possible
include all tautomer and force user to identify which one is correct
capture ambiguities
include data provenance and display to user
extend residue name to more than 3 characters

Question 2:

1. Should be able to see author's density map
2. Should be able to see map coefficient file
3. Strain energy is till not clear enough from scientific point of view. Something should be re-consider in future, but not for now.

Question 3:

1. Include ambiguities
2. Include more additional annotation information
3. Allow one deposition to include multiple different states and multiple different conformations
4. Multiple models in file must be defined.
5. What's the value to have multiple models?
6. Is it possible to provide initial or partial refined models?
7. Make easy for user to deposit additional information such as partial model, no atomic coordinate model
8. Provide GUI to force use to input all components whether or not seen in map
9. ED statistics should use standard calculated values
10. Allow annotation at all levels (chain vs. residue vs. atom)

Question 4:

1. For non-crystallographer review, provide additional introduction summary review, highlights the problems, what should be looking into in each section

Question 5:

1. Add wiki as one element
2. Provide all reports, do not re-refine structures
3. Oppose versioning for third party to re-refine all structures
4. Re-refine only if original authors agree

Question 6:

1. Include all hydrogens
    a. Help refinement
    b. Help to identify tautomer and chirality
2. Support ring opening and chemical modification
    a. Do not know how to capture dynamic (transition) states of chemicals
    b. Capture the experiments case by case
    c. Chemical description should be enough

My summary:

More mandatory annotation of ligand
- Information about how samples were prepared
- Ligand definition (dictionary used in refinement)
- Additional refinement restraints
- Ligand quality
- Fit quality
- Irregular conformation
- Intermolecular interaction
- Provide example and control vocabularies in deposition system such as
  - Ligand exists but not seen in density
  - Ligand exists but only seen partially in density
  - Ligand exists and seen clearly in density

Better way to deliver annotation data (other than REMARK)
- Make user aware of missing data
- Make user aware of interpretation of structural data conflicts to biological view
- Work with graphical software developers to better display data

Include all hydrogens if possible (include zero occupancy hydrogen)
- Help refinement
- Help to identify tautomer and chirality

Validation report
- Include density map
- Include atom level RSR
- Color code 2D ligand image
- Provide possible solution to fix problem(s)
- Provide secure link for journal review to see map

Additional notes for Thursday breakout session:

MN – Marc Nicklaus
PW – Pamela Williams
KC – Kirk Clark
MM – Matt Miller
JM – Joe Marcotrigiano
SS – Steven Sheriff
AJ – Andrzej Joachimiak
AP – Anil Padyana
JH – Jorg Hendle
WT – Wolfram Tempel
HY – Huanwang Yang
ZF – Zukang Feng
CG – Colin Groom
HMB – Helen Berman
SKB – Stephen Burley

All transcribed statements are (highly) paraphrased.

PW:
Astex database constructed by former GSK personnel
All structures are annotated for chemists
The Astex viewer is central to the database
Chemists must view electron density maps, including the initial maps
- omit maps, structures overlaid in viewer

Numbers: 1 data set / 10 minutes; 100 data sets / month at synchrotron; c. 500 data sets / project
Use Buster, REFMAC for refinement

WT:
Areas of ambiguous density or areas not initially thought important can turn out to be important.

PW:
Screening against fragment library, c. 400 from a library of 1600
Library is continually refined, with targeted sets added depending on the target

AJ:
With libraries, success rate is low, ambiguity high.

CG:
Do they look at Fobs(fragment)-Fobs(native)?

PW:
No, we use Fo-Fc now, though we used to.

HY:
Difference maps.

MN:
Resolution?

PW:
2.5 Angstrom resolution is okay; 2.0 sufficient when ligands are present; usually want around 1.6 for an apo- structure.

SS:
An apo- structure is very important.
Can accidentally fit something that isn't there.
Don't look at Fo-Fo much.

CG:
In the PDB, ligand structures with accompanying apo- structures are rare.

AJ:
Recommend apo- for all structures with ligands?

SS:
Apo- structure can be difficult to obtain.

PW:
Back-soaking is possible, but it is hard to know whether the active site has been emptied.

AJ:
Some proteins need to be unfolded to release bound ligands.

PW:
One ligand can be mistaken for another. The experimental information available in the PDB is often insufficient to know if the ligand has been correctly identified.
Suggestion? More experimental information in deposition....

KC:
Novartis, Biophysics: X-ray, NMR, EM
Define global standards within Novartis (9 structure groups to standardize)
Determine tools necessary to prevent problems.
All 9 groups using the same software now.

CG:
Is there a standards document?

KC:
A philosophy:
Have a gold standard structure
Lots of annotation – database as both repository and communication tool (communication through database)

Minimalistic database
Cross-references to notebooks
SMILES, internal codes
Integration hooks
Allows different groups to take different approaches
(no in-house X-ray generator, only synchrotron)
Look at structure in density
For ligands, ab initio calculations to start in order to get bond lengths and angles correct (get chemistry right; Jaguar)

AJ:
Ambiguity in ligand position is related to resolution. High resolution resolves alternate conformations.
1.7 A clear, 2.5 A unclear – publish both

PW:
Don't usually solve more than one structure for a particular protein/ligand.

HY:
How do you determine the high resolution limit?

AJ:
Rmerge and I/sigmaI for highest resolution shell

HY:
These values are rising (Rmerge) and falling (I/sigmaI) in the PDB.

MN:
What is your everyday range of resolutions?

KC:
1.3-3.3 A
Resolution of 1.8 A doesn't usually show alternate sidechain conformation
Having one very good structure is nice.

MN:
Use structures for ideas or quantitative evaluation?

SS:
Every structure is for idea generation.
Good to have "non-embarrassing" structure.
There are are 7 to 8-year old structures he might like to redo in order to use newer ligand tools.

JM:
PDB protein validation good.
What is a "done" structure?

WT:
Done when there are no more interpretable features in the difference map.
A 2.5 A structure today is different than a 2.5 A structure from a few years ago (water placement).
Have a script that adds a resolution shell, then re-checks parameters
Interpretation model independent of the map is dangerous.
Communication between groups is very important.

MM:
Protein crystallographer at Rutgers
For a bad ligand, a remark is nice, but no one reads these remarks. Need different ways to bring this information to the attention of the user.

JM:
Also for sidechains modeled in the absence of density.

ZF:
There is the question of removing side chains versus assigning them at zero occupancy.

AJ:
Comments are necessary for all ligands, regardless of whether they are good or bad.

MN:
There needs to be quality annotation in the file itself; a confidence indicator down to the atomic level (in the ATOM records).

ZF:
mmCIF can support this.

JH:
There needs to be program development to display this information.

PW:
Maps are generated differently by different investigators – it is difficult to know what the original interpreted map looks like.

ZF:
Upload of author map.

WT:
What happens with a structure becoming cluttered w/ annotations...perhaps they should just look at the map.

KC:
A non-expert needs to have the appropriate level of concern communicated to them.

SS:
Bristol Myers Squibb
4/5 in group use Buster, which has a better report than the PDB
Use MolProbity
Always put hydrogen atoms on ligands, even if they are zero occupancy – this avoids the tautomer problem.
Will take resolution as far as it will go (did 1 Angstrom structure recently), but even 3 Angstrom is better than no structure.

MN:
Which tautomer do you pick? (chairing IUPAC group on revising InChI tautomer)

SS:
No good answer. Programs know proteins better than ligands. Hydrogens on ligands help.

NM:
Presence of H atoms force consideration of chemical plausibility.

LUNCH

AJ:
Solved structures from 4 A to 0.38 A – same difficulty
If there are ligand problems, return to experimental phases.
Missing groups can result from radiation damage.

KC:
Features lost to radiation damage can be seen by processing only the first 1/3 of the data set.

PW:
There is often no quality control for ligands – people will model in what they think is there and mistakes can be made.

MN:
1. Determine the correct identity of the ligand – how to annotate if ligand was thought to be something else.
2. What to do if there is isomerization?

An additional annotation:
“Biological view is different than what is shown.” or
“Radiation damage has resulted in the ligand shown.”

(numerous specific experiments discussed)

KC:
Solve structures for (1) idea generation and (2) confidence building (that the picture matches the function)

AJ:
Structures to discover new things; looking for the unexpected.

AP:
Agios, previously Boehringer Ingelheim
Useful to capture: the ligand as added, map before the ligand is modeled,
Problem for computation: modeling programs do not use electron density maps well

SS:
PyMol normalizes maps

MN:
What annotations are missing in the PDB?

AP:
Ligand quality, fit quality (correlation coefficient)
Irregular conformations
Intermolecular interactions
(hydrogens, tautomerization)
Should be succinct

MN:
Other metrics of ligand quality?

SS:
Atom-by-atom correlation coefficients?

MN:
All metrics we have are not optimal, especially for very high resolution structures.

SS, AJ:
A way to show differential order

WT:
Isn't this correlated to occupancy and B-factor?

SS:
Some, but not entirely
“Religious war”: zero occupancy side chains vs. omitted side chain atoms

JH:
Better to capture annotations than to not have them.

AP:
GIGO “Garbage in, garbage out”
Hiccup, Prodrug propagate errors, generate erroneous dictionaries

WT:
Haven't seen problems from Prodrug, at least for the purpose of protein crystallography

SS:
Important to know if its wrong or just an outlier.

PW:
Can't judge a ligand against CCDC if it came from CCDC in the first place

WT:
Mogul makes mistakes

AP:
The ligand in the model may not match the ligand that was added to sample.

JH:
Structural genomics, SGX, Lilly
"3 stages of validation"
Database is integral to the company
validation system – second crystallographer has to sign off on structure solved by a crystallographer (every structure, internal)

PW:
Astex does this prior to PDB deposition.

JH:
Hi-throughput has created second-tier structures where only a 10-Angstrom sphere around the active site/ligand is evaluated. These are fully refined before going into the PDB, though.
crystallographers = # different structures produced (some standardization is required: global, individual validation parameters, guidelines)
At Lilly, can get N=3 (three structures for each system); multiple structures are good confirmation, but sometimes too much information
Variation for a particular protein is ~0.5 Angstrom.

PW:
The more potent the inhibitor, the higher the resolution.

SS:
This is seen for all combinations of resolution, inhibitors.

JH:
In the Lilly database, a comment field is available. Over time, common comments are given specific fields.
Example: racemic mixture goes in, only one enantiomer binds

PW:
Can override chemical database if the chemists make a mistake.

AJ:
Non-specifically bound ligands?

PW:
Compound stacking

SS:
We fit all of these, regardless of location

JH:
Annotation of specific versus nonspecific

WT:
JMOL should show density and ligand environment.

KC:
Validation versus value-added

JH:
Mandatory to have annotation field for ligand regardless of correctness and quality

SS:

Nice, but there should be suggestions and/or examples

general:
Pull-down menus

JH:
Add hydrogens to ligand (for sake of geometry), but occupancies should be consistent with resolution.

SS:
Don't like adding hydrogens if resolution doesn't warrant it, unless geometry is aided (does add them to ligands).
Oliver Smart recommends to add hydrogens at less than 2.0 Angstrom resolution.

HY:
Would add hydrogens with occupancy 1.0, regardless.

PW:
A PDB flag for radiation damage

WT:
Molprobity tool for ligands (conformational analysis)
Chemically sensitive moeities

JH:
take geometry library for ligands (non-standard heterogroups)

SS:
PDB should be building ultimate ligand dictionary
B-factors, initial omit maps, final omit maps
2D diagrams that color-code bonds and angles for quality (better than current tables in the wwPDB validation report)

AP:
Validation tool should provide information on how to correct problems

MN:
Question #1: dictionaries or QM calculations?

KC:
Already do QM in vacuo (Jaguar) already for some – works well.
QM requires initial guesses.

AP, SS:
Valuable to make tools available to the community.

MN:
Ask the PDB to have other models along with the "structure"

WT:
Could be provided within a wiki context: ancillary models, etc.

JM:
Question #5:
Re-refinement problem – need versioning

WT:
Publication is a branch point.

JM:
The problem is communication.

MN:
This is why version management is important

(Philosophical discussion on the viability of the publication-based model of research funding)

WT:
PDB as basis for wiki; wiki contains additional info and PDB is the immutable resource.

(Long discussion of versioning, resolution unclear)

Question 4: Empty ligand
Disordered ligand density / uncertainty

PW:
If not certain of binding modes, structure is set aside and not validated
Give it to computational chemist, design a new lead that breaks symmetry, solve its structure.
Missing portion of ligand:
deposited internally with full ligand with B-factors and occupancies that reflect the missing portion

WT,SS:
Only ordered atoms modeled, but it is identified as full ligand
Mass spec run to check if intact

KC:
NMR to convince you that it's there
Either way works if teams work closely

(Philosophical discussion of internecine conflict regarding zero occupancy versus not modelling)

HMB:
Annotation – matrix that gives a level of confidence:

1. validated by other techniques, but not visible
2. validated and fully modeled
3. partially observed

PW,SS:
Only deposit coordinates after approval but before submission

WT:
Validation report should include density maps or reviewers should be able to maps

MM:
Viewing maps may be impractical

PW,KC:
May be no crystallographers reviewing

KC:
Secure link to animated, rotating density map?

ZF:
Color-coded 2D diagram of ligand a la Buster report

MN:
InChI = identifier; SMILES = representation

Recommendations:
BUSTER report should be a template with orthogonal view of map.
Atom-by-atom correlation coefficients mapped to 2D diagram.