Chapter 25

Validation of Protein–Ligand Crystal Structure Models: Small Molecule and Peptide Ligands

Edwin Pozharski, Marc C. Deller, and Bernhard Rupp

Abstract

Models of target proteins in complex with small molecule ligands or peptide ligands are of significant interest to the biomedical research community. Structure-guided lead discovery and structure-based drug design make extensive use of such models. The bound ligands comprise only a small fraction of the total X-ray scattering mass, and therefore particular care must be taken to properly validate the atomic model of the ligand as experimental data can often be scarce. The ligand model must be validated against both the primary experimental data and the local environment, specifically: (1) the primary evidence in the form of the electron density, (2) examined for reasonable stereochemistry, and (3) the chemical plausibility of the binding interactions must be inspected. Tools that assist the researcher in the validation process are presented.

Key words Crystal structure, Protein–ligand complexes, Structure model validation, Small molecule ligands, Peptide ligands

1 Introduction

Accurate atomic models of protein–ligand complexes, as deposited in the Protein Data Bank (PDB [1]), are essential for many computational methods including ligand docking and structure-based drug design [2–4]. Rigorous validation of the quality of the atomic model is vital to ensure that these computational methods produce meaningful results [5–8].

Bound ligands, primarily small molecules and short peptides, DNA fragments, or related analogues, have a small mass relative to the protein target, and hence make proportionally minor contribution to the net X-ray scattering intensities. *Global* reciprocal space statistics, or indicators such as the often cited R-values, therefore contain little information specific to the ligand. Additionally, overall real space measures such as r.m.s. deviations (RMSD) from target values, particularly for bond lengths and bond angles, are also often quoted for the structure as a whole, and generally do not specifically reflect the quality of the ligand. Furthermore, ligands

Alexander Wlodawer et al. (eds.), *Protein Crystallography: Methods and Protocols*, Methods in Molecular Biology, vol. 1607,
DOI 10.1007/978-1-4939-7000-1_25,

bound to a protein are susceptible to a whole host of complicating factors including partial occupancies, enhanced mobility, conformational flexibility, and poor protein stability [9]. All of these factors act to further reduce the discrete scattering contributions of the ligand. It is therefore necessary to conduct a careful inspection of the ligand at the *local* level using *real space* quality indicators. The primary evaluation criteria that should be inspected include:

- The fit of the ligand model to the primary data in the form of the electron density.
- The compliance of the ligand model with prior expectations of reasonable stereochemistry.
- Plausible chemistry within the ligand–protein binding site interactions.

The ability to conduct strict evidence-based validation has revealed a number of instances in which ligands purportedly bound to a target protein are insufficiently supported by the primary evidence (i.e., electron density) [7, 10]. Although there may be other lines of supporting evidence (e.g., biochemical data, database annotations, functional assignments, and protein fold) for the binding of a particular ligand to the protein under study, it is essential to ensure that the placement and conformation (i.e., the pose) of any ligand in a structure model is supported by the electron density. Furthermore, it is essential that the ligand (and the protein environment to which it binds) adhere to well established guidelines of plausible stereochemistry, and suggest meaningful ligand–protein contacts.

Occasionally the quality of the electron density does not allow complete determination of the specific ligand pose, while the presence of some bound chemical entity (frequently disordered) is evident. There is currently no consensus within the structural biology community on how such cases should be described when depositing a structural model in the PDB. Therefore, without the context of the electron density, one should be aware of possible over-interpretation of ligands found in deposited structures.

2 Examination Against the Primary Evidence

The primary evidence supporting the atomic model of a crystal structure is the electron density. The electron density is obtained via Fourier reconstruction from the diffraction data (several tens to hundreds of thousands of X-ray reflection intensities) and the phase angle of each reflection [11, 12].

It is often the case that experimental phases are not used when solving the structure of a protein in complex with a small molecule or peptide ligand. In many cases the preferred method is Molecular

Replacement (MR) in which the structure of an apo-protein is placed in the crystal structure [13, 14]. These MR techniques are traditionally followed by manual adjustments of the model using local real space refinement on a graphics workstation and further rounds of computational reciprocal space refinement against the experimental diffraction data. While the traditional electron density maps obtained by this process (e.g., $2mF_o - DF_c$) are inherently biased, difference electron density maps (e.g., $mF_o - DF_c$) are almost always sufficiently clear to indicate both the presence and location of a bona fide ligand. Therefore, it is essential that difference electron density maps, specifically those calculated with the ligand omitted from the model, are consulted during the model building and ligand validation procedure.

2.1 Displaying Electron Density

Display programs such as *COOT* [8, 15] or *PyMol* [16] allow for the downloading of all the necessary data for electron density reconstruction from data repositories and web sites such as the wwPDB [17], *PDB_REDO* [18, 19], or EDS [20]. The data downloaded from these sites contains the appropriately weighted structure factor amplitudes and phases for each observed X-ray reflection. The reconstructed electron density is displayed together with the atom positions recorded in the associated model file, identified with a unique, 4-symbol PDB ID. The electron density is generally displayed as a three-dimensional contour map of density values, scaled in levels of standard deviations (sigma, σ) from the electron density mean. The standard $2mF_o - DF_c$ map (Fig. 1) is reconstructed from maximum likelihood coefficients obtained during global reciprocal space maximum posterior refinement of the model against the diffraction data [10, 22].

2.2 Real Space Correlation and Real Space R-Value

The fit of the model to the electron density is quantified by the Real Space Correlation Coefficient (RSCC) [23], the Real Space R-Value (RSR) [24], or more complex composite measures such as the Local Ligand Density Fit (LLDF) [19], just to mention a few. These values can be calculated on a local, per-residue basis, and can be obtained from the PDB validation report (http://wwpdb-validation.wwpdb.org/validservice). Additionally, many of these real space values can be displayed on a per-residue basis in *COOT*. Many software packages are available to assist the researcher in the analysis of the real space fit of the ligand to the electron density and several are highlighted in Table 1.

The main output of most real space fit algorithms is a metric that reflects the fit of the model to the electron density. For example, a perfect fit of the model to the electron density would result in an RSCC of 1.0, and values between 0.8 and 1.0 are commonly considered as good. The lower the RSCC, the more uncertainty in the positioning of the model with respect to the electron density. The RSR value is somewhat scale dependent [23], and in an

Table 1
Protein–ligand validation software

Package	Description	URL	Reference
Twilight	Small molecule and peptide ligand validation	http://bit.ly/1shcwu4	[7, 10, 21]
COOT	Molecular graphics package for model building and validation	http://bit.ly/1wJNWBy	[8, 15]
MolProbity	All atom clash scores and other validation statistics	http://bit.ly/1o1PuHW	[25]
VHELIBS	Fit of ligands and binding sites	http://bit.ly/1t5szxl	[26]
ValLigURL server	Compare conformations of ligands in the PDB	http://bit.ly/1v80yoS	[27]
MotiveValidator and *ValidatorDB*	Interactive web-based validation of ligands and residues	http://bit.ly/1tNd7Vs	[28]
PDB_REDO	Updated and optimized X-ray structure models and maps	http://bit.ly/1pVYQQA	[18, 19]
wwPDB	wwPDB Validation Server	http://bit.ly/1xdfoZN and http://bit.ly/1si1ZeL	[19]
PRIVATEER	Checks glycan nomenclature and stereochemistry	http://bit.ly/1Rc6C5e	[29]
LIGPLOT	Ligand–protein interaction diagrams	http://bit.ly/1qwZ7cS	[30]
EDS	Electron density server	http://bit.ly/1Ddnlkg	[20]
EDSTATS	Statistical quality indicators of electron density maps	http://bit.ly/1xv0VI8	[23, 31]
OVERLAPMAP	Average of two maps	http://bit.ly/ZZUSk3	[32]
BUSTER-TNT	Refinement of proteins and ligands	http://bit.ly/1w8NX1Q	[33]
WHAT_IF WHAT_CHECK	Protein and ligand verification tools	http://bit.ly/1F0j19Y	[31]

attempt to overcome this issue, the associated measure used by the PDB, the RSRZ value, is a normalized statistic that quantifies how many standard deviations a particular residue deviates from the mean electron density fit. Figure 1 illustrates how the RSCC is generally better for high resolution data and lower B-factors of the ligand (specifically Occupancy Weight Adjusted B-factors; OWAB, as detailed in Fig. 1).

2.3 Difference and Omit Electron Density Maps

Difference electron density maps are essential for the full and accurate validation of any protein–ligand model. A typical difference electron density map, with coefficients of $mF_o - DF_c$, can be simultaneously displayed with a traditional $2mF_o - DF_c$ map using programs like *COOT*. Difference electron density maps reveal *positive*

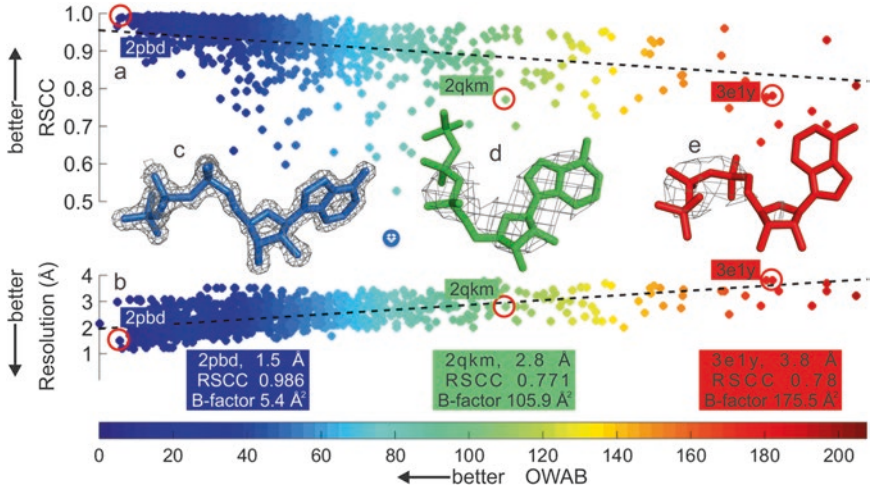

Fig. 1 (**a**) Real Space Correlation Coefficient (RSCC) and typical electron density in relation to (**b**) resolution and occupancy weighted B-factor (OWAB). In general, the higher the resolution of the data and the lower the B-factors of the ligand, the better the fit of the ligand model to the electron density. Data points are shown in each plot for adenosine triphosphate (ATP) bound structures deposited in the PDB (**c**–**e**). RSCC is a metric used to determine the *local* measure of ligand fit to the electron density and OWAB is a *local* measure of the displacement of the atoms of the ligand. RSCC and OWAB values were calculated using *Twilight* [10, 21]. These plots demonstrate how the *local* RSCC and OWAB metrics of the ligand depend closely on *global* metrics such as the resolution of the diffraction data. The ATP is shown as sticks (**c**–**e**) and the maximum likelihood $2mF_o - DF_c$ electron density map is shown as a grey mesh contoured at 2σ. Reproduced with permission from [5]

difference density where parts of the ligand model are missing, and *negative difference density* where the ligand model is present but is not supported by electron density.

Omit difference electron maps are a second type of difference electron density map that can be calculated as a powerful validation of ligand presence. In fact, the most definitive proof positive for ligand density is obtained through inspection of positive omit difference maps. These omit electron density maps are obtained by refinement of the model *sans* the part to be examined, i.e., without including the ligand in question in the refinement. If the data are obtained from a crystal structure with the ligand present, strong positive difference density, in the shape of the ligand, will be present.

Omit difference maps are not directly available from public databases, and therefore need to be generated for each specific protein–ligand complex. While there is nothing particularly difficult about this process, it does require general knowledge of file formats and crystallographic software. The ligand in question is first deleted from the structural model, either by editing the PDB file directly or by using model editing tools available in *COOT* and *PyMol*. The resulting ligand-free model is then subjected to a cycle of standard crystallographic refinement. For thorough examination, and to confirm that the difference electron density for the ligand in question is robustly reproduced, it is recommended to

employ multiple refinement programs (e.g., *REFMAC* [34], *phenix.refine* [35], or *BUSTER-TNT* [33]). All modern crystallographic refinement programs produce output files that contain map coefficients suitable for calculation of both standard $2mF_o - DF_c$ maps and $mF_o - DF_c$ difference maps that can be directly imported into *COOT* to display the resulting omit electron density.

2.4 Model Bias

Calculation of difference omit electron density maps is particularly important for ligand validation due to the phenomenon of model bias. Traditional electron density maps, as available from public databases such as EDS and PDB-REDO, are constructed using phases originating from a model that includes the ligand in question. It is therefore not surprising that some spurious electron density is always present in the model map that may seem to confirm the presence of the ligand in the structure. Figure 2 illustrates the phenomenon of model bias by comparing, side-by-side, the electron density maps from the original PDB deposition (biased $2mF_o - DF_c$ map from EDS) with those from corrected sources utilizing difference omits maps (correct $mF_o - DF_c$ map from *PDB_REDO* or *BUSTER-TNT*).

2.5 Incomplete Ligand Models

We define incomplete models as those in which peptides, or any other small molecules, are enthusiastically modeled beyond what is clearly traceable in the electron density. Enthusiastic ligand modeling results in poor quality scores for parts of the ligand, or of the peptide molecule, that are not correctly modeled. As a result, local quality scores, which are typically averaged over the whole ligand, are relatively good, giving the false appearance of a correct model. Furthermore, overall quality measures of electron density, such as ligand RSCC, RSR, RSRZ, and the LLDF, are also artificially lowered by the presence of otherwise correct regions in the model. In addition, the percentile of Ramachandran plot outliers [25, 37] can be very low for short peptide ligands. These examples highlight that the only reliable method for the validation of a protein–ligand complex is inspection of the ligand, on a local basis, and its fit to the electron density. Using these methods it is possible to determine which parts of the ligand model are experimentally defined and consistent with the electron density. It is important to note that limiting the ligand model to the pieces that are visible in the electron density, while crystallographically honest and at first glance in the spirit of parsimony, can lead to incorrect functional assumptions or hypotheses. For example, modeling of only a small portion of a large peptide ligand (e.g., only five modeled residues of a 33 residue peptide) will often result in a model requiring independent supporting evidence to determine the correct register of the peptide [7].

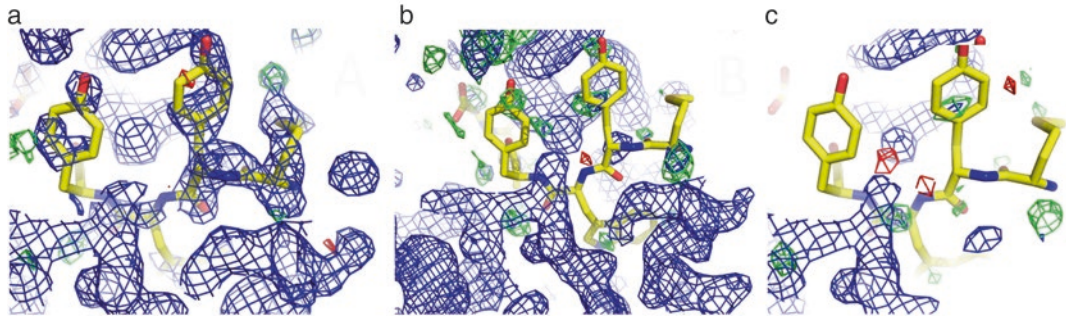

Fig. 2 Evidence of model bias in electron density maps. Shown are (**a**) $2mF_o - DF_c$, (**b**) and (**c**) $mF_o - DF_c$ electron density maps contoured at 1σ (*blue*) of the same region of a concanavalin A/peptide complex at a resolution of 1.93 Å (PDB ID:1jw6) [36]. Electron density maps were calculated by (**a**) EDS, (**b**) *PDB_REDO* and (**c**) *BUSTER-TNT* after refinement of the model with the peptide molecule omitted. The peptide is shown as *yellow sticks* and the protein model has been omitted for clarity. Figures were rendered with *PyMol* [16]. Reproduced with permission from [7]

3 Examination Against Prior Expectations

Basic empirical epistemology requires that an atomic model be examined for agreement with the primary experimental evidence. Also, it is important that the atomic model be weighted by its compliance with independently acquired prior knowledge; a well-refined ligand model built into unambiguous electron density will, in general, exhibit reasonable stereochemistry and plausible ligand–protein interactions.

3.1 Stereochemical Restraints

Presumptions about correct stereochemistry are incorporated in maximum posterior refinement in the form of stereochemical restraints [22, 38]. Stereochemical restraints effectively act as data points, stabilizing the otherwise underdetermined, or only weekly over-determined, macromolecular model refinement. The stereochemistry of the ligand is described in a restraint file. The restraint file contains general well-established information such as the bond lengths, bond angles, planarity restraints, or preferred torsions of the ligand. For small molecules and new chemical entities, particular great care must be taken to ensure that the specific chemistry is also correctly described, including delocalization, tautomerization, and charge state. If the restraints for the ligand are incorrectly described, the crystallographic data, and its contribution to the refinement target function, may not be strong enough to guide the correct placement, pose, and stereochemistry of the ligand; in such cases any resulting deviation from the electron density may not be evident from analysis of standard real space metrics.

In the case of peptide ligands, backbone torsion angles are not restrained, and therefore they provide an excellent geometric real space cross-validation of the model [7]. Peptide ligand models

that lack good supporting electron density will almost always refine to implausible high energy conformations; these are easily detectable as outliers in a Ramachandran plot (Fig. 3). The percentile of Ramachandran outliers for peptide ligands is readily available from PDB validation reports (http://wwpdb-validation.wwpdb.org/validservice).

Another effective validation tool is to compare the B-factors of the ligand with those of the protein atoms in the immediate vicinity of the ligand binding site. This can be carried out by inspection of the model in *PyMol* or *COOT*. It is expected that the B-factors of interacting parts of the ligand and the protein will have similar B-factors; significant differences may be indicative of partial occupancy or, in some cases, of incorrectly built ligand molecules.

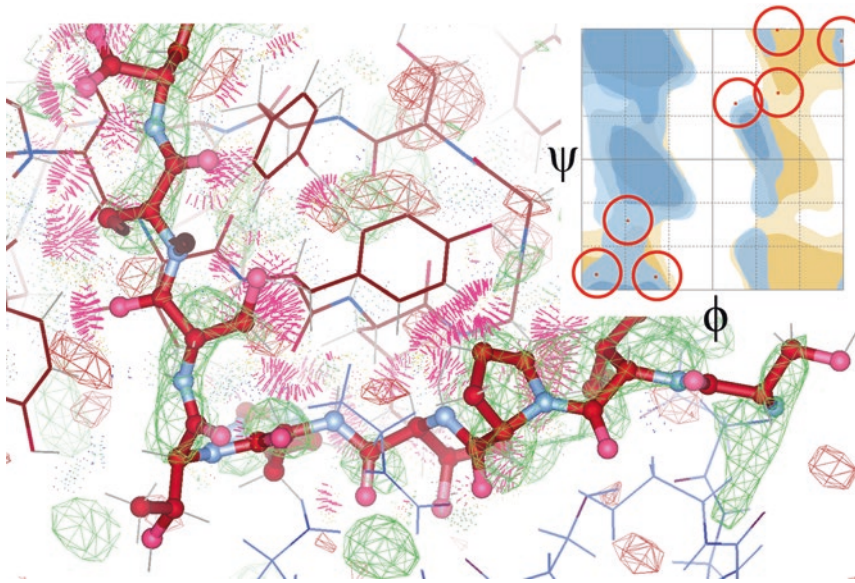

Fig. 3 An improbable peptide ligand. The main panel shows the model of a peptide antigen (ball-and-stick model) bound to a Fab antibody fragment (thin sticks) (PDB entry 2a6i, chain P, [39]). The positive $mF_o - DF_c$ omit difference electron density (calculated with the ligand omitted during maximum likelihood map calculation) is shown as a green grid contoured at 2.5σ above the mean density. In this example, the peptide model should be surrounded by clear positive difference electron density resembling the distinct shape of the peptide. However, only discontinuous fragments of positive difference electron density are visible, which can be explained in part by ordered solvent (the *round green spheres* are typical for water molecules). In addition to the poor fit to the electron density, the antigen model has a multitude of unreasonably close contacts with the Fab (visualized as *red spikes*) and has an implausibly high-energy backbone conformation as denoted by the backbone torsion angles located in unfavourable regions of the Ramachandran plot (*top right insert*). The image was prepared using *COOT* [8] and the $mF_o - DF_c$ map reconstructed from maximum likelihood coefficients computed by *REFMAC* [34]. The backbone torsion angles and interatomic clashes were calculated using *MolProbity* [37]. Updated figure with permission from [40]

3.2 Clash Scores and Binding Site Chemistry

A reasonably placed ligand will have clearly defined interactions with amino acid residues of the target protein. Common problems include steric clashes between a poorly placed ligand and the target protein (Fig. 3), and the absence of sensible contacts between the ligand and the target.

An all-atom close contact analysis (i.e., clash score) can be computed using *MolProbity* and visualized using *COOT* [37]. The program displays a list of bad atom contacts and a visual display of dots and spikes representing the contacts. Figure 3 highlights the crystal structure of an implausible Fab-Antigen complex (PDB entry 2a6i, chain P, [39]). In this example, the combination of poor agreement with the primary evidence (i.e., the electron density), together with violation of any established prior expectations, suggests that the presence of the ligand, as modeled, is unlikely.

Some validation programs, notably *VHELIBS* [26], also evaluate the quality of the interactions in the ligand binding site and rank the ligand on the basis of these quality scores. A summary of validation tools is provided in Table 1.

4 Practice of Ligand Evaluation

The degree of ligand validation that is required depends on the context in which the protein–ligand model will be used. For example, validation of a protein–ligand structure for use in structure-based lead discovery will typically require a higher level of scrutiny, and generally a resolution of around 2 Å or better. Conversely, less scrutiny is required when validating a protein–ligand complex that is being used to analyze global features of the ligand binding site for use in techniques such as structure-guided mutagenesis.

Figure 4 summarizes a useful decision tree for ligand inspection and validation of protein–ligand models. The following steps are a recommended best practice procedure for ligand validation.

4.1 Data and Program Setup

- Does the selected structure model provide enough detail, particularly resolution, to serve your interest?
- Download the full PDB validation report (https://www.ebi.ac.uk/pdbe/).
- Download the coordinates and electron density from EDS or *PDB_REDO* in *COOT*. If the electron density is not available, full validation is not possible.
- Turn on symmetry related atoms.
- Turn on environment distances.
- Select the ligand and zoom-in. If you need an omit map, proceed to the next section, otherwise inspect the $2mF_o - DF_c$ electron density.

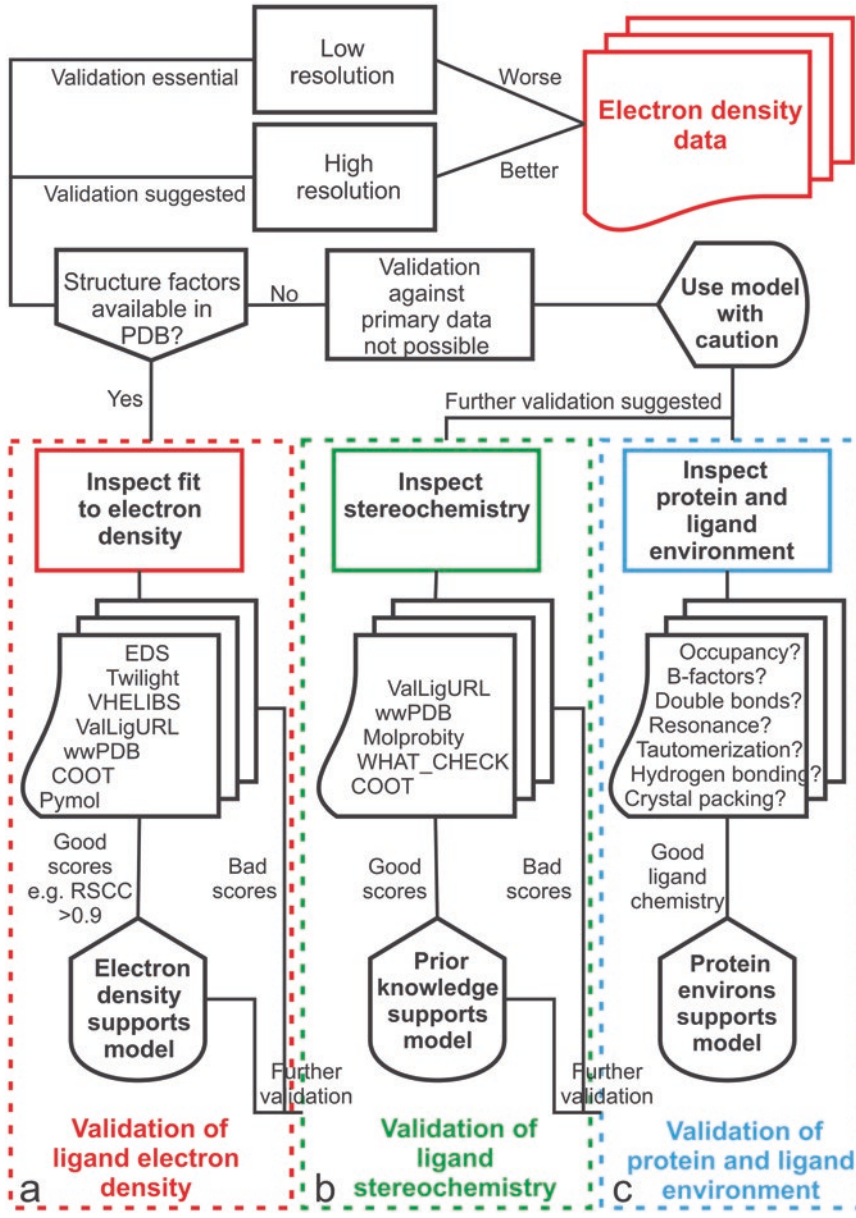

Fig. 4 Decision tree for interpretation of crystallographic data and validation of protein–ligand models. A pathway of recommended examinations starts with the electron density data (*top right*). Important steps include, (**a**) validation of the ligand electron density with a particular focus on ensuring the electron density supports the ligand model, (**b**) validation of the ligand stereochemistry and confirmation that the ligand model is supported by prior expectations and (**c**) validation of the protein and ligand environment and confirmation that the environment supports the ligand model. Reproduced with permission from [5]

4.2 Prepare an Omit Map

- Download the coordinates from the PDB (https://www.ebi.ac.uk/pdbe/).
- Download the reflection data from the PDB and convert it to mtz format using the CCP4i structure factor utilities (http://www.ccp4.ac.uk/examples/tutorial/html/intro-tutorial.html).
- Alternatively, download the mtz file corresponding to the "fully optimized structure" from *PDB_REDO* (http://www.cmbi.ru.nl/*PDB_REDO*). Another tool for converting experimental data from the PDB into the proper input format is phenix.cif_as_mtz from the PHENIX software package (https://www.phenix-online.org/documentation/reference/cif_as_mtz.html).
- Delete the ligand from the model. This can be done by manually editing the text of the PDB file or by using *COOT*. Ligand atom records in the PDB file are identified by "HETATM" and the 3-letter ligand identifier. The PDB file might contain additional LINK/CISPEP records for the ligand in the header section which also need to be removed. Atoms belonging to the protein are denoted by lines beginning with "ATOM", these lines and the "CRYST1" lines need to be retained.
- Run a single cycle of refinement, with the ligand-omitted model, as a starting structure. It is often informative to run several different refinement programs as the results may differ. *REFMAC*, *phenix.refine* and *BUSTER-TNT* are widely used crystallographic refinement programs and simple graphical user interfaces are available for *REFMAC* (http://ccp4wiki.org/~ccp4wiki/wiki/index.php?title=Refinement_with_*REFMAC*5) and *phenix.refine* (https://www.phenix-online.org/documentation/reference/refine_gui.html).
- The output mtz file of the refinement procedure contains the $2mF_o - DF_c$ omit map (default blue) and $mF_o - DF_c$ difference omit map (green, positive difference density, red negative difference density) that can be inspected in *COOT* (https://www2.mrc-lmb.cam.ac.uk/personal/pemsley/coot/).

4.3 Electron Density Inspection

Inspect the $2mF_o - DF_c$ electron density. Does the blue grid surround the model? If not, then it is likely that further validation is required.

- Inspect the omit $mF_o - DF_c$ difference electron density. Import the original PDB model and examine the omit difference density around the ligand molecule. It is common to view the difference electron density at a 3σ level. It may be informative to calculate an omit map after deleting a single protein residue in the binding site—this will provide a good estimate of the magnitude of the difference omit density due to bona fide structural elements.
- Evaluate how much of the ligand molecule can be clearly placed within the electron density map. Partial fits are a fairly

common situation for peptide ligands, lipid molecules and glycosylation sites. In these cases, the most plausible explanation for the partial fit to the electron density is disorder in the ligand, as a result of weak interactions with the protein, or poor occupancy of the ligand binding site.

- Evaluate whether the ligand molecule can be fit into the electron density in a different orientation or conformation.
- Evaluate whether the electron density allows for alternative explanations. Examples include: (1) strings of structured water molecules bound to the ligand binding site, (2) components of the mother liquor or cryo-protectant (e.g., short polyethylene glycol fragments), and (3) terminal fragments or loop regions of a symmetry related protein molecule.
- Examine the "residue density fit" graph in *COOT*, paying particular attention to the ligand and residues within the ligand binding site (https://www2.mrc-lmb.cam.ac.uk/Personal/pemsley/*COOT*/web/docs/*COOT*.html#Validation-Graphs).

4.4 Environmental Examination

- Are symmetry-related molecules involved in ligand binding? If yes, further evaluation using biochemical or biophysical techniques may be required to address if these interactions are real or crystal packing artifacts.
- Are there multiple copies of the same molecule present in the asymmetric unit (ASU, non-crystallographic symmetry)? If yes, then all of the ligand binding sites within the ASU should be examined and validated. Can different conformations or occupancies of the ligand be explained by nonequivalent environments, plasticity, or crystal packing?
- Examine the bond distances within the environment of the ligand binding site. These can be displayed using *COOT*. Are the non-covalent interactions plausible?
- Compute the clash score via the validation menu in *COOT*. Are there any significant clashes between the ligand and the target protein? Can they be explained or perhaps corrected?

5 Tools

There are several tools available for the validation of ligands (Table 1). They include standalone programs, server-based applications on the web, and databases containing collections of ligands scored by various quality criteria. As always, the quality and the suitability of a protein–ligand structure for a given purpose is context-sensitive; not every protein–ligand structure is fit for structure based lead discovery, and not every missing piece of a ligand is reason for panic. As already mentioned before: *trust, but verify* [5].

6 Outlook and New Techniques

Several recent developments in structural biology, such as ultrafast electron diffraction (UED), double electron–electron resonance (DEER), atomic force microscopy, cryo-electron microscopy (cryo-EM), and X-ray free-electron lasers (XFEL) [41, 42] pose an extra set of challenges for ligand validation, as they generally produce structures of lower resolution, typically in the 4–10 Å range. It is essential that special emphasis be placed on the ligand restraints that are applied. New hybrid techniques allow probing the dynamics of ligand binding on new timescales, and in the case of XFEL, analysis of pico- and femto-scale dynamics are possible [43]. Much of the chemistry, and in particular ligand stereochemistry, is yet to be explored on such timescales and new definitions of ligand stereochemistry will be required to ensure that accurate restraints are maintained for structures determined using these techniques.

Structural biology lies at the forefront of biomedical research and drug discovery and it is essential that the structural biology community gets serious about effective ligand validation. It requires effort at all levels to ensure that protein–ligand structures deposited in the PDB, and used by the scientific community, are of the highest quality [44]. It is essential that validation methods continue to evolve and embrace new hybrid techniques. Mandatory validation efforts and task forces, such as those initiated by the PDB, will be essential to this effort [19].

Acknowledgments

BR receives partial support from the Austrian Science Foundation (FWF) under project P28395-B26.

References

1. Berman H, Henrick K, Nakamura H (2003) Announcing the worldwide Protein Data Bank. Nat Struct Biol 10:980
2. Blundell TL, Jhoti H, Abell C (2002) High-throughput crystallography for lead discovery in drug design. Nat Rev Drug Discov 1:45–54
3. Blundell TL, Patel S (2004) High-throughput X-ray crystallography for drug discovery. Curr Opin Pharmacol 4:490–496
4. Congreve M, Murray CW, Blundell TL (2005) Structural biology and drug discovery. Drug Discov Today 10:895–907
5. Deller MC, Rupp B (2015) Models of protein–ligand crystal structures: trust, but verify. J Comput Aided Mol Des 29:817–836
6. Pozharski E, Weichenberger CX, Rupp B (2012) Techniques, tools and best practices for ligand electron-density analysis and results from their application to deposited crystal structures. Acta Crystallogr D Biol Crystallogr 69:150–167
7. Weichenberger C, Pozharski E, Rupp B (2016) Twilight reloaded: the peptide experience. Acta Crystallogr D Biol Crystallogr 72: 211–222
8. Debreczeni JE, Emsley P (2012) Handling ligands with Coot. Acta Crystallogr D Biol Crystallogr 68:425–430
9. Deller MC, Kong L, Rupp B (2016) Protein stability: a crystallographer's perspective. Acta Crystallogr F Struct Biol Commun 72:72–95

10. Pozharski E, Weichenberger CX, Rupp B (2013) Techniques, tools and best practices for ligand electron-density analysis and results from their application to deposited crystal structures. Acta Crystallogr D Biol Crystallogr 69:150–167
11. Rupp B (2009) Biomolecular crystallography: principles, practice, and application to structural biology. Garland Science, New York
12. Rhodes G (2006) Crystallography made crystal clear. Academic Press, London, UK
13. Rossmann M (ed) (1972) The molecular replacement method. Gordon and Breach Science Publishers, New York
14. Evans P, McCoy A (2008) An introduction to molecular replacement. Acta Crystallogr D Biol Crystallogr 64:1–10
15. Emsley P, Lohkamp B, Scott WG et al (2010) Features and development of Coot. Acta Crystallogr D Biol Crystallogr 66:486–501
16. DeLano WL (2008) The PyMOL molecular graphics system. DeLano Scientific, Palo Alto, CA
17. Dutta S, Burkhardt K, Swaminathan GJ et al (2008) Data deposition and annotation at the Worldwide Protein Data Bank. In: Kobe B, Guss M, Huber T (eds) Structural proteomics: high-throughput methods. Humana Press/ Springer, New York
18. Joosten RP, Womack T, Vriend G et al (2009) Re-refinement from deposited X-ray data can deliver improved models for most PDB entries. Acta Crystallogr D Biol Crystallogr 65:176–185
19. Read RJ, Adams PD, Arendall WB 3rd et al (2011) A new generation of crystallographic validation tools for the protein data bank. Structure 19:1395–1412
20. Kleywegt GJ, Harris MR, Zou J-Y et al (2004) The Uppsala Electron-Density Server. Acta Crystallogr D Biol Crystallogr 60:2240–2249
21. Weichenberger CX, Pozharski E, Rupp B (2013) Visualizing ligand molecules in twilight electron density. Acta Crystallogr F Struct Biol Commun 69:195–200
22. Tronrud D (2004) Introduction to macromolecular refinement. Acta Crystallogr D Biol Crystallogr 60:2156–2168
23. Tickle IJ (2012) Statistical quality indicators for electron-density maps. Acta Crystallogr D Biol Crystallogr 68:454–467
24. Brändén CI, Jones TA (1990) Between objectivity and subjectivity. Nature 343:687–689
25. Ramachandran GN, Ramakrishnan C, Sasisekharan V (1963) Stereochemistry of polypeptide chain configurations. J Mol Biol 7:95–99
26. Cereto-Massague A, Ojeda MJ, Joosten RP et al (2013) The good, the bad and the dubious: VHELIBS, a validation helper for ligands and binding sites. J Cheminform 5:36
27. Kleywegt GJ, Harris MR (2007) ValLigURL: a server for ligand-structure comparison and validation. Acta Crystallogr 63:935–938
28. Varekova RS, Jaiswal D, Sehnal D et al (2014) MotiveValidator: interactive web-based validation of ligand and residue structure in biomolecular complexes. Nucleic Acids Res 42:W227–W233
29. Agirre J, Iglesias-Fernandez J, Rovira C et al (2015) Privateer: software for the conformational validation of carbohydrate structures. Nat Struct Mol Biol 22:833–834
30. Laskowski RA, Swindells MB (2011) LigPlot+: multiple ligand-protein interaction diagrams for drug discovery. J Chem Inf Model 51:2778–2786
31. Joosten RP, te Beek TA, Krieger E et al (2011) A series of PDB related databases for everyday needs. Nucleic Acids Res 39:D411–D419
32. Winn MD, Ballard CC, Cowtan KD et al (2011) Overview of the CCP4 suite and current developments. Acta Crystallogr D Biol Crystallogr 67:235–242
33. Smart OS, Womack TO, Flensburg C et al (2011) Better ligand representation in BUSTER protein-complex structure determination. Acta Crystallogr A 67:C134
34. Murshudov GN, Skubak P, Lebedev AA et al (2011) REFMAC5 for the refinement of macromolecular crystal structures. Acta Crystallogr D Biol Crystallogr 67:355–367
35. Afonine PV, Grosse-Kunstleve RW, Echols N et al (2012) Towards automated crystallographic structure refinement with phenix. refine. Acta Crystallogr D Biol Crystallogr 68:352–367
36. Zhang Z, Qian M, Huang Q et al (2001) Crystal structure of the complex of concanavalin A and hexapeptide. J Protein Chem 20(5):423–429
37. Chen VB, Arendall WB III, Headd JJ et al (2010) MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr D Biol Crystallogr 66:12–21
38. Konnert J (1976) A restrained-parameter structure-factor least-squares refinement procedure for large asymmetric units. Acta Crystallogr A 32:614–617
39. Sethi DK, Agarwal A, Manivel V et al (2006) Differential epitope positioning within the germline antibody paratope enhances promiscuity in the primary immune response. Immunity 24:429–438

40. Rupp B (2016) Only seeing is believing: the power of evidence and reason. Adv Biochem (Postępy Biochemii) 62:250
41. Wakatsuki S (2014) Structural biology applications of synchrotron radiation and X-ray free-electron lasers. In: Jaeschke E, Khan S, Schneider RJ, Hastings BJ (eds) Synchrotron light sources and free-electron lasers: accelerator physics, instrumentation and science applications. Springer International Publishing, Switzerland, pp 1–39
42. Lander GC, Saibil HR, Nogales E (2012) Go hybrid: EM, crystallography, and beyond. Curr Opin Struct Biol 22:627–635
43. Neutze R (2014) Opportunities and challenges for time-resolved studies of protein structural dynamics at X-ray free-electron lasers. Philos Trans R Soc London B Biol Sci 369:20130318
44. Rupp B, Wlodawer A, Minor W et al (2016) Correcting the record of structural publications requires joint effort of the community and journal editors. FEBS J 283(24):4452–4457