Analysis of the Quality of Macromolecular Structures

Dinakar M. Salunke*

International Centre for Genetic Engineering and Biotechnology

Aruna Asaf Ali Marg, New Delhi-110 067. India.

*Corresponding author email: dinakar.salunke55@gmail.com

Abstract

Structure determination utilizing X-ray crystallography involves collection of diffraction data, determination of initial phases followed by iterative rounds of model building and crystallographic refinement to improve the phases and minimize the differences between calculated and observed structure factors. At each of these stages, a variety of statistical filters exist to ensure appropriate validation. Biologically important observations often come from interpretations of signals that need to be carefully deciphered from noise and therefore human intervention is as important as the automated filters. Currently, all structural data are deposited in the Protein Data Bank and this repository is continuously evolving to incorporate possible new improvements in macromolecular crystallography. The journals that publish data arising from structural studies modulate their policies to take cognizance of new improved methodologies. The PDB and journals have evolved an accepted protocol to ensure the integrity of crystallographic results. As a result, the quality of available data and interpretations are becoming better over the years. However, there have been periodic efforts by some individuals who misuse validation mechanisms to selectively target published research through spurious challenges. These actions do more harm to the field of structural biology and runs counter to their claim to cleanse the system. The scientific systems in structural biology are robust and capable of self-correction and unwarranted vigilantism is counterproductive.

Introduction

As in other fields, scientific data in biology is peer-reviewed and documented. It is available to other scientists so that new experiments could be further designed to provide next level of insights into biological processes under scrutiny. Therefore, quality control of all scientific data in biology is crucial. The structural biology community in the world has been at the forefront of the efforts to render open access to data and resources. Macromolecular crystallography involves crystallization, X-ray diffraction data collection, and determination of phases to generate an electron density map, which is used to build an initial model. This preliminary model is then subjected to iterative rounds of model building and refinement to generate a final model that can be analyzed to obtain deep insights regarding the biological function of the macromolecule(s) under scrutiny. Although the steps are straightforward, and currently blessed with a large degree of automation, a significant number of crystals are of a quality that renders structure determination, model building or crystallographic refinement a tricky exercise. To extract useful information from such datasets requires careful refinement, regular examination of the maps and a deviation from the practice of “black box” crystallography. The contents of the asymmetric unit have to be modeled appropriately using basic principles of protein structure and physical chemistry. This is particularly important if the molecules adopt more than one conformation. The proper estimation of the asymmetric unit contents leads to maps that are clear enough to build a model that can be used to elucidate the structural basis for molecular function.

A traditional approach to model building and crystallographic refinement is beneficial for these challenging datasets. Model building and crystallographic refinement have to be done gradually to ensure that the phases improve steadily and provide the highest probability for new features to appear in the electron density map. Even small changes in the model have to be followed by crystallographic refinement and computation of a new map. The processes of model building and refinement have to be repeated a large number of times to reduce the noise and improve the clarity in the maps. Errors in describing the contents of the asymmetric unit or excessively hasty refinement- as is done due to automation- will not give rise to a decipherable map and may lead to the erroneous judgment that the quality of data from these crystals is inadequate to enable structure determination and refinement. The process of obtaining useful information from these challenging datasets is labor intensive and different from the current practice of “black-box” crystallography where the electron density maps are examined only few times and modeling/ligand-fitting and crystallographic refinement are

carried out mostly by highly automated softwares. Also, there are further considerations in the refinement of protein structures such as, those involving conformational changes in the protein that occur due to ligand binding. The observation of such structural changes is also an important factor for establishing the presence of a ligand in the structure.

Generally, an experiment in biological crystallography is carried out in the context of a hypothesis, and the data should help prove or disprove the hypothesis. It is important to highlight that most practicing macromolecular crystallographers address a problem in biology, design biochemical experiments, execute them and then carry out the crystallographic analysis. The purification of adequate amounts of the target protein to high homogeneity and subsequent crystallization requires considerable amount of effort, and the researcher has to struggle quite a lot even before reaching the crystallographic experimental stage. With regards to macromolecular complexes associated with biology and medicine, the degree of difficulty to obtain structures is further enhanced. Indeed, as more and more complex and physiologically relevant molecular systems are selected for scrutiny, the challenges faced by crystallographers' will only amplify. The crystallographic skills are of utmost importance in dealing with these challenges.

Parameters for Quality Analysis

Electron Density Map: After one has determined the correct set of phases, the interpretation of the electron density map is the most critical step in terms of elucidation of macromolecular structure from diffraction data. Obviously the guiding principle is the molecular stereochemistry and if the protein sequence is known and the diffraction data are available at very high resolution, the process currently is near automatic. However, at relatively lower resolutions, the experience of a crystallographer becomes paramount. One relies hugely on the immediate chemical neighborhood of the polypeptide chain while building the model into electron density. The presence and fit of ligands in a crystal structure is judged on the basis of many inter-related factors. Several kinds of maps are routinely used to model the structure including $2Fo-Fc$ and simulated annealing $2Fo-Fc$ composite omit maps. Delete-refine $2Fo-Fc$ and delete-refine $Fo-Fc$ maps are used periodically for examination of the positive electron density corresponding to the ligand model being built (Bhat, 1988; Brunger et al., 1997; Read, 1986). Model building is followed by refinement and the change in R_{free} and R_{cryst} is continuously monitored to ensure correct modeling of the ligand and prevent any wrong fit or over-interpretation. The overall fit of the model to the electron

density is evaluated by the R-factors and the local fit is assessed through careful examination of the electron density maps. The goodness of fit locally can be evaluated using the Real Space Correlation Coefficient in majority of cases. However, these calculations may be considered cautiously in special cases, such as when the bound ligand shows multiple conformations or low occupancy. The excessive dependence on any single factor or automated programs can be detrimental to an unbiased evaluation of the model fit.

Electron density maps are generally analyzed at a variety of different σ cut-offs and there aren't any fixed rules for setting contour levels when building models in delete-refine *Fo-Fc* electron density maps or the other electron density maps. Experienced crystallographers critically evaluate results during the refinement process to ensure that the signal-to-noise distinction is adequately handled, and go back and forth at different contour levels rather than sticking to a specified σ cut-off while analyzing different electron density maps simultaneously. It is important to emphasize that many structures, which are crystallographically sound and physiologically interesting, have ligands visualized at the contour levels less than 3.0σ . In fact, the crystallography community has recognized that precisely defining the contour level for ligand building is not possible (Adams et al., 2016; Read and Kleywegt, 2009). We provide many examples from among the biologically relevant protein-ligand complexes reported with visualization of electron density at σ cut-offs of lower than 3.0 (Table 1). Figure 1 shows delete-refine *Fo-Fc* maps of ligands contoured at 3.0σ from two structures deposited in the Protein Data Bank (PDB). It is evident that neither of these structures would have seen the light of the day if they were subjected to 3.0σ cut-off criterion. Overall, the signal can be distinguished from noise at σ cut-off lower than 3.0 in many structures.

B-Factors: Temperature factor or B-factor is an important parameter while monitoring the quality of model during the structure refinement. However, an appropriate theoretical model for refinement of B-factors that will enable universally acceptable refinement has not yet been possible, and the available strategies are constantly evolving (Merritt, 2012). The crystallographic refinement involving temperature factors (B individual vs. B group) ought to be based on resolution and the observation to parameter ratio rather than automated procedures. Refinement of B-factors of individual atoms is acceptable at high resolution, but the similar treatment of B-factors at lower resolution leads to artifacts (Merritt, 2012). The quality of data is required to be viewed in the context of specific problem under scrutiny and its functional implications. Indeed, depending on the resolution and the context, skilled crystallographers usually adopt appropriate strategies to avoid model bias.

Recently, it has been suggested that differences in the B-factors of the ligand and the protein imply that the ligand is not bound to the protein (Stanfield et al., 2016a). This contention is basically erroneous as there are a number of situations wherein the B-factors of the ligand and the protein can be dissimilar. One reason could be that the affinity of the ligand binding is relatively low due to which not all macromolecules in the crystal have ligand bound to them and this will be reflected in the B-factors. The B-factors would also be higher if the ligand does not have rigid geometry and the contacts with the protein are limited. The resolution of a number of structures of functionally important complexes, such as those of the helicase and primase (PDB: 2R6C) involved in DNA replication is low (~ 4 Å) and the atoms refine with very high B-factors (~ 150 Å²) (Bailey et al., 2007). In comparison to the intrinsically flexible molecules such as linear peptides, rigid ligands are expected to have better defined electron density and much lower B-factors unless they bind in multiple conformations. It is anticipated that flexible ligands such as peptides may bind in more than one conformation for the backbone and the bound peptides may exhibit further variation in the side chain rotamers. To avoid constitutive activation/repression of corresponding biological processes, the majority of interactions between biological molecules have to be necessarily transient. The inherent dynamic nature of macromolecular interactions associated with physiological processes is often reflected even in static crystal structures. As a result, the application of excessively stringent parameters that ignore the biological context of the molecular interactions under scrutiny will render unnecessary rejection of the vast majority of protein-ligand complex structures.

Table 2 lists a few from the many examples to support the argument that B factors in the case of physiologically relevant bound ligands are higher than those of the protein. Most of the ligands cited here, which are either peptides or nucleic acids, exhibit higher degree of freedom in backbone conformation. Notably, due to variations in side chain conformations, peptides show even greater degree of freedom. This degree of freedom is considerably reduced when ligand binds protein with higher affinity through multiple sites of stabilizing interactions.

Ramachandran Parameters: Ramachandran map has been a time-tested method of validating protein stereochemistry. Theoretically, except due to structural constraints, the protein model should abide by the limits set according to Ramachandran map. However, it is important to recognize that the structures deposited in the PDB are experimentally obtained data and could reflect typical constraints of such data. It has been established that the crystallographic refinement will be more sensitive to Ramachandran parameters at higher

resolution as compared to that at lower resolution (Morris et al, 1992; Gunasekaran et al., 1996). During the early years, the PDB had a larger proportion of structures at low resolution as the data were often collected at home sources. Therefore, PROCHECK (Laskowski, 1993) which was based on the earlier PDB, was generous with respect to the allowed region of the Ramachandran map (Read et al., 2011). In the recent past, a large number of high resolution structures have been determined due to the availability of high intensity X-rays at synchrotron sources. This large database of high resolution structures is utilized by MolProbity (Chen et al., 2010), and this program is much more stringent in assessing stereochemical quality. Therefore, the structures that are at moderate resolution and were earlier assessed using PROCHECK, when subjected to MolProbity will naturally exhibit a large number of residues in the disallowed region. Conversely, it is also to be emphasized that good stereochemistry does not necessarily imply well-determined structure as stereochemistry could be guided by imposing tight constraints (Morris et al., 1992).

Analyses of high-resolution protein structures have shown that some residues can be present in the disallowed regions of the Ramachandran map (Gunasekaran et al., 1996; Morris et al., 1992). In some cases, these residues have been shown to be important for the function of the protein (Gunasekaran et al., 1996). It is possible that an individual phi, psi outlier is conformationally correct and arises due to biological/structural relevance in a given context. The presence of residues in the disallowed regions of the Ramachandran map is possible due to interactions formed between the backbone atoms of that residue with side chains of neighboring residues. The residue at the 51st position in the light chain is present in the disallowed region in a number of antibody Fab structures (Nair et al., 2000).

It is observed that the resolution of the structure and the number of outliers are correlated (Morris et al., 1992). It is important to note that Ramachandran parameters couldn't be directly correlated with presence/absence of the electron density, despite the emphasis to the contrary (Stanfield et al., 2016b). For low-resolution datasets, the corresponding structure could refine very well even if the stereochemistry is compromised. Therefore, a small fraction of residues is commonly observed in disallowed region of the Ramachandran map in the intermediate/low-resolution structures deposited in PDB.

Protein Data Bank Depositions

The current policy of the Protein Data Bank to present the validation reports along with the data is very useful. The depositors are provided with these validation reports and have

the opportunity of reviewing their structures before depositing. What is really appreciable is that the PDB does not sit on the judgment regarding the structures. It is left to the users to decide whether to use the data or otherwise, based on the quality parameters in the validation reports and the relevance to the biological problem under scrutiny. Many journals now ask for a submission for the report along with the manuscript to enable reviewers to use it along with electron density map figures in the manuscript to make an informed judgment about the quality and relevance of the structure to the major findings in the manuscript. Such a policy is perfectly in order and will be very useful, as the authors will have the opportunity to provide a rationale for the deviations, if any, with respect to the validation reports prior to the decision of acceptability or otherwise is made.

Supporting Biochemical Evidence

Currently, structural data is almost always supported by appropriately designed biochemical experiments. Macromolecular folding leads to surface complementary which is not incorporated in refinement methods (Read et al., 2011). Therefore, biochemical methods assessing intra- and intermolecular interactions are useful for independent validation of structural observations. Importance of biological relevance in macromolecular crystallography is such that a number of structures are never deposited or documented that do not conform with relevant biochemical/genetic experiments.

While various validation mechanisms implemented by PDB ensure quality of the deposited structural data, the one-size-fits-all approach does not hold ground, particularly with challenging datasets wherein the ligand exhibits low occupancy or multiple conformations. Due to the diversity of ligand binding affinities, protein thermodynamics and crystallographic resolution, it is not prudent to apply benchmarks of uniform stringency on structural data to assess quality.

We would like to reinforce that while dealing with biological systems several variables are critical and therefore all cases should not be subjected to uniform criteria, which may be restrictive and arbitrary. Such criteria would result in limited applicability and immense loss of genuine structural data of high biological relevance. Considering the constraints that the macromolecular crystallographic experimentations impose, one has to carefully distinguish between signal and noise. It needs to be emphasized that structural biologists determine macromolecular structures to address specific questions in biology. The structure needs to be of a quality adequate enough to answer questions that are addressed in that particular study.

Analyses using multiple parameters ought to be used, as is done by many experienced macromolecular crystallographers, some of whom are cited above. If one is to blindly follow restrictive criteria, a lot of biologically relevant structural data will go into oblivion. If this happens in case of drug-discovery studies, it would be particularly unfortunate. The erroneous strategy of seeking only high quality structural data for ligands may be especially lethal for fragment-based drug-discovery studies (Murray and Blundell, 2010). Application of automated and excessively stringent parameters will lead to unnecessary rejection of low-affinity binders that can be evolved through chemical modification to obtain binders with high affinity and specificity.

Spurious Challenges

The deposited data corresponding to published works are sometimes abused to mount spurious challenges. In fact, work from our laboratory has been subjected to such attacks (Sethi et al., 2006; Khan and Salunke, 2012, Tapryal et al., 2013; Khan and Salunke, 2014). To understand how the humoral immune system generates enough receptors that have the cumulative capacity to recognize any and all non-self antigens, we had characterized the interaction of different antibodies with peptides unearthed using a phage display library (Manivel et al., 2000; Manivel et al., 2002). Our efforts showed that flexibility in terms of backbone and side chain conformations in the peptide antigens and the antibody paratopes enables the germline antibodies to bind different antigens with adequate affinity. These studies involved structures of Fab fragments of antibodies with the low-affinity peptides, which were obtained through the refinement procedures outlined in the Introduction. These studies were questioned by a group of people who seem to be incapable of dealing with challenging datasets (Stanfield et al., 2016a&b). In spite of adequately responding to all the questions raised and the journal closing the subject by publishing their queries and our rebuttals (Salunke et al., 2016 a&b), the group did not stop misrepresenting our data in different ways.

It is not correct to post-facto apply arbitrary criteria for judging selectively the data which have been obtained using established methodologies and crystallographic experience and which have gone through due review process and are accepted based on the standard norms. It is acceptable if someone wants to analyze the entire PDB based on any new criteria. But cherry-picking specific structures with ulterior agenda and then cleverly misrepresenting the data, is by itself a fraud and needs to be stopped. The challengers appear to cleverly exploit the fact that it is relatively simple to carry out flawed refinement and generate electron density

maps that lack clarity. The harmful intentions of these individuals can be gauged from the fact that their initial efforts to get the research papers retracted were all anonymous and when journal refused to accept their argument unilaterally, they were forced to go public with their point of view. It is unfortunate when people sitting on the judgment on other people's data do not apply similar yardsticks to their own structures. This is evident when one re-analyses some of the published structures (Viola et al., 2007; Wilder et al., 2010).

It has been stated that the standard approach according to modern practice is to inspect the difference electron density omit map contoured at 3.0σ (Stanfield et al., 2016a). This is absolutely invalid. To justify 3.0σ cut-off for ligand fitting, the reasoning given subsequently (Weichenberger et al. 2017), is because it is the default contouring level in the display program Coot. To say the least, this is absolutely arbitrary. There ought to be more rational crystallographic/statistical logic for insisting on 3.0σ . It is a routine practice to vary the σ cut-off while interpreting electron density map; this is something every experienced crystallographer practices and reports. Use of delete-refine *Fo-Fc* maps at levels less than 3.0σ , often at 1.5σ contour level for visualizing the ligand model is routinely practiced (Table 1). In fact, the ligand model could not have been interpreted in the two structures (PDBs: 2J9N and 3LK0) without visualizing *Fo-Fc* electron density maps below 3σ cut-off as shown in Figure 1 (Viola et al., 2007; Wilder et al., 2010).

To state that the overall B factors of neighboring components in a crystal structure cannot differ drastically is sheer hypocrisy (Stanfield et al., 2016a), considering that two of the co-authors have published structures where the B values of the ligands (PDBs: 2J9N and 3LK0, 3LK1 and 3KL5) are very high compared to neighboring amino acid residues of the proteins to which they bind (Viola et al., 2007; Wilder et al., 2010). Figure 1 shows that two of these structures (PDBs: 2J9N, 3LK0) when viewed at 3.0σ cutoff fail to show interpretable electron density in delete-refine *Fo-Fc* maps. It would be required to look for positive omit difference density at significantly lower σ cutoff (Figure 1) while the B-factors are relatively high (Table 2). It is strange that claims are made that the present validation methods are not adequate to judge correctness of data and yet certain structures are being selectively targeting by citing the validation reports based on the very same methods (Weichenberger et al. 2017). The reality is, validation tools are fine, but there are twilight zones where the judgment based on crystallographic experience is far more vital than the automated tools. The sound judgment and abilities of the experienced crystallographer take precedence over the application of excessively stringent automated criteria to prevent the loss of precious information of

biological relevance.

It is important to note that we are dealing with not mature but germline antibodies or antibodies with polyspecificity which bind to the antigens with relatively lower affinity as the binding site is potentially flexible and therefore, as expected, antigen atoms exhibit higher B-factors (Khan and Salunke, 2012, 2014; Sethi et al., 2006; Tapryal et al., 2013). Germline antibodies possess unique thermodynamic properties that are physiologically relevant (Manivel et al., 2000). These thermodynamic properties are the result of flexibility in the germline antibody paratope, which results in an inherently lower binding affinity. Further, this flexibility is the basis for the cross-reactivity exhibited by germline antibodies, which is the primary step in antibody recognition of the antigenic universe. The situation in our structures is similar and is not unexpected given the biology of germline antibodies. The criticism of our data is unwarranted as the aforementioned researches have different and generous set of standards for their own structures but impose excessively high quality benchmarks that are incompatible with the resolution and biology on the existing norms practiced by other crystallographers in describing their structures.

The potential damage that spurious challenges can cause can be gauged from the situation involving structural studies with DNA polymerase α (Nair et al., 2004). The investigators realized that the asymmetric unit contains two molecules of the enzyme bound to each end of the substrate DNA and this entire complex occupies two conformations- related by a 180° inversion in the crystal. Gradual model building and crystallographic refinement using this description of the asymmetric unit led to electron density maps that showed the presence of a Hoogsteen base pair in the active site (Nair et al., 2004). This finding was challenged by an individual who chose to describe the contents of the asymmetric unit differently and subjected the data to “black-box” refinement and this exercise led to maps that showed no clear density in the active site (Wang, 2005). In the above example, subsequent biochemical and structural studies showed unambiguously that the initial finding was correct and above reproach (Nair et al., 2005; Johnson et al. 2005; Johnson et al., 2006; Nair et al., 2006). The entire episode highlights the strength of approach described in the Introduction and illustrates how an important finding could have been consigned to obscurity due to erroneous description of contents of the asymmetric unit and excessively hasty refinement.

Ramachandran map is indeed an excellent yardstick for quality of protein structures. Yet, a detailed analysis of the PDB shows that there exist plenty of structures in the PDB that violate the Ramachandran parameters beyond generally accepted limits. There are more than 200 structures in PDB with around 10% or greater Ramachandran outliers. Table 3 shows a

list of randomly selected structures among them. For example, Duffy-binding-like domain from *Plasmodium knowlesi* (PDB: 2C6J) and the nucleosome assembly protein S from *Plasmodium falciparum* (PDB: 3KYP) show Ramachandran outliers at the level of 10% and 15%, respectively (Gill et al., 2010; Singh et al., 2006). Also, structure of a nitrogenease from *Azotobacter vinelandii* (PDB: 1NIP) exhibits Ramachandran outliers as high as 12% (Georgiadis et al., 1992). Wlodawer et al. (1989) have deposited in PDB a structure with Ramachandran outliers above 10% (Table 3). These represent much higher level of Ramachandran outlier than in any structures from our laboratory (Khan and Salunke, 2012, Khan and Salunke, 2014; Sethi et al., 2006; Tapryal et al., 2013).

A view point article (Rupp et al., 2016) has been published essentially to criticize policies of the PDB and the journal editors for not giving in to the one sided coercion. A recent publication also continues to harp on the same issues under the guise of describing a new version of software (Weichenberger et al., 2017). In science, if there is a dispute about interpretation of data or about the conclusions made in a paper, it is by no means scientific misconduct or fraud. We believe that the journal editors follow normal practice by responding to the comments received and seeking explanations from the authors on the papers published in their journals. In some cases, the editors decide to publish both the letters and the rebuttals and in the other cases the editors reject the spurious challenges based on the responses of the authors. In either case, due process of scientific debate is followed. History is full of examples where there have been strong views on the results of others, sometimes proved valid and at other times not. In this specific matter, it is sad to see some colleagues conflate disagreement however strong with misconduct and then to follow this up by stirring a witch-hunt. The journal editors, reviewers and the PDB understand that the final judgment regarding quality of macromolecular structures is largely dependent on the context of the biological problem under scrutiny.

While it requires enormous skills and experience to obtain biochemically meaningful interpretations from macromolecular crystallographic data, it is extremely easy to carry out erroneous refinement and generate fuzzy maps to create unnecessary and invalid controversies regarding deposited structures. It is also clear that structural information, for example, regarding flexible ligands that bind to macromolecules with micromolar affinity cannot be subjected to the same criteria as that of nanomolar affinity protein-ligand complexes. Such sweepingly uniform default approaches will lead to unnecessary rejection at the initial stages itself of potentially useful data that may otherwise contain important discoveries causing enormous damage to the field. Left to the vigilante groups, many important researches with

implications to public health and biotechnology would be stillborn. In addition, the actions of such people result in journal editors becoming unnecessarily suspicious of structural papers and this will lead to several disastrous consequences, the worst being that it will drive many young scientists away from utilizing the method of macromolecular crystallography.

Acknowledgements

I would like to thank my colleagues and friends, many of whom are accomplished macromolecular crystallographers, for help and discussions.

References

- Adams, P.D., Aertgeerts, K., Bauer, C., Bell, J.A., Berman, H.M., Bhat, T.N., Blaney, J.M., Bolton, E., Bricogne, G., Brown, D., *et al.* (2016). Outcome of the First wwPDB/CCDC/D3R Ligand Validation Workshop. *Structure* *24*, 502-508.
- Adams, P.D., Afonine, P.V., Bunkóczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.-W., Kapral, G.J., and Grosse-Kunstleve, R.W. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D: Biological Crystallography* *66*, 213-221.
- Bailey, S., Eliason, W.K., and Steitz, T.A. (2007). Structure of hexameric DnaB helicase and its complex with a domain of DnaG primase. *Science* *318*, 459-463.
- Bazzicalupi, C., Ferraroni, M., Bilia, A.R., Scheggi, F., and Gratteri, P. (2013). The crystal structure of human telomeric DNA complexed with berberine: an interesting case of stacked ligand to G-tetrad ratio higher than 1:1. *Nucleic Acids Research* *41*, 632-638.
- Bhat, T.N. (1988). Calculation of an Omit Map. *Journal of Applied Crystallography* *21*, 279-281.
- Bhatt, T.K., Yogavel, M., Wydau, S., Berwal, R., and Sharma, A. (2010). Ligand-bound structures provide atomic snapshots for the catalytic mechanism of D-amino acid deacylase. *Journal of Biological Chemistry* *285*, 5917-5930.
- Brunger, A.T., Adams, P.D., and Rice, L.M. (1997). New applications of simulated annealing in X-ray crystallography and solution NMR. *Structure* *5*, 325-336.
- Cheetham, G.M., Knegt, R.M., Coll, J.T., Renwick, S.B., Swenson, L., Weber, P., Lippke, J.A., and Austen, D.A. (2002). Crystal structure of aurora-2, an oncogenic serine/threonine kinase. *Journal of Biological Chemistry* *277*, 42419-42422.
- Chen, V.B., Arendall, W.B., 3rd, Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S., and Richardson, D.C. (2010). MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* *66*, 12-21.

- Cingolani, G., Bednenko, J., Gillespie, M.T., and Gerace, L. (2002). Molecular basis for the recognition of a nonclassical nuclear localization signal by importin β . *Molecular cell* 10, 1345-1353.
- Das, K., Sarafianos, S.G., Clark, A.D., Boyer, P.L., Hughes, S.H., and Arnold, E. (2007). Crystal structures of clinically relevant Lys103Asn/Tyr181Cys double mutant HIV-1 reverse transcriptase in complexes with ATP and non-nucleoside inhibitor HBY 097. *Journal of molecular biology* 365, 77-89.
- Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallographica Section D: Biological Crystallography* 60, 2126-2132.
- Georgiadis, M.M., Komiya, H., Chakrabarti, P., Woo, D., Kornuc, J.J., and Rees, D.C. (1992). Crystallographic structure of the nitrogenase iron protein from *Azotobacter vinelandii*. *Science* 257, 1653-1659.
- Gill, J., Kumar, A., Yogavel, M., Belrhali, H., Jain, S.K., Rug, M., Brown, M., Maier, A.G., and Sharma, A. (2010). Structure, localization and histone binding properties of nuclear-associated nucleosome assembly protein from *Plasmodium falciparum*. *Malaria Journal* 9, 90.
- Gunasekaran, K., Ramakrishnan, C., and Balaram, P. (1996). Disallowed Ramachandran conformations of amino acid residues in protein structures. *Journal of Molecular Biology* 264, 191-198.
- Han, Q., Robinson, H., Cai, T., Tagle, D.A., and Li, J. (2009). Biochemical and structural properties of mouse kynurenine aminotransferase III. *Molecular and Cellular Biology* 29, 784-793.
- Johnson, R.E., Prakash, L., Prakash, S. (2005). Biochemical evidence for the requirement of Hoogsteen base pairing for replication by human DNA polymerase ϵ . *Proceedings of the National Academy of Sciences* 102, 10466-10471.
- Johnson, R.E., Haracska, L., Prakash, L., Prakash, S. (2006). Role of Hoogsteen edge hydrogen bonding at template purines in nucleotide incorporation by human DNA polymerase ϵ . *Molecular and Cellular Biology* 26, 6435-6441.
- Keitel, T., Kramer, A., Wessner, H., Scholz, C., Schneider-Mergener, J., and Hohne, W. (1997). Crystallographic analysis of anti-p24 (HIV-1) monoclonal antibody cross-reactivity and polyspecificity. *Cell* 91, 811-820.
- Khan, T., and Salunke, D.M. (2012). Structural elucidation of the mechanistic basis of degeneracy in the primary humoral response. *The Journal of Immunology* 188, 1819-1827.
- Khan, T., and Salunke, D.M. (2014). Adjustable locks and flexible keys: plasticity of epitope-paratope interactions in germline antibodies. *The Journal of Immunology* 192, 5398-5405.
- Kinoshita, T., Nakanishi, I., Warizaya, M., Iwashita, A., Kido, Y., Hattori, K., and Fujii, T. (2004). Inhibitor-induced structural change of the active site of human poly (ADP-ribose) polymerase. *FEBS letters* 556, 43-46.

- Laskowski, R.A., MacArthur, M. W., Moss, D. S. and Thornton J. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography* 26.
- Leyrat, C., Renner, M., Harlos, K., Huiskonen, J.T., and Grimes, J.M. (2014). Drastic changes in conformational dynamics of the antiterminator M2-1 regulate transcription efficiency in Pneumovirinae. *Elife* 3, e02674.
- Love, M.L., Szebenyi, D.M., Kriksunov, I.A., Thiel, D.J., Munshi, C., Graeff, R., Lee, H.C., and Hao, Q. (2004). ADP-ribosyl cyclase; crystal structures reveal a covalent intermediate. *Structure* 12, 477-486.
- Lyu, S.Y., Liu, Y.C., Chang, C.Y., Huang, C.J., Chiu, Y.H., Huang, C.M., Hsu, N.S., Lin, K.H., Wu, C.J., Tsai, M.D., *et al.* (2014). Multiple complexes of long aliphatic N-acyltransferases lead to synthesis of 2,6-diacylated/2-acyl-substituted glycopeptide antibiotics, effectively killing vancomycin-resistant enterococcus. *Journal of the American Chemical Society* 136, 10989-10995.
- Manivel, V., Bayiroglu, F., Siddiqui, Z., Salunke, D.M., and Rao, K.V. (2002). The primary antibody repertoire represents a linked network of degenerate antigen specificities. *The Journal of Immunology* 169, 888-897.
- Manivel, V., Sahoo, N.C., Salunke, D.M., and Rao, K.V. (2000). Maturation of an antibody response is governed by modulations in flexibility of the antigen-combining site. *Immunity* 13, 611-620.
- Merritt, E.A. (2012). To B or not to B: a question of resolution? *Acta Crystallographica Section D: Biological Crystallography* 68, 468-477.
- Morris, A.L., MacArthur, M.W., Hutchinson, E.G., and Thornton, J.M. (1992). Stereochemical quality of protein structure coordinates. *Proteins* 12, 345-364.
- Murray, C.W., and Blundell, T.L. (2010). Structural biology in fragment-based drug design. *Current Opinion in Structural Biology* 20, 497-507.
- Nair, D.T., Johnson, R.E., Prakash, L., Prakash, S., and Aggarwal, A.K. (2005). Human DNA polymerase ϵ incorporates dCTP opposite template G via a G.C + Hoogsteen base pair. *Structure* 13, 1569-1577.
- Nair, D.T., Johnson, R.E., Prakash, S., Prakash, L., and Aggarwal, A.K. (2004). Replication by human DNA polymerase- ϵ occurs by Hoogsteen base-pairing. *Nature* 430, 377-380.
- Nair, D.T., Johnson R.E., Prakash L., Prakash S., Aggarwal A.K. (2006). Hoogsteen base pair formation promotes synthesis opposite the 1,N6-ethenodeoxyadenosine lesion by human DNA polymerase ϵ . *Nature Structural and Molecular Biology* 13, 619-625.
- Nair, D.T., Singh, K., Sahu, N., Rao, K.V., and Salunke, D.M. (2000). Crystal structure of an antibody bound to an immunodominant peptide epitope: novel features in peptide-antibody recognition. *The Journal of Immunology* 165, 6949-6955.

- Oberfield, J.L., Collins, J.L., Holmes, C.P., Goreham, D.M., Cooper, J.P., Cobb, J.E., Lenhard, J.M., Hull-Ryde, E.A., Mohr, C.P., and Blanchard, S.G. (1999). A peroxisome proliferator-activated receptor γ ligand inhibits adipocyte differentiation. *Proceedings of the National Academy of Sciences* *96*, 6102-6106.
- Read, R.J. (1986). Improved Fourier Coefficients for Maps Using Phases from Partial Structures with Errors. *Acta Crystallographica Section A* *42*, 140-149.
- Read, R.J., Adams, P.D., Arendall, W.B., 3rd, Brunger, A.T., Emsley, P., Joosten, R.P., Kleywegt, G.J., Krissinel, E.B., Luthe, T., Otwinowski, Z., *et al.* (2011). A new generation of crystallographic validation tools for the protein data bank. *Structure* *19*, 1395-1412.
- Read, R.J. and Kleywegt, G. J. (2009). Case-controlled structure validation. *Acta Crystallographica Section D: Biological Crystallography* *65*, 140-147.
- Rupp, B., Wlodawer, A., Minor, W., Helliwell, J.R., and Jaskolski, M. (2016). Correcting the record of structural publications requires joint effort of the community and journal editors. *The FEBS Journal* *283*, 4452-4457.
- Salgado, P.S., Makeyev, E.V., Butcher, S.J., Bamford, D.H., Stuart, D.I., and Grimes, J.M. (2004). The structural basis for RNA specificity and Ca²⁺ inhibition of an RNA-dependent RNA polymerase. *Structure* *12*, 307-316.
- Salunke, D.M., Khan. T., Gaur. V., Tapryal. S., Kaur. K. (2016a) Response to Comment on Three X-ray Crystal Structure Papers. *Journal of Immunology*. 196:524-528.
- Salunke, D.M., Khan T., Gaur, V., Tapryal, S., Kaur, K. (2016b) Response to Additional Comment on Three X-ray Crystal Structure Papers. *Journal of Immunology*. 196:530.
- Santini, S., Monchois, V., Mouz, N., Sigoillot, C., Rousselle, T., Claverie, J.M., and Abergel, C. (2008). Structural characterization of CA1462, the *Candida albicans* thiamine pyrophosphokinase. *BMC Structural Biology* *8*, 33.
- Sethi, D.K., Agarwal, A., Manivel, V., Rao, K.V., and Salunke, D.M. (2006). Differential epitope positioning within the germline antibody paratope enhances promiscuity in the primary immune response. *Immunity* *24*, 429-438.
- Shiba, T., Kametaka, S., Kawasaki, M., Shibata, M., Waguri, S., Uchiyama, Y., and Wakatsuki, S. (2004). Insights into the phosphoregulation of beta-secretase sorting signal by the VHS domain of GGA1. *Traffic* *5*, 437-448.
- Singh, S.K., Hora, R., Belrhali, H., Chitnis, C.E., and Sharma, A. (2006). Structural basis for Duffy recognition by the malaria parasite Duffy-binding-like domain. *Nature* *439*, 741-744.
- Snyder, G.A., Brooks, A.G., and Sun, P.D. (1999). Crystal structure of the HLA-Cw3 allotype-specific killer cell inhibitory receptor KIR2DL2. *Proceedings of the National Academy of Sciences* *96*, 3864-3869.

- St John, F.J., Hurlbert, J.C., Rice, J.D., Preston, J.F., and Pozharski, E. (2011). Ligand bound structures of a glycosyl hydrolase family 30 glucuronoxylan xylanohydrolase. *Journal of Molecular Biology* 407, 92-109.
- Stanfield, R., Pozharski, E., and Rupp, B. (2016a). Comment on Three X-ray Crystal Structure Papers. *The Journal of Immunology* 196, 521-524.
- Stanfield, R., Pozharski, E., and Rupp, B. (2016b). Additional Comment on Three X-ray Crystal Structure Papers. *The Journal of Immunology* 196, 528-530.
- Tapryal, S., Gaur, V., Kaur, K.J., and Salunke, D.M. (2013). Structural evaluation of a mimicry-recognizing paratope: plasticity in antigen-antibody interactions manifests in molecular mimicry. *The Journal of Immunology* 191, 456-463.
- Turk, B.E., Wong, T.Y., Schwarzenbacher, R., Jarrell, E.T., Leppla, S.H., Collier, R.J., Liddington, R.C., and Cantley, L.C. (2004). The structural basis for substrate and inhibitor selectivity of the anthrax lethal factor. *Nature Structural & Molecular Biology* 11, 60-66.
- Viola, R., Carman, P., Walsh, J., Miller, E., Benning, M., Frankel, D., McPherson, A., Cudney, B., and Rupp, B. (2007). Operator-assisted harvesting of protein crystals using a universal micromanipulation robot. *Journal of Applied Crystallography* 40, 539-545.
- Wang, J. (2005). DNA polymerases: Hoogsteen base-pairing in DNA replication? *Nature* 437, E6-7; discussion E7.
- Wang, J., Song, J., Franklin, M., Kamtekar, S., Im, Y., Rho, S., Seong, I., Lee, C., Chung, C., and Eom, S. (2001). Crystal structures of the HsIVU peptidase-ATPase complex reveal an ATP-dependent proteolysis mechanism. *Structure* 9, 177-184.
- Weichenberger C.X., Pozharski, E., and Rupp, B. (2017). Twilight reloaded: the peptide experience. *Acta Crystallographica Section D: Biological Crystallography* doi:10.1107/s205979831601620x (In press)
- Wilder, P.T., Charpentier, T.H., Liriano, M.A., Gianni, K., Varney, K.M., Pozharski, E., Coop, A., Toth, E.A., Mackerell, A.D., and Weber, D.J. (2010). In vitro screening and structural characterization of inhibitors of the S100B-p53 interaction. *International Journal of High Throughput Screening* 2010, 109-126.
- Wlodawer, A., Miller, M., Jaskolski, M., Sathyanarayana, B.K., Baldwin, E., Weber, I.T., Selk, L.M., Clawson, L., Schneider, J., and Kent, S.B. (1989). Crystal Structure of Synthetic HIV-Protease. *Science* 245, 616.
- Xu, R., de Vries, R.P., Zhu, X., Nycholat, C.M., McBride, R., Yu, W., Paulson, J.C., and Wilson, I.A. (2013). Preferential recognition of avian-like receptors in human influenza A H7N9 viruses. *Science* 342, 1230-1235.
- Ye, D., Wei, M., McGuire, M., Huang, K., Kapadia, G., Herzberg, O., Martin, B.M., and Dunaway-Mariano, D. (2001). Investigation of the catalytic site within the ATP-grasp domain of *Clostridium symbiosum* pyruvate phosphate dikinase. *Journal of Biological Chemistry* 276, 37630-37639.

Zajonc, D.M., Crispin, M.D., Bowden, T.A., Young, D.C., Cheng, T.Y., Hu, J., Costello, C.E., Rudd, P.M., Dwek, R.A., Miller, M.J., *et al.* (2005). Molecular mechanism of lipopeptide presentation by CD1a. *Immunity* 22, 209-219.

Zubieta, C., Ross, J.R., Koscheski, P., Yang, Y., Pichersky, E., and Noel, J.P. (2003). Structural basis for substrate recognition in the salicylic acid carboxyl methyltransferase family. *The Plant Cell* 15, 1704-1716.

Figure 1

Figure 1: Delete refine *Fo-Fc* maps for representative structures. (A) Maps for one of the two ligands in the asymmetric unit in PDB: 3LK0 (Wilder et al., 2010), contoured at 3.0σ and 1.8σ and (B) maps for two peptides in the asymmetric unit in PDB: 2J9N (Viola et al., 2007) contoured at 3.0σ and 2.0σ. Positive and negative electron densities are shown in green and red colors, respectively. Maps were generated after deleting the ligands and refining the structures in PHENIX (Adams et al., 2010). Figures were generated using COOT (Emsley and Cowtan, 2004).

Table 1: Contour levels of *F_o-F_c* electron density maps for visualizing ligands in representative structures in PDB.

Sl. No	PDB code	Resolution (Å)	σ cut-off	Reference
1	3R6R	2.4	1.5	(Bazzicalupi et al., 2013)
2	3KO4	2.7	2.5	(Bhatt et al., 2010)
3	3E2Y	2.3	1.5	(Han et al., 2009)
4	2E2Z	2.8	1.5	(Han et al., 2009)
5	2G9Z	2.0	1.5	(Santini et al., 2008)
6	2IAJ	2.5	2.0	(Das et al., 2007)
7	1XZ0	2.8	1.8	(Zajonc et al., 2005)
8	1UVI	2.2	1.5	(Salgado et al., 2004)
9	1UVM	2.0	1.5	(Salgado et al., 2004)
10	1R15	2.4	1.7	(Love et al., 2004)
11	1PWW	2.8	2.0	(Turk et al., 2004)
12	1UJJ	2.6	1.5	(Shiba et al., 2004)

Table 2: Comparison of protein and ligand B-factors deposited in a representative structure in PDB.

Sl. No	PDB code	B-factors as reported in deposited coordinates		Resolution	Reference
		(Å ²)	(Å ²)		
		Protein	Ligand		
1	4CSA	74.2	122.9	2.3	(Leyrat et al., 2014)
2	4MFL	31.8	79.8	1.9	(Lyu et al., 2014)
3	4MFP	38.2	83.9	2.2	(Lyu et al., 2014)
4	4N60 (Chain A)	79.0	140.8	2.9	(Xu et al., 2013)
5	3KL5	58.3	94.9	2.6	(St John et al., 2011)
6	3LK0	43.5	113.5	2.0	(Wilder et al., 2010)
7	3LK1	49.8	91.3	1.8	(Wilder et al., 2010)
8	2J9N	12.9	69.4	1.5	(Viola et al., 2007)
9	1UVI	32.7	121.1	2.2	(Salgado et al., 2004)
10	1UVM	43.0	112.6	2.0	(Salgado et al., 2004)
11	1PWW	44.8	99.5	2.8	(Turk et al., 2004)
12	1HI6	45.1	79.5	2.6	(Keitel et al., 1997)

Table 3 Statistics of Ramachandran outliers in some randomly selected entries from PDB.

Sl No	PDB code	Resolution (Å)	Ramachandran Outlier (%)	Reference
1	3KYP	2.8	15	(Gill et al., 2010)
2	2C6J	3.0	10	(Singh et al., 2006)
3	1UK0	3.0	14	(Kinoshita et al., 2004)
4	1M6E	3.0	13	(Zubieta et al., 2003)
5	1MUO	2.9	13	(Cheetham et al., 2002)
6	1M5N	2.9	12	(Cingolani et al., 2002)
7	1G4A	3.0	13	(Wang et al., 2001)
8	1JDE	2.8	10	(Ye et al., 2001)
9	2DL2	3.0	12	(Snyder et al., 1999)
10	4PRG	2.9	11	(Oberfield et al., 1999)
11	1NIP	2.9	12	(Georgiadis et al., 1992)
12	3HVP	2.8	11	(Wlodawer et al., 1989)