A Histogram-based Outlier Profile for Atomic Structures Derived from Cryo-Electron Microscopy

Lin Chen[†]

Department of Computer Science
Valdosta State University
Valdosta, GA 31698
lichen@valdosta.edu

Jing He

Department of Computer Science
Old Dominion University
Norfolk, VA 23529
jhe@cs.odu.edu

ABSTRACT

As more atomic structures are determined from cryo-electron microscopy (cryo-EM) density maps, validation of such structures is an important task. We report findings after analyzing the change of cryo-EM structures in a comparison between those released by December 2016 and those released between 2017 and 2019. The cryo-EM models created from density maps with resolution better than 6 Å were divided into six data sets. A histogram-based outlier score (HBOS) was implemented and validation reports were collected from the Protein Data Bank. The results suggest that the overall quality of EM structures released after December 2016 is better than that of structures released before 2017. The conformation qualities of most residue types might have been improved, except for Leucine, Phenylalanine, and Serine in high-resolution datasets (higher than 4 Å). We observe that structures solved from 0-4 Å resolution density maps have an almost identical HBOS profile as that of 4-6 Å resolution density maps.

CCS CONCEPTS

• **Applied computing** → **Life and medical sciences** → **Computational biology** → Molecular structural biology

KEYWORDS

protein structure; cryo-electron microscopy; validation; statistics; X-ray; machine learning; side-chain; anomaly

[†]Corresponding author

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.

ACM-BCB '19, September 7–10, 2019, Niagara Falls, NY, USA

© 2019 Association for Computing Machinery.

ACM ISBN 978-1-4503-6666-3/19/09...\$15.00

<https://doi.org/10.1145/3307339.3343865>

ACM Reference format:

Lin Chen, Jing He. 2019. A Histogram-based Outlier Profile for Atomic Structures Derived from Cryo-electron Microscopy. *In the 10th ACM Conference on Bioinformatics, Computational Biology, and Health Informatics (ACM-BCB 2019)*. ACM, Niagara Falls, NY, USA, ACM, New York, NY, 6 pages. <https://doi.org/10.1145/3307339.3343865>.

1. INTRODUCTION

Cryo-electron microscopy (cryo-EM) is an essential method to determine three-dimensional atomic structures of proteins and some RNA and DNA molecules [1-6]. Cryo-EM technique is not limited by the crystallization step that is required in X-ray crystallography technique, and hence it suits broader types of molecules [7]. Molecules are quickly frozen in solution and imaged in near-native conditions [8]. As of May 30, 2019, RCSB Protein Data Bank (PDB) (<https://www.rcsb.org/>) contains 3203 atomic structures, about 2% of 152,327 structures at RCSB PDB, solved using the cryo-EM technique, which. The number of cryo-EM deposition increases steadily from 11 in 2000 to 846 in 2018 [9]. Although both X-ray and cryo-EM build protein models from their electron density maps, they have different sample preparation, different data collection, and different data processing. As more atomic structures are derived from cryo-EM density data, it is critical to validate the quality of structure models. Developing strategies to validate such models is a challenging task.

The Protein Data Bank community convened three Validation Task Force (VTF) to develop standards, formats, and specifications for the proteins from X-ray [10], NMR [11], and EM [12] respectively. The EM VTF report published in 2012 summarized suggestions from EM VTF participants [12]. The report recommended validating both the cryo-EM density maps and their derived models. Validation of EM density maps involves creating standards for assessing image resolution [13], reporting visible structural features in accordance with the claimed resolution [14]. Validation of models involves creating criteria for assessing conformation without density maps [15], assessing a model with regard to the density maps, and assessing with regard to additional data. EM VTF recommended the development of the corresponding software [16].

Although the resolution of a density map provides a rough estimation of the quality of the data, it has been found that local resolutions provide more precise measurement [17]. Though

searching cryo-EM proteins by resolution had been removed from RCSB PDB in early 2018, the resolution of each structure exists in its information webpage and XML metadata (ftp://ftp.rcsb.org/pub/pdb/validation_reports/). In a cryo-EM density map with about 3 Å resolution [18-21], the overall conformations of sidechains are generally distinguishable. Since the early 1990s, various methods have been developed for X-ray crystallography to validate experimental data, protein models, and the fit between experimental data and protein models [22-27]. At the end of 2017, OneDep validation system [28] was launched. The validation reports are divided into three categories for X-ray, NMR, and cryo-EM respectively at PDB.

The overall quality of each structure and individual residues are analyzed using MolProbity and a few other software [27, 29-31] and a validation report is available for download from PDB [28]. The overall quality section in an EM validation report contains the clashscore, Ramachandran outliers, sidechain outliers, and RNA backbone. Each of these metrics is compared with entire PDB archive and EM archive respectively. The percentile ranks are provided in each report. The clashscore is derived from the number of clashes per thousand atoms. A clash is counted when two non-donor-acceptor atoms have overlap over 0.4 Å [32]. The Ramachandran outliers of a protein are derived from the percentage of Ramachandran outliers with respect to the total number of residues in the protein model. A residue is counted as a Ramachandran outlier if the combination of backbone torsion angle Phi (ϕ) and Psi (ψ) is in a region in which less than 0.5% reference data points are observed from near 8000 X-ray structures (Top8000 datasets) [31, 33]. A sidechain of a residue is counted as an outlier if it is in a region that has less than 0.3% of the reference data from Top8000 set [31]. The outliers of an individual residue are detected by comparing the bond, angle, torsion angle, and contact distances of a residue with its ideal values [28]. In the property plot section of the validation report, a residue is colored green, yellow, orange, or red depending on situations where no outlier, one type of outlier, two types of outliers, or three and above types of outliers exist. Note that the outliers may not represent errors in the model. Instead, those outliers may be genuine and unusual, which may be of biological interest. Validation reports are regenerated annually for the entire public archive of PDB.

We previously observed that block length is a simple but sensitive measure of sidechain conformations [34]. We introduced two new distance measures in the characterization of sidechain conformations [35, 36]. In this paper, we report initial findings using a histogram-based-outlier score (HBOS) method when it is applied to different cryo-EM datasets. The HBOS quantification method shows slightly different profiles for different cryo-EM structure datasets. In particular, newly solved cryo-EM structures show a profile closer to that of the reference data set of high-quality X-ray structures than those cryo-EM structures downloaded earlier. Our results also suggest changes in conformation quality over the last four years for individual residue types.

2. METHOD

Due to the availability of super-high-resolution atomic structures derived from X-ray crystallography, a reference data set, X-ray-1.5, was constructed. The reference data set contains 9131 protein structures that are solved using X-ray crystallography and have the resolution better than or equal to 1.5 Å. At resolution 1.5 Å, major atoms in a protein are located precisely. The protein structures in X-ray-1.5 were downloaded from the RCSB PDB website (<https://www.rcsb.org/>) with the sequence similarity less than 90% in March 2018. Since non-crystallographic symmetry is commonly seen in cryo-EM density maps [37], a chain with 95% sequence identity with any other chains in each protein has been ignored.

Figure 1. The five features ϕ , ψ , χ_1 , $d_{sidechain}$ and d_{block} of a protein residue.

Six cryo-EM datasets were analyzed. Three of them contain protein structures solved from cryo-EM density map with resolution better than 4 Å. The three cryo-EM data sets, EM-0-4-2016, EM-0-4-2018, and EM-0-4-2019, have 215, 288, and 465 proteins respectively. The other three cryo-EM data sets contain protein structures solved from cryo-EM density map with a resolution between 4 and 6 Å. They are EM-4-6-2016, EM-4-6-2018, EM-4-6-2019, which contain 163, 160, and 225 proteins respectively. The data in EM-0-4-2016 and EM-4-6-2016 were downloaded in December 2016, and therefore they represent a snapshot of the conformations in the database before that time. The data in EM-0-4-2018 and EM-4-6-2018 were released between January 2017 to March 2018, and those in EM-0-4-2019 and EM-4-6-2019 were released from April 2018 to May 2019. The structures of the six cryo-EM data sets were also downloaded from RCSB PDB.

Five features were used to describe the conformation of a protein residue - backbone torsion angle Phi (ϕ) and Psi (ψ), side-chain torsion angle (χ_1), sidechain length ($d_{sidechain}$), and block length (d_{block}) (Figure 1). The torsional angle ϕ is formed by atom C in the previous residue in the protein sequence chain and atom N, atom CA, atom C in the current residue. The ψ is formed by atom N, atom CA, atom C in the current residue and atom N in the next residue in the sequence. The χ_1 angle is the first torsion angle in the side-chain, which is formed by the N, CA, CB, and CG.

The range of three torsion angles is $0^\circ - 360^\circ$, instead of $-180^\circ - +180^\circ$ in the Ramachandran plot [38]. $d_{sidechain}$ is the distance between CA atom on the backbone and the mass centroid of the sidechain. d_{block} is the distance between the CA atom on backbone and mass centroid of the distal block of a specific residue. The blocks in a residue are defined in Chen et al. [34]. 18 of 20 residues were used since glycine (GLY) and alanine (ALA) have no χ_1 due to their small size of side-chains. Histogram-based outlier score (HBOS) of each residue was calculated (1). HBOS is an unsupervised model based on the idea of Naïve Bayes (see more details in Chen et al. [36]).

$$HBOS_j(v_1, v_2, v_3, v_4, v_5) = \sum_{i=1}^5 HBOS_j(v_i) = \sum_{i=1}^5 \log\left(\frac{1}{npdf_{i,j}(v_i)}\right) \quad (1)$$

Let $npdf_{i,j}(v_i)$ be the normalized density function value for feature i and residue j when $i = v_i$. For example, $npdf_{d_{Block,Lys}}(3.0)$ is the function value when $d_{Block} = 3.0$ Å for Lysine. The HBOS score of a residue is the summation of the five HBOS values from the five features. A residue with a high HBOS has a low probability of occurrence and its conformation is unfavorable. A residue with HBOS score of 10 and more is considered as an outlier in this paper.

For each protein, its validation report is accessible in PDB. Besides the pdf version of the validation report for each deposited protein, RCSB PDB maintains metadata in XML format. For each outlier residue labeled by HBOS, the corresponding XML file was downloaded from ftp://ftp.rcsb.org/pub/pdb/validation_reports/ using a web crawler, and the outlier information was extracted. The metadata contains the mark of “OUTLIER” for unpopular configurations. Different types of outliers are noted in the report, such as Ramachandran, rotamer, clashes, bond length or bond angle, and fit in electron density [28]. Each residue has a score based on how many types of outliers, among the five types of outliers, have been found for the residue. For example, a residue has the mark of rotamer outliers and two atom clashes, and it has a score 2. If no outlier is detected (score 0), a green color is assigned for the residue in the validation report. The Python scripts of the web crawler and the outlier detector have been deposited to Github at <https://github.com/lin-chen-VA/CSBW2019>. The Python source code and installation tutorial of a Chimera visualization tool have been deposited to <https://github.com/lin-chen-VA/chimeraplugin>.

3. RESULTS

HBOS was calculated for each residue of a protein. In general, a larger HBOS score represents a higher risk of being an outlier. We investigated the distribution of HBOS scores for each of the seven data sets. A probability histogram was derived by normalizing the distribution by the area under the curve (Figure 2). In each dataset, HBOS scores of residues in those chains that have less than 95% sequence similarity with other chains in the same protein are collected. Since X-ray-1.5 set contains high-quality protein structures, we used its probability histogram as a reference. The probability histogram for X-ray-1.5 (red solid) has

Figure 2. Probability histogram plots for seven data sets. (A) The probability histogram of residue HBOS scores in X-ray-1.5 (red solid), EM-0-4-2016 (blue solid), EM-4-6-2016 (blue dash line), EM-0-4-2018 (Cyan solid), EM-4-6-2018 (Cyan dash line), EM-0-4-2019 (Magenta solid), and EM-4-6-2019 (Magenta dash line). For clear viewing, four curves in (A) are shown in (B), and three curves in (A) are shown in (C).

a peak at HBOS value of 0.9 with a probability of 0.04 (red curve in Figure 2). For the two data sets downloaded in 2016, EM-0-4-2016 (blue solid) and EM-4-6-2016 (blue dash line), the highest probability values are (near HBOS score of 1.1) 0.03 and 0.028 respectively, which are lower than that of X-ray-1.5 dataset. For X-ray-1.5 data set, it is scarce to have HBOS score greater than 10, since only 0.088% of the entire data set have such HBOS scores. However, 0.737% of EM-0-4-2016 and 1.44% of EM-4-6-2016 set have HBOS scores greater than 10. The higher outlier percentage

of EM-4-6-2016 than that of EM-0-4-2016 may relate to the difference in resolution ranges. It is generally harder to determine sidechain conformation from a lower resolution density map. We observe that the curve of EM-0-4-2018 (cyan solid) and EM-0-4-2019 (magenta solid) are closer to the curve of X-ray-1.5 than that of EM-0-4-2016 (Figure 2 (B)). It suggests that the quality of cryo-EM protein models has improved since 2016. It is noted that the curves of EM-0-4-2019 and EM-4-6-2019 are highly overlapped (Figure 2 (C)). In fact, almost identical curves were observed for EM-0-4-2018 and EM-4-6-2018. This suggests that the resolution difference, 0 to 4 Å versus 4

using the knowledge gained from cryo-EM density maps in the modeling process.

We investigated those residues that are suggested as outliers from HBOS scores but not by OneDep system. An HBOS outlier is one with a score greater than 10, and a green label in OneDep is a residue with zero types of outliers identified. The value of each bar in Figure 3 represents the number of HBOS outliers per 1000 same type of residues in a dataset.

From 2016 to 2019, most of the residue types in 0-4Å-resolution datasets might have been modeled more accurately over time (Figure 3A), since outlier rate of the 2016 data set shows a higher bar than those of 2018 and 2019 respectively. As an example, the height of Arginine (ARG) bar is 0.56 in EM-0-4-2016 set, which is much higher than 0.08 in EM-0-4-2018 set and 0 in EM-0-4-2019 set. The sidechain quality of ARG might have been improved over time. The chart in Figure 4(A) shows that HBOS method detects less number of outliers in the 2019 data set than in the 2016 data set for 15 of the 18 residues. This may suggest improved quality in sidechain conformations. The three exceptions are LEU, PHE, and SER, with LEU and PHE having significantly higher bars in the 2019 data set (Figure 4(A)). Phenylalanine (PHE) is a residue with large sidechain. The score of PHE declines from 0.59 in EM-0-4-2016 to 0.02 in EM-0-4-2018 and bounces up to 1.24 in EM-0-4-2019. Somehow more PHE outlier conformations may have been introduced in the modeling process in 2019. It is noted that Leucine (LEU) has a higher score in EM-0-4-2018 and EM-0-4-2019 than EM-0-4-2016. More LEU outliers are observed after 2016 and most of them are located on long loops or loop between two secondary structures.

Figure 3. HBOS outliers that are not OneDep outliers. The number of outliers per 1000 same type of residues is shown for those residues that are labeled as outliers based on the HBOS score but are labeled green in OneDep system. (A) Three data sets with resolutions between 0 and 4 Å: EM-0-4-2016, EM-0-4-2018, and EM-0-4-2019; (B) Three data sets with resolutions between 4 Å and 6 Å: EM-4-6-2016, EM-4-6-2018, EM-4-6-2019.

Å to 6 Å resolutions, does not make much difference in producing HBOS outliers. It is possible that residue conformations are mostly assigned using known protein structures, rather than

Figure 4. Four Leucine (LEU) examples (A) LEU with index 17 in chain A of protein 6g72 (PDB ID), not an HBOS outlier; (B) LEU with index 167 in chain L of protein 6j0n (PDB ID), an HBOS outlier due to sidechain length; (C) LEU with index 409 in chain X of protein 6j0n (PDB ID), an HBOS outlier due to sidechain length;

(D) LEU with index 320 in chain A of protein 6fe8 (PDB ID), an HBOS outlier due to backbone torsion angle.

In Figure 3(B), the numbers of outliers of ARG, ASN, ASP, CYC, GLN, GLU, HIS, ILE, LYS, PHE, THR, TRP, and TYR drop over the years. What is interesting is the different behavior of PHE in two resolution ranges. The number of PHE outliers that are considered normal in OneDep system reduced from 2016 to 2018 in the 4–6 Å resolution sets (Figure 3(B)), but it increased in the 0–4 Å resolution sets (Figure 3(A)). It is possible that PHE conformations in 4–6 Å resolution density maps are predominantly built by choosing known PHE conformations in PDB, but those in 0–4 Å resolution density maps are mostly built using the density. Most of the residue types in the 4–6 Å resolutions (Figure 3(B)) have a higher number of outliers than the corresponding residue types in the 0–4 Å resolutions (Figure 3(A)) except for ARG. This aligns with the observation that it is in general harder to determine structure precisely from a lower resolution density map.

In order to visualize conformations of LEU outliers, one favorite LEU conformation (Figure 4(A)) in both OneDep and HBOS and three HBOS outliers are shown (Figure 4(B), 4(C), 4(D)). All four examples are labeled green in OneDep validation report. In Figure 4(B) and 4(C), the HBOS scores of $d_{sidechain}$ and d_{block} are abnormal. $d_{sidechain}$ and d_{block} of the normal configuration of LEU (in Figure 4(A)) are 2.58 Å and 3.1 Å respectively, which fall in the high favorable region of sidechain size. In Figure 4(B), the bending of the plane formed by atoms CG-CD1-CD2 toward up direction leads CD1 and CD2 drifting far away from the backbone. The values of $d_{sidechain}$ and d_{block} are 2.79 Å and 3.25 Å respectively, which are unusually long. In this case, though ϕ , ψ , and χ angle values are still in the acceptable bin of the OneDep system, the probability of observing such conformation is near zero. Similarly unusual, the plane of CG-CD1-CD2 is bent toward the backbone (Figure 4(C)). The values of 2.38 Å and 2.78 Å for $d_{sidechain}$ and d_{block} are significantly shorter than favorite values. In Figure 4(D), nine residues are missed in the model between residue 320 and residue 330. Missing segments in the model provides extra space to allocate the sidechain of LEU indexed 320. The whole sidechain appears to be "pushed" into the density cloud nearby to gain a better fitting. The model has reasonable $d_{sidechain}$, d_{block} , and χ values, but it has an unfavorable ψ value on the backbone.

4. CONCLUSION

The HBOS score was calculated using five features, including two distances, $d_{sidechain}$, and d_{block} , that are not used in OneDep system. The five features are combined to calculate HBOS scores. We explored the use of HBOS score as a complementary measure beyond the validation report produced in OneDep system. In the comparison of probability histograms of HBOS scores among the seven datasets, we observed that those structures derived from 0–4 Å resolution density maps and were released recently may have improved quality than those released by December 2016. The HBOS distribution profile also suggests

the risk in quality for those structures solved from 4–6 Å resolution density maps. In an analysis of those residues that are considered outliers based on HBOS scores but generally based on OneDep scores, our results suggest that the overall conformation quality of most types of residues has improved since 2017 except for LEU, PHE, and SER in those models derived from 0–4 Å resolution density maps.

5. ACKNOWLEDGMENTS

The work in this paper is supported by NSF DBI-1356621, NIH R01-GM062968.

REFERENCES

- [1] Zhang, X., Jin, L., Fang, Q., Hui, W. H. and Zhou, Z. H. 3.3 Å cryo-EM structure of a nonenveloped virus reveals a priming mechanism for cell entry. *Cell*, 141, 3 (2010), 472–482.
- [2] Peng, L., Ryazantsev, S., Sun, R. and Zhou, Z. H. Three-dimensional visualization of gammaherpesvirus life cycle in host cells by electron tomography. *Structure*, 18, 1 (2010), 47–58.
- [3] Narayanan, A., Vago, F. S., Li, K., Qayyum, M. Z., Yernool, D., Jiang, W. and Murakami, K. S. Cryo-EM structure of *Escherichia coli* σ (70) RNA polymerase and promoter DNA complex revealed a role of σ non-conserved region during the open complex formation. *J Biol Chem*, 293, 19 (2018), 7367–7375.
- [4] Serwer, P., Wright, E. T., Demeler, B. and Jiang, W. States of phage T3/T7 capsids: buoyant density centrifugation and cryo-EM. *Biophys Rev*, 10, 2 (2017), 583–596.
- [5] Chen, J., Zhang, C., Zhou, Y., Zhang, X., Shen, C., Ye, X., Jiang, W., Huang, Z. and Cong, Y. A 3.0-Ångstrom Resolution Cryo-Electron Microscopy Structure and Antigenic Sites of Cocksackievirus A6-Like Particles. *J Virol*, 92, 2 (2018), e01257–01217.
- [6] Jiang, W. and Tang, L. Atomic cryo-EM structures of viruses. *Curr Opin Struct Biol*, 46 (2017), 122–129.
- [7] Renaud, J.-P., Chari, A., Ciferri, C., Liu, W.-t., Rémy, H.-W., Stark, H. and Wiesmann, C. Cryo-EM in drug discovery: achievements, limitations and prospects. *Nature Reviews Drug Discovery*, 17 (06/08/online 2018), 471.
- [8] Kühlbrandt, W. The Resolution Revolution, 343, 6178 (2014), 1443–1444.
- [9] RCSB, <https://www.rcsb.org/stats/growth/em>, accessed on May 30, 2019.
- [10] Read, R. J., Adams, P. D., Arendall, W. B., 3rd, Brunger, A. T., Emsley, P., Joosten, R. P., Kleywegt, G. J., Krissinel, E. B., Lütteke, T., Otwinowski, Z., Perrakis, A., Richardson, J. S., Sheffler, W. H., Smith, J. L., Tickle, I. J., Vriend, G. and Zwart, P. H. A new generation of crystallographic validation tools for the protein data bank. *Structure*, 19, 10 (2011), 1395–1412.
- [11] Montelione, G. T., Nilges, M., Bax, A., Güntert, P., Herrmann, T., Richardson, J. S., Schwieters, C. D., Vranken, W. F., Vuister, G. W., Wishart, D. S., Berman, H. M., Kleywegt, G. J. and Markley, J. L. Recommendations of the wwPDB NMR Validation Task Force. *Structure*, 21, 9 (2013), 1563–1570.
- [12] Henderson, R., Sali, A., Baker, M. L., Carragher, B., Devkota, B., Downing, K. H., Egelman, E. H., Feng, Z., Frank, J., Grigorieff, N., Jiang, W., Ludtke, S. J., Medalia, O., Penczek, P. A., Rosenthal, P. B., Rossmann, M. G., Schmid, M. F., Schröder, G. F., Steven, A. C., Stokes, D. L., Westbrook, J. D., Wriggers, W., Yang, H., Young, J., Berman, H. M., Chiu, W., Kleywegt, G. J. and Lawson, C. L. Outcome of the first electron microscopy validation task force meeting. *Structure*, 20, 2 (2012), 205–214.
- [13] Sousa, D. and Grigorieff, N. Ab initio resolution measurement for single particle structures. *Journal of Structural Biology*, 157, 1 (2007/01/01/ 2007), 201–210.

- [14] Penczek, P. A. Resolution measures in molecular electron microscopy. *Methods in enzymology*, 482 (2010), 73-100.
- [15] Schwede, T., Sali, A., Honig, B., Levitt, M., Berman, H. M., Jones, D., Brenner, S. E., Burley, S. K., Das, R., Dokholyan, N. V., Dunbrack, R. L., Jr., Fidelis, K., Fiser, A., Godzik, A., Huang, Y. J., Humblet, C., Jacobson, M. P., Joachimiak, A., Krystek, S. R., Jr., Kortemme, T., Kryshtafovych, A., Montelione, G. T., Moul, J., Murray, D., Sanchez, R., Sosnick, T. R., Standley, D. M., Stouch, T., Vajda, S., Vasquez, M., Westbrook, J. D. and Wilson, I. A. Outcome of a workshop on applications of protein models in biomedical research. *Structure*, 17, 2 (2009), 151-159.
- [16] Baker, T. S. and Henderson, R. *Electron cryomicroscopy*. Springer Netherlands, City, 2001.
- [17] Kucukelbir, A., Sigworth, F. J. and Tagare, H. D. Quantifying the local resolution of cryo-EM density maps. *Nature Methods*, 11 (11/10/online 2013), 63.
- [18] Rawson, S., Bisson, C., Hurdiss, D. L., Fazal, A., McPhillie, M. J., Sedelnikova, S. E., Baker, P. J., Rice, D. W. and Muench, S. P. Elucidating the structural basis for differing enzyme inhibitor potency by cryo-EM, 115, 8 (2018), 1795-1800.
- [19] Autzen, H. E., Myasnikov, A. G., Campbell, M. G., Asarnow, D., Julius, D. and Cheng, Y. Structure of the human TRPM4 ion channel in a lipid nanodisc, 359, 6372 (2018), 228-232.
- [20] Kater, L., Thoms, M., Barrio-Garcia, C., Cheng, J., Ismail, S., Ahmed, Y. L., Bange, G., Kressler, D., Berninghausen, O., Sinning, I., Hurt, E. and Beckmann, R. Visualizing the Assembly Pathway of Nucleolar Pre-60S Ribosomes. *Cell*, 171, 7 (2017), 1599-1610.e1514.
- [21] Guo, J., She, J., Zeng, W., Chen, Q., Bai, X.-c. and Jiang, Y. Structures of the calcium-activated, non-selective cation channel TRPM4. *Nature*, 552 (12/06/online 2017), 205.
- [22] Adams, P. H. Z. a. R. W. G.-k. a. P. D. *Xtriage and Fest: automatic assessment of X-ray data and substructure structure factor estimation*. City, 2005.
- [23] Jones, T. A., Zou, J.-Y., Cowan, S. W. and Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallographica Section A*, 47, 2 (1991), 110-119.
- [24] Laskowski, R. A., MacArthur, M. W., Moss, D. S. and Thornton, J. M. PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Crystallogr*, 26, 2 (1993), 283-291.
- [25] Hooft, R. W. W., Vriend, G., Sander, C. and Abola, E. E. Errors in protein structures. *Nature*, 381, 6580 (1996/05/01 1996), 272-272.
- [26] Bruno, I. J., Cole, J. C., Kessler, M., Luo, J., Motherwell, W. D. S., Purkis, L. H., Smith, B. R., Taylor, R., Cooper, R. I., Harris, S. E. and Orpen, A. G. Retrieval of Crystallographically-Derived Molecular Geometry Information. *Journal of Chemical Information and Computer Sciences*, 44, 6 (2004/11/01 2004), 2133-2144.
- [27] Kleywegt, G. J., Harris, M. R., Zou, J.-y., Taylor, T. C., Wahlby, A. and Jones, T. A. The Uppsala Electron-Density Server. *Acta Crystallographica Section D*, 60, 12 Part 1 (2004), 2240-2249.
- [28] Gore, S., Sanz García, E., Hendrickx, P. M. S., Gutmanas, A., Westbrook, J. D., Yang, H., Feng, Z., Baskaran, K., Berrisford, J. M., Hudson, B. P., Ikegawa, Y., Kobayashi, N., Lawson, C. L., Mading, S., Mak, L., Mukhopadhyay, A., Oldfield, T. J., Patwardhan, A., Peisach, E., Sahni, G., Sekharan, M. R., Sen, S., Shao, C., Smart, O. S., Ulrich, E. L., Yamashita, R., Quesada, M., Young, J. Y., Nakamura, H., Markley, J. L., Berman, H. M., Burley, S. K., Velankar, S. and Kleywegt, G. J. Validation of Structures in the Protein Data Bank. *Structure*, 25, 12 (2017), 1916-1927.
- [29] Yang, H., Peisach, E., Westbrook, J. D., Young, J., Berman, H. M. and Burley, S. K. DCC: a Swiss army knife for structure factor analysis and validation. *J Appl Crystallogr*, 49, Pt 3 (2016), 1081-1084.
- [30] Afonine, P. V., Klaholz, B. P., Moriarty, N. W., Poon, B. K., Sobolev, O. V., Terwilliger, T. C., Adams, P. D. and Urzhumtsev, A. New tools for the analysis and validation of cryo-EM maps and atomic models. *Acta crystallographica. Section D, Structural biology*, 74, Pt 9 (2018), 814-840.
- [31] Williams, C. J., Headd, J. J., Moriarty, N. W., Prisant, M. G., Videau, L. L., Deis, L. N., Verma, V., Keedy, D. A., Hintze, B. J., Chen, V. B., Jain, S., Lewis, S. M., Arendall, W. B., 3rd, Snoeyink, J., Adams, P. D., Lovell, S. C., Richardson, J. S. and Richardson, D. C. MolProbity: More and better reference data for improved all-atom structure validation. *Protein science : a publication of the Protein Society*, 27, 1 (2018), 293-315.
- [32] 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. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta crystallographica. Section D, Biological crystallography*, 66, Pt 1 (2010), 12-21.
- [33] Hintze, B. J., Lewis, S. M., Richardson, J. S. and Richardson, D. C. Molprobity's ultimate rotamer-library distributions for model validation. *Proteins*, 84, 9 (2016), 1177-1189.
- [34] Chen, L. and He, J. A distance- and orientation-dependent energy function of amino acid key blocks. *Biopolymers*, 101, 6 (2014), 681-692.
- [35] Chen, L., He, J., Sazzed, S. and Walker, R. An Investigation of Atomic Structures Derived from X-ray Crystallography and Cryo-Electron Microscopy Using Distal Blocks of Side-Chains, 23, 3 (2018), 610.
- [36] Chen, L. and He, J. Using Combined Features to Analyze Atomic Structures derived from Cryo-EM Density Maps. In *Proceedings of the Proceedings of the 2018 ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics* (Washington, DC, USA, 2018). ACM, [insert City of Publication], [insert 2018 of Publication].
- [37] He, J., Schmid, M. F., Zhou, Z. H., Rixon, F. and Chiu, W. Finding and using local symmetry in identifying lower domain movements in hexon subunits of the herpes simplex virus type 1 B capsid. *J Mol Biol*, 309, 4 (Jun 15 2001), 903-914.
- [38] Ramachandran, G. N., Ramakrishnan, C. and Sasisekharan, V. Stereochemistry of polypeptide chain configurations. *Journal of Molecular Biology*, 7, 1 (1963/07/01/ 1963), 95-99.