Title

Distribution of evaluation scores for the models submitted to the second cryo-EM model challenge

Authors

Kryshtafovych, Andriy
Monastyrskyy, Bohdan
Adams, Paul D
et al.

Publication Date

2018-10-01

DOI

10.1016/j.dib.2018.08.214

Peer reviewed

Data in Brief 20 (2018) 1629–1638

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

Distribution of evaluation scores for the models submitted to the second cryo-EM model challenge

Andriy Kryshtafovych [a,*], Bohdan Monastyrskyy [a], Paul D. Adams [b,c], Catherine L. Lawson [d], Wah Chiu [e]

[a] Genome Center, University of California, Davis, 451 Health Sciences Drive, Davis, CA 95616, USA
[b] Molecular Biophysics & Integrated Bioimaging, LBNL, CA 94720, USA
[c] Department of Bioengineering, University of California Berkeley, CA 94720, USA
[d] Institute for Quantitative Biomedicine and Research Collaboratory for Structural Bioinformatics, Rutgers, The State University of New Jersey, 174 Frelinghuysen Road, Piscataway, NJ 08854, USA
[e] Department of Bioengineering, Microbiology and Immunology and Photon Science, Stanford University, James H. Clark Center, MC5447, 318 Campus Drive, Stanford, CA 94305-5447, USA

ARTICLE INFO

Article history:
Received 24 July 2018
Received in revised form
24 August 2018
Accepted 31 August 2018
Available online 8 September 2018

ABSTRACT

142 protein structure models were submitted to second Cryo-EM model challenge (2015–2016). Accuracy of the models was evaluated with 54 evaluation scores. Results of the descriptive statistical analysis of the scores are provided in this article.

Specifications table

Subject area	*Structural Biology*
More specific subject area	*Cryo-EM Models*
Type of data	*Figures*

DOI of original article: https://doi.org/10.1016/j.jsb.2018.07.006
* Corresponding author.
E-mail address: akryshtafovych@ucdavis.edu (A. Kryshtafovych).

https://doi.org/10.1016/j.dib.2018.08.214

How data was acquired	*Computational analysis*
Data format	*Analyzed*
Experimental factors	*None*
Experimental features	*None*
Data source location	*Rutgers University*
Data accessibility	http://model-compare.emdatabank.org/data/scores http://model-compare.emdatabank.org/em_score_boxplots.cgi

Value of the data

- The data reveal ranges of evaluation scores for models submitted to cryo-EM model challenge and provide descriptive statistics.
- The data show differences in accuracy of cryo-EM models generated with different modeling techniques.
- The data can serve as a benchmark for future cryo-EM modeling challenges.
- The data can be compared with the data in other structure modeling experiments (e.g., CASP [1]).

1. Data

142 protein structure models were submitted on eight modeling targets of the second Cryo-EM model challenge (2015–2016) [2]. A computational system was developed to estimate the accuracy of the models [3]. Each model was evaluated using a suite of 15 software packages and, as a result, 54 accuracy scores were generated per model. All scores are presented in tables and graphs of the dedicated web infrastructure (http://model-compare.emdatabank.org). Some scores were analyzed in the accompanying paper [3]. This article provides results of the descriptive statistical analysis of the complete set of evaluation scores.

2. Materials and methods

2.1. Model types

Each model submitted to the second cryo-EM model challenge was accompanied by basic information about the modeling technique used. This information indicated whether model was built *ab initio* by fitting coordinates to density maps or by optimizing already available related structures. Based on this information, models were binned into two categories: *ab initio* or optimized. Figs. 1–3 in this paper show distributions of the evaluation scores separately for *ab initio* and optimized models, while Figs. 4 and 5 show the distributions for all models in one dataset (see description of the data presented in the figures in Section 4.4 below).

2.2. Box plots

Data in the Figs. 1–5 are presented as box plots. Box boundaries correspond to the $Q_1 = 25$th (bottom) and $Q_3 = 75$th (top) percentiles in the data; the horizontal line inside the box corresponds to the median (Q_2). The height of the box defines the interquartile range (IQR = Q_3-Q_1). The height of the whiskers shows the range of the values outside the interquartile range, but within 1.5 IQR. The dots correspond to outliers, i.e. values outside the 1.5 IQR range.

2.3. Evaluation tracks and packages

Submitted models were evaluated in four evaluation tracks:

(1) directly from the model coordinates, i.e. without referring to density maps or other available structures (software packages used: *MolProbity 4.4* [4], *phenix.model_vs_map* module (from *PHENIX 1.11.1-2575 distribution* [5]), *DFIRE* [6], *ProQ3* [7], *QMEAN* [8]);
(2) comparing the model coordinates to those of reference structures (software packages used: *LGA* [9], *TM* [10,11], *LDDT* [12], *CAD* [13], *QS-score* [14], *IFaceCheck* [15], *phenix.chain_comparison* module [5]);
(3) checking fit of the model coordinates to the experimental 3DEM density maps (software packages used: *TEMPy 1.0* [16,17], *EMRinger* [18], *phenix.model_vs_map* module [5]);
(4) comparing coordinates of each model to those of other submitted models (software packages used: *Davis-QAconsensus* [19]).

2.4. Score distributions

Detailed explanation of the evaluation scores is provided in the accompanying paper [3].

Fig. 1 illustrates distributions of evaluation scores calculated directly from model coordinates (evaluation track 2.3.1). Panel (A) shows scores calculated on representative model subunits, while panel (B) – on whole multimeric structures. The figure includes box plots for the following scores:

Panel (A):

- Molprb(mon) – MolProbity's combined MPscore;
- Molprb(mon):rot_out – MolProbity's rotamer outlier score;
- Molprb(mon):clash – MolProbity's clash score;
- Molprb(mon):ram_fv – MolProbity's Ramachandran favored score;
- Molprb(mon):ram_out – MolProbity's Ramachandran outlier score;
- Log(-DFIRE) – logarithm of negative DFIRE energy score;
- ProQ3 – score from the ProQ3 program ran with default parameters;
- QMEAN – score from the QMEAN program ran with default parameters.

Panel (B):

- [Molprb(mult):ram_fv /Molprb(mult):clash /Molprb(mult):ram_out /Molprb(mult):rot_out] – MolProbity's [Ramachandran favored score /clash score /Ramachandran outlier score /rotamer outlier] score;
- [PHENIX:bond_rmsd /PHENIX:angle_rmsd /PHENIX:planar_rmsd /PHENIX:chiral_rmsd /PHENIX:dihedr_rmsd] – the RMSD on [bond /angle /planarity /chirality /dihedral angle] deviations calculated with the *phenix.model_vs_map* module;
- PHENIX:bond_max – the maximum deviation of bond distances from ideal values (in Å);
- PHENIX:angle_max – the maximum deviation of angles from ideal values (in deg.);
- PHENIX:planar_max – the maximum deviation of peptide bond planarity from ideal values (in deg.);
- PHENIX:chiral_max – the maximum deviation of chirality score from ideal values;
- PHENIX:dihedr_max – the maximum deviation of dihedral angles from ideal values (in deg.).

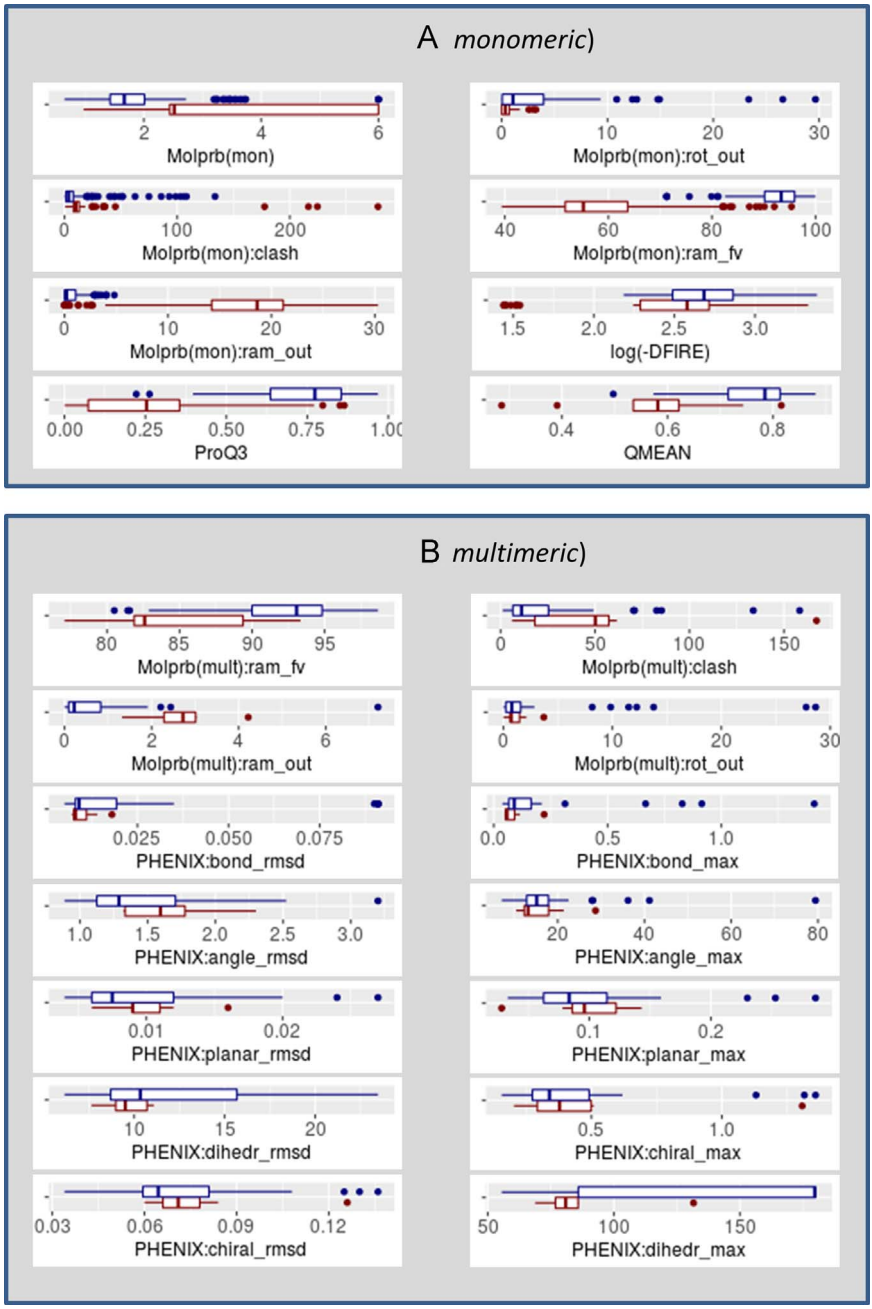

Fig. 1. Distribution of the evaluation scores calculated exclusively from the model coordinates for different types of models. For each measure (specified in the *x*-axis title), a blue boxplot shows the score distribution for models built starting from reference structure, while a red boxplot – for models built ab initio. Panel (A) shows evaluation scores for models' subunits (monomeric evaluation mode), while panel (B) for whole multimeric models.

Fig. 2 illustrates distributions of evaluation scores calculated by comparing models with reference structures (evaluation track 2.3.2). Panel (A) shows scores calculated on representative model subunits, while panel (B) – on whole multimeric structures. The figure includes box plots for the following scores:

Panel (A):

- GDT_TS, GDT_HA – scores from the LGA package ran with parameters: '-3 -sda -d:4');
- LGA_S – score from the LGA package ran with parameters: '-4 -sia -d:4');
- RMSD – root mean square deviation on Cα atoms of the representative chain (as reported by the LGA package);
- TMscore, TMalign – scores from the TM package ran with default parameters;
- CAD – CAD_aa variant of the CAD score calculated on all atoms;
- LDDT – a score from the LDDT package run with 15 Å inclusion radius.

Panel (B):

- [QS-global /QS-best] – Quaternary Structure score calculated on [all interfaces /best interface] with the *QS-score* package;
- [QS:LDDT /QS:RMSD] – [LDDT and /Cα RMSD] scores calculated on all chains by the *QS-score* package;
- IFaceCheck:F1_max – maximum *F1* statistics from among those calculated on all interfaces calculated with the *IFaceCheck* package;
- IFaceCheck:Jd_min – the minimum *Jaccard distance* from among those calculated on all interfaces;
- IFaceCheck:prec_max – maximum *precision* from among those calculated on all interfaces;
- IFaceCheck:recall_max – maximum *recall* from among those calculated on all interfaces;
- IFaceCheck:RMSD_min – minimum RMSD on target interface atoms from among those calculated on all interfaces;
- [IFaceCheck:F1_avg /IFaceCheck:Jd_avg /IFaceCheck:prec_avg /IFaceCheck:recall_avg /IFaceCheck:RMSD_avg] – the [F1 /Jaccard distance /precision /recall /interface RMSD] scores averaged over all interfaces;
- [PHENIX:CA-score /PHENIX:seq_match] –scores generated with the *phenix.chain_comparison* module.

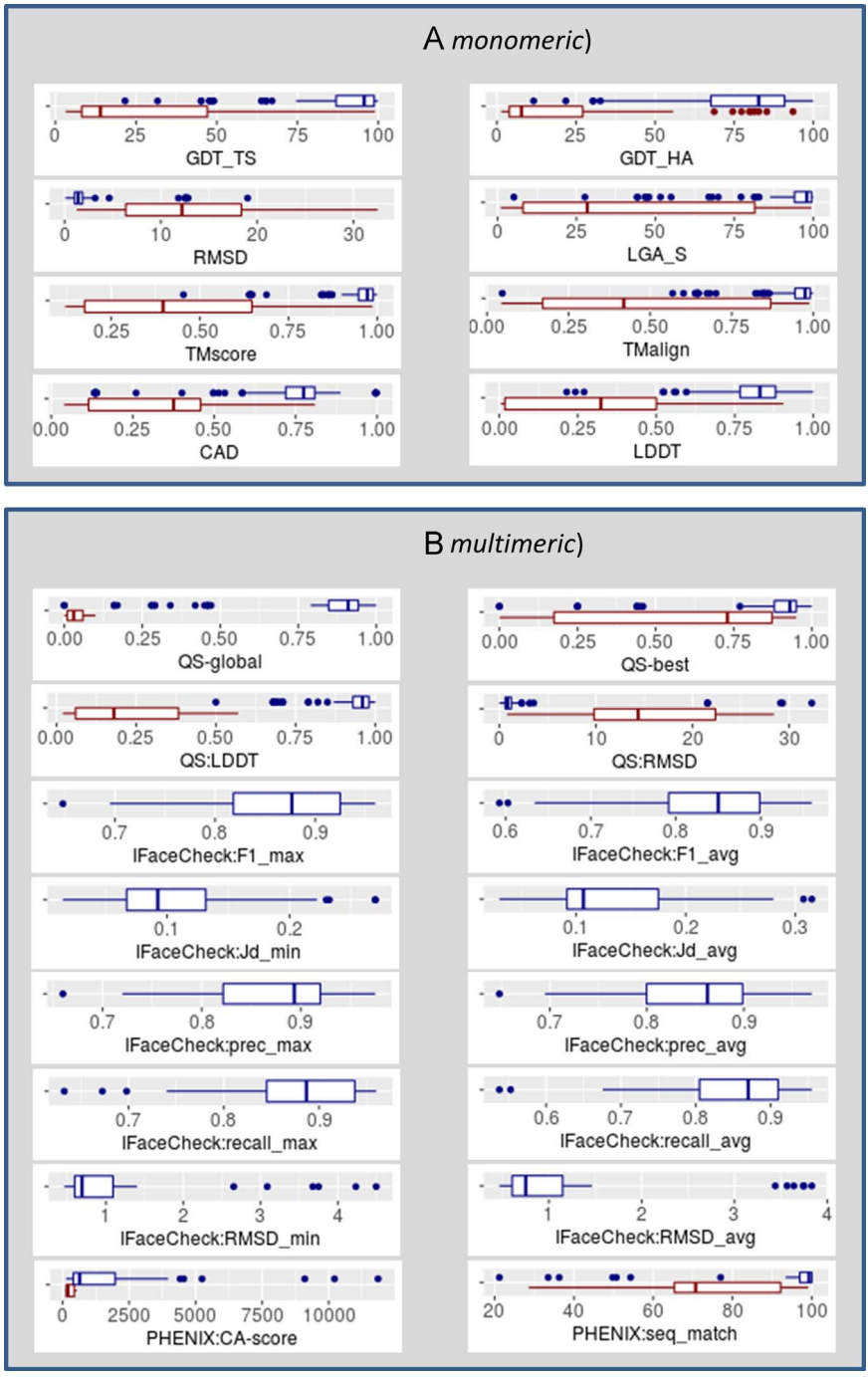

Fig. 2. Distribution of the evaluation scores calculated by comparing models with reference structures. For each measure (specified in the *x*-axis title), a blue boxplot shows the score distribution for models built starting from reference structure, while a red boxplot – for models built *ab initio*. Panel (A) shows evaluation scores for models' subunits (monomeric evaluation mode), while panel (B) for whole multimeric models.

Fig. 3 illustrates distributions of evaluation scores estimating fit of models to density maps (panel A, evaluation track 2.3.3) and similarity of models to other submitted models (panel B, evaluation track 2.3.4). The figure includes box plots for the following scores:

Panel (A):

- [PHENIX:overall_FSC /PHENIX:boxCC] – the [overall Fourier Shell Correlation in reciprocal Fourier space /per-chain box cross-correlation] calculated with the *phenix.model_vs_map* module;
- [TEMPy:CCC /TEMPy:LAP /TEMPy:ENV /TEMPy:MI] – TEMPY's [cross-correlation coefficient /Laplacian-filtered cross-correlation /envelope /mutual information] scores;
- EMRinger – EMRinger score calculated using the *phenix:emringer* module.

Panel (B):

- Davis-QA – a model consensus score calculated by averaging the GDT_TS scores from pairwise comparisons of the model to all others.

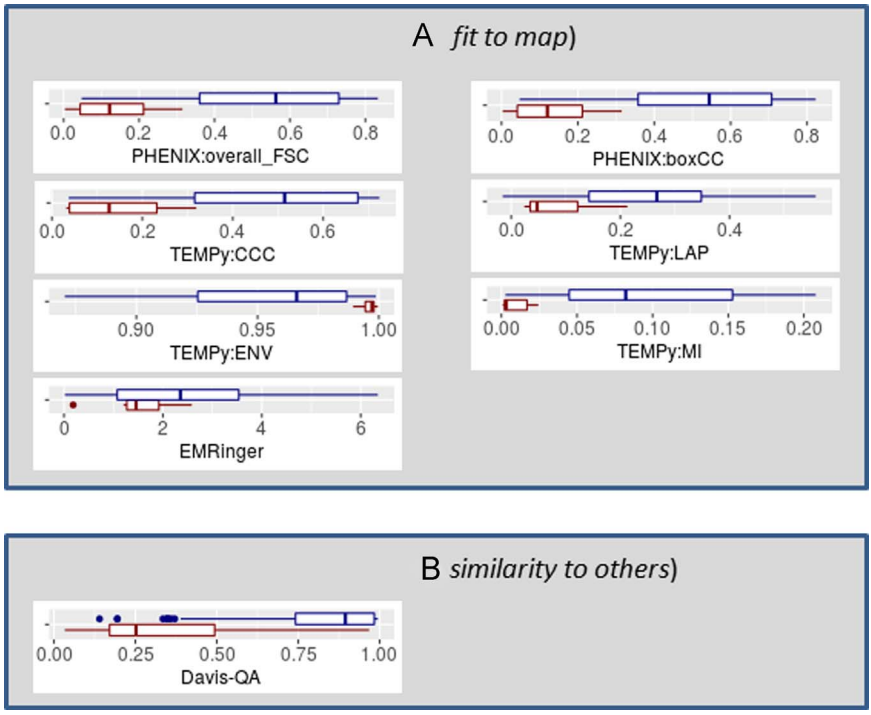

Fig. 3. Distribution of the evaluation scores estimating fit of models to density maps (panel A) and similarity of models to other submitted models (panel B). For each measure (specified in the *x*-axis title), a blue boxplot shows the score distribution for models built starting from reference structure, while a red boxplot – for models built *ab initio*.

Figs. 4 and 5 illustrate distributions of evaluation scores presented in Figs. 1–3 when all models (optimization and *ab initio*) are grouped in one dataset. Score names are as described above for Figs. 1–3.

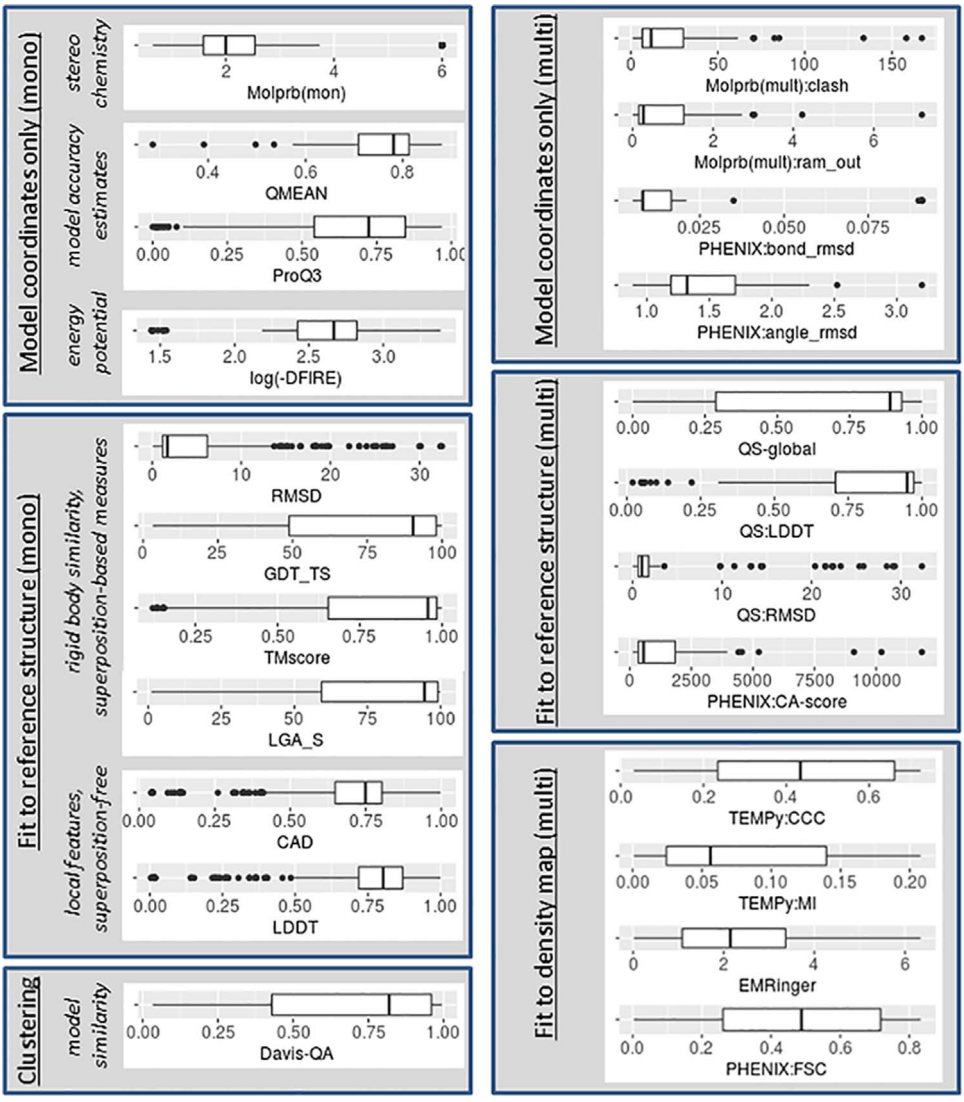

Fig. 4. Distribution of the evaluation scores shown in Fig. 8 of Ref. [3] for all submitted models (optimization and *ab initio* models pulled together). Left set of boxplots shows scores from multimeric evaluations, right set – from monomeric ones.

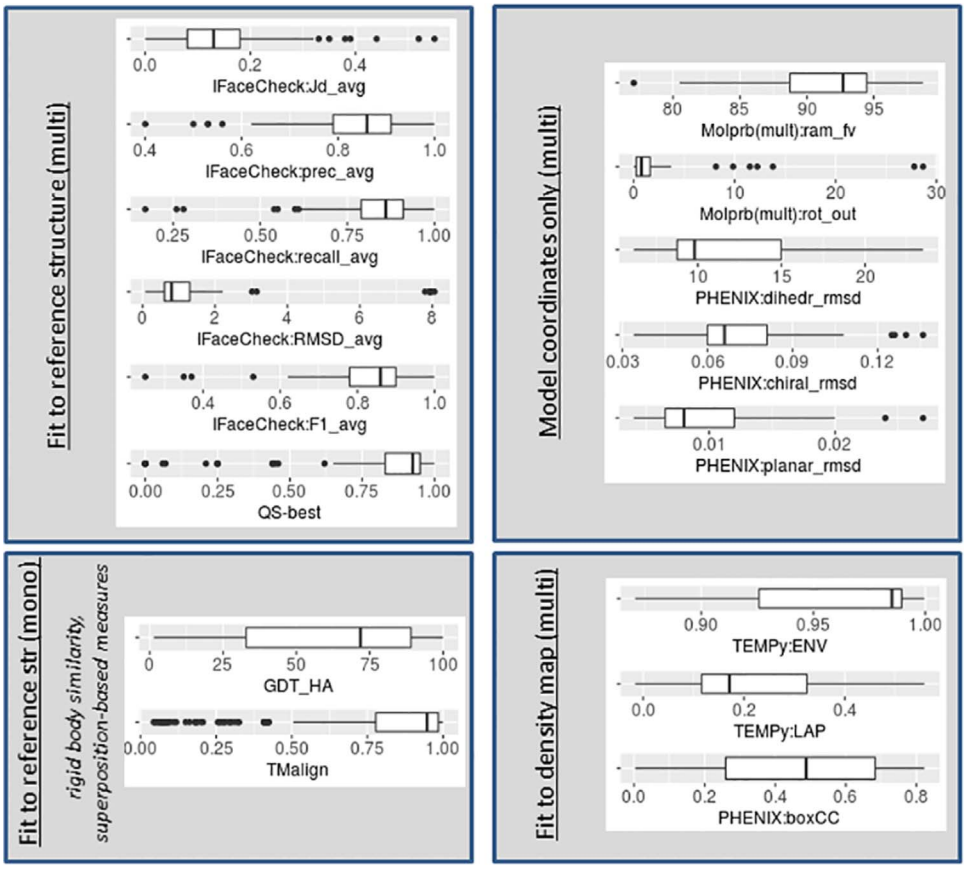

Fig. 5. Distribution of the evaluation scores not included in Fig. 4 for all submitted models.

Acknowledgements

Authors acknowledge support of the National Institute of General Medical Sciences, USA: Grant R01GM079429 to WC, and Grant P01GM063210 to PDA.

Transparency document. Supplementary material

Transparency data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.dib.2018.08.214.

References

[1] J. Moult, K. Fidelis, A. Kryshtafovych, T. Schwede, A. Tramontano, Critical assessment of methods of protein structure prediction (CASP)—round XII, Proteins Struct. Funct. Bioinforma. 86 (2018) 7–15. https://doi.org/10.1002/prot.25415.

[2] C. Lawson, A. Kryshtafovych, W. Chiu, P. Adams, A. Brünger, G. Kleywegt, A. Patwardhan, R. Read, T. Schwede, M. Topf, P. Afonine, J. Avaylon, M. Baker, T. Braun, W. Cao, S. Chittori, T. Croll, F. DiMaio, B. Frenz, S. Grudinin, A. Hoffmann, C. Hryc, A.P. Joseph, T. Kawabata, D. Kihara, B. Mao, D. Matthies, R. McGreevy, H. Nakamura, S. Nakamura, L. Nguyen, G. Schroeder, M. Shekhar, K. Shimizu, A. Singharoy, O. Sobolev, E. Tajkhorshid, I. Teo, G. Terashi, T. Terwilliger, K. Wang, I. Yu, H. Zhou, R. Sala, CryoEM models and associated data submitted to the 2015/2016 EMDataBank model challenge, Zotero (2018), https://doi.org/10.5281/zenodo.1165999.

[3] A. Kryshtafovych, P.D. Adams, C.L. Lawson, W. Chiu, Evaluation system and web infrastructure for the second cryo-EM model challenge, J. Struct. Biol. 204(1) (2018) 96–108. https://doi.org/10.1016/j.jsb.2018.07.006.
[4] V.B. Chen, W.B. Arendall 3rd, J.J. Headd, D.A. Keedy, R.M. Immormino, G.J. Kapral, L.W. Murray, J.S. Richardson, D. C. Richardson, MolProbity: all-atom structure validation for macromolecular crystallography, Acta Crystallogr. D Biol. Crystallogr. 66 (2010) 12–21. https://doi.org/10.1107/S0907444909042073S0907444909042073 (pii).
[5] P.D. Adams, P.V. Afonine, G. Bunkoczi, V.B. Chen, I.W. Davis, N. Echols, J.J. Headd, L.W. Hung, G.J. Kapral, R.W. Grosse-Kunstleve, A.J. McCoy, N.W. Moriarty, R. Oeffner, R.J. Read, D.C. Richardson, J.S. Richardson, T.C. Terwilliger, P.H. Zwart, PHENIX: a comprehensive python-based system for macromolecular structure solution, Acta Crystallogr. D Biol. Crystallogr. 66 (2010) 213–221. https://doi.org/10.1107/S0907444909052925S0907444909052925 (pii).
[6] H. Zhou, Y. Zhou, Distance-scaled, finite ideal-gas reference state improves structure-derived potentials of mean force for structure selection and stability prediction, Protein Sci. 11 (2002) 2714–2726. https://doi.org/10.1110/ps.0217002.
[7] K. Uziela, N. Shu, B. Wallner, A. Elofsson, ProQ3: improved model quality assessments using Rosetta energy terms, Sci. Rep. 6 (2016) 33509. https://doi.org/10.1038/srep33509srep33509 (pii).
[8] P. Benkert, M. Kunzli, T. Schwede, QMEAN server for protein model quality estimation, Nucleic Acids Res. 37 (2009) W510–W514. https://doi.org/10.1093/nar/gkp322 (doi:gkp322 [pii]).
[9] A. Zemla, LGA: a method for finding 3D similarities in protein structures, Nucleic Acids Res. 31 (2003) 3370–3374 ⟨http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=12824330⟩.
[10] Y. Zhang, J. Skolnick, TM-align: a protein structure alignment algorithm based on the TM-score, Nucleic Acids Res. 33 (2005) 2302–2309, https://doi.org/10.1093/nar/gki524 (doi:33/7/2302 [pii]).
[11] Y. Zhang, J. Skolnick, Scoring function for automated assessment of protein structure template quality, Proteins 57 (2004) 702–710. https://doi.org/10.1002/prot.20264.
[12] V. Mariani, M. Biasini, A. Barbato, T. Schwede, lDDT: a local superposition-free score for comparing protein structures and models using distance difference tests, Bioinformatics 29 (2013) 2722–2728. https://doi.org/10.1093/bioinformatics/btt473btt473 (pii).
[13] K. Olechnovic, E. Kulberkyte, C. Venclovas, CAD-score: a new contact area difference-based function for evaluation of protein structural models, Proteins 81 (2013) 149–162. https://doi.org/10.1002/prot.24172.
[14] M. Bertoni, F. Kiefer, M. Biasini, L. Bordoli, T. Schwede, Modeling protein quaternary structure of homo- and hetero-oligomers beyond binary interactions by homology, Sci. Rep. 7 (2017) 10480. https://doi.org/10.1038/s41598-017-09654-810.1038/s41598-017-09654-8 (pii).
[15] A. Lafita, S. Bliven, A. Kryshtafovych, M. Bertoni, B. Monastyrskyy, J.M. Duarte, T. Schwede, G. Capitani, Assessment of protein assembly prediction in CASP12, Proteins Struct. Funct. Bioinforma. 86 (2018) 247–256. https://doi.org/10.1002/prot.25408.
[16] D. Vasishtan, M. Topf, Scoring functions for cryoEM density fitting, J. Struct. Biol. 174 (2011) 333–343. https://doi.org/10.1016/j.jsb.2011.01.012S1047-8477(11)00025-6 (pii).
[17] I. Farabella, D. Vasishtan, A.P. Joseph, A.P. Pandurangan, H. Sahota, M. Topf, a Python library for assessment of three-dimensional electron microscopy density fits, J. Appl. Crystallogr. 48 (2015) 1314–1323. https://doi.org/10.1107/S1600576715010092vg5014 (pii).
[18] B.A. Barad, N. Echols, R.Y. Wang, Y. Cheng, F. DiMaio, P.D. Adams, J.S. Fraser, EMRinger: side chain-directed model and map validation for 3D cryo-electron microscopy, Nat. Methods 12 (2015) 943–946. https://doi.org/10.1038/nmeth.3541nmeth.3541 (pii).
[19] A. Kryshtafovych, B. Monastyrskyy, K. Fidelis, CASP prediction center infrastructure and evaluation measures in CASP10 and CASP ROLL, Proteins Struct. Funct. Bioinforma. 82 (2014) 7–13. https://doi.org/10.1002/prot.24399.