Estimating Residue-Specific Accuracies of Protein Structure Models Using Molecular Dynamics Simulations

Jing-Hao Hu,[1,2] Sang-Ni Xun,[1] Hao-Nan Wu,[1] Yun-Dong Wu,[1,2] Fan Jiang[1,*]

1. Laboratory of Computational Chemistry and Drug Design, State Key Laboratory of Chemical Oncogenomics, Peking University Shenzhen Graduate School, Shenzhen, 518055, China.

2. College of Chemistry and Molecular Engineering, Peking University, Beijing, 100871, China.

[*]Correspondence and requests for materials should be addressed.

ABSTRACT

Estimating the accuracy of a structure model is very crucial to promote the usefulness of protein structure prediction methods. Currently, a vast majority of successful model quality assessment (or model accuracy estimation, MAE) methods are knowledge-based. Based on molecular dynamics (MD) simulation with a recently developed residue-specific force field (RSFF2), we develop a method for absolute MAE at per-residue level. Using a training set of 31 models and a test set of 24 models from different proteins, the MAE performance of our MD-based method can reach or even exceed the state-of-the-art single-model MAE method within a short simulation time (less than one nanosecond). In addition, a simple combination of knowledge-based method with the MD-based method can obtain more accurate MAE than any of the constituent methods.

Introduction

Recently, the cost of obtaining new protein sequences has dramatically decreased due to significant progresses in high-throughput DNA sequencing technology.[1-2] On the other hand, to obtain high-resolution three-dimensional (3D) structures experimentally is still quite expensive, despite numerous effects. Therefore, there is a huge and increasing gap between the number of known structures and sequences.[3-5] Due to the importance of protein structures in modern biological research and drug discovery,[6-8] protein structure prediction (PSP) has become a cost-effective and high-throughput way to provide structure information for biological and pharmaceutical studies.[9-10]

There are various methods for PSP,[11-16] including homology modelling,[17] threading,[18] fragment assembly[19-20] and *ab initio* folding.[21-22] The predicted protein models can span a broad range of accuracies and are suitable for different applications.[7] They usually have lower qualities compared with experimental structures, significantly limiting their usefulness, especially for certain important applications including structure-based drug design.[23-26] Besides, a single PSP method can often generate a number of structure models, whereas it is very difficult to pick out the most accurate one(s).[27] The discrimination of accurate models from decoys has been a key factor to further increase the accuracy of PSP.[28] Thus, a reliable method of ranking models is urgently needed, which we called *relative* model accuracy estimation.[28-29]

Before using a model, it can be very beneficial to check whether it is accurate enough to avoid the risk of misuse. Thus, there is also a need for *absolute* model accuracy estimation (or simply called MAE) to estimate the discrepancy between a predicted model and the (unknown)

actual native structure, which can be performed on global or local level. The *global* MAE evaluates the overall quality of a model, while the *local* MAE estimates the error at residue-specific or atomic level. Basically, local MAE is more useful because that global score can often be inferred from summarizing local scores. Moreover, the local prediction can provide accuracy information of different regions, which can be used as a guide for further refinement.[28]

Some early MAE tools can evaluate a protein model by checking the correctness of stereochemistry, such as WHAT-CHECK,[30] PROCHECK,[31] and MolProbity.[32] Besides, many knowledge-based potentials derived from statistical analysis of features in known protein structures can also discriminate near-native structures from decoys (relative MAE).[33,34] Recently, there are developments of absolute MAE methods such as the ProQ series[28, 35-37] using supervised machine learning on multiple features including evolutionary information, residue environment compatibility or components from knowledge-based potentials. Besides these *single-model* methods, there are also consensus-based (clustering) methods, in which the quality prediction result of a model is correlated to its similarity to other models, such as Pcons.[11] Recently, hybrid methods that combine single-model and consensus-based approaches have also been developed.[16, 38]

The MAE was introduced into the Critical Assessment of Techniques for Protein Structure Prediction (CASP) experiment as an independent category since 2006, traditionally called model quality assessment (MQA).[39-44] As far as we know, a vast majority of successful MAE methods are knowledge-based.[45] However, with rapid increasing of computing power, the physics-based methods have shown more and more promising future. Especially, molecule

dynamics (MD) simulations, as a useful tool for studying protein dynamics, have been successfully applied in *ab initio* protein folding[21-22] and structure refinement.[46-49] The successful use of MD simulation in MAE has not been reported, but we think that it may work based on a physically sound assumption: native structure (free energy minimum) of a protein can be well kept during a MD simulation, but a non-native structure is less stable and will deviate from the initial structure.

Here, we evaluate the applicability of MD simulations in single-model absolute MAE at residue level. As an accurate force field is an essential requirement, our recently developed residue-specific force field RSFF2[50] was used, which is an improvement of AMBER ff99SB[51] by fitting conformational distributions from PDB coil library using residue-specific torsional parameters.[52-54] Previous studies have shown that RSFF2 can well reproduce conformational behaviors of peptides and proteins,[55] and predict the crystal structures of cyclic peptides accurately.[22] To investigate the applicability of MD simulations in MAE, we first performed MD simulations of a series of target models to find suitable patterns that are related to residue-wise model quality. Then, we compared our MD-based MAE method with other popular single-model methods and investigated whether a combination of MD-based and knowledge-based methods can achieve better performance. Finally, we further analyzed the MD trajectories to explain our findings and make connections to MD-based model refinement.

Material and Methods

Protein training and test datasets.

Thirty-one single-domain targets from the refinement categories of CASP 8-10, excluding those with missing residues in corresponding experimental structures, were chosen as the training datasets. Each model corresponds to a specific protein without redundancy, and the size of these models ranges from 63 to 235 residues. The chosen targets with diverse secondary structures have a large coverage range (1.3 - 7.7 Å) of Cα-RMSD to corresponding experimental structures. Meanwhile, twenty-four targets from the refinement categories of CASP 11 were chosen for the test datasets based on the same criteria above. All targets and experimental structures were downloaded from the CASP website.

MD simulations.

All simulations were carried out using the *Gromacs* 4.5.4 software[51, 55] with RSFF2 force field.[50] The hydrogen atoms of each target were added by Gromacs under neutral pH. Each protein was merged in an octahedron box with TIP3P water model.[56]. The distance between the protein and the box was 4.5 Å. Na+ or Cl- ions were added to neutralize the protein charges and to achieve an ion concentration of 0.05 mol/L. The long-range electrostatic interactions were calculated by PME method. The cutoffs of electrostatic and van der Waals interactions were at 9 Å. The integration step time was 3 fs. All bonds with hydrogen atoms were constrained using the LINCS. The Berendsen method was used to couple temperature and pressure (1 atm). For each target, after a 5000-step energy minimization, a 200 ps NPT simulation was performed with temperature increasing from 10 to 298 K within the first 40 ps, to equilibrium the solvent and box volume. During this process, all Cα atoms were restrained by 10 kJ/mol/Å^2 harmonic potentials. Then, a NVT productive simulation was employed. Parallel simulations were carried

out from different initial velocities. The structures were stored every 1.5 ps during the simulations.

Results

Simulated structural deviations correlate well with per-residue actual errors.

MD simulations of thirty-one protein models (training dataset) were performed. For each protein at each of the five temperatures (330K, 360K, 380K, 400K and 430K), five parallel MD simulations were carried out for 3 ns. As shown in Figure 1a, residue-wise RMSD (in time-average) values from MD simulations (defined in eq. 1) correlate with corresponding actual residue-wise errors (defined in eq. 2), with Pearson's correlation coefficient r = 0.64, which is slightly higher than that from a state-of-the-art MAE method ProQ3D[37] based on deep learning (Figure 1b).

$$rmsd_i^{\mathrm{MD}} = \sqrt{\frac{1}{T}\sum_{t=0}^{T}\left(x_{i,t} - x_{i,\mathrm{init}}\right)^2} \quad (1)$$

$$e_i^{\mathrm{actual}} = \left|x_{i,\mathrm{expt}} - x_{i,\mathrm{init}}\right| \quad (2)$$

To apply eq. 1, all structure frames in a MD trajectory was optimally superposed to the initial model. Here, $x_{i,\mathrm{init}}$ and $x_{i,t}$ are positions of Cα atom on i^{th} residue at simulation time t=0 (the initial model) and t=t, respectively. T is the trajectory length (number of frames) used for the calculation. In eq. 2, $x_{i,\mathrm{expt}}$ is the Cα position of ith residue in the experimental structure. The distance was also calculated after optimally superposition.

Interestingly, all different simulation temperatures (330K – 430K) give quite similar

residue-wise RMSD distributions (Figure 1c), which slightly larger and dispersed RMSD from higher temperature. As expected, longer simulation times make RMSD distributions shift to larger values (Figure 1d). MD simulations of 0.03 ns can give a RMSD distribution with peak position similar to the actual error distribution, but it is narrower. There is not a proper simulation time and temperature that can give RMSD distribution similar to the actual error distribution, indicating that certain kind of normalization is needed to make absolute MAE prediction from MD-derived RMSD values.

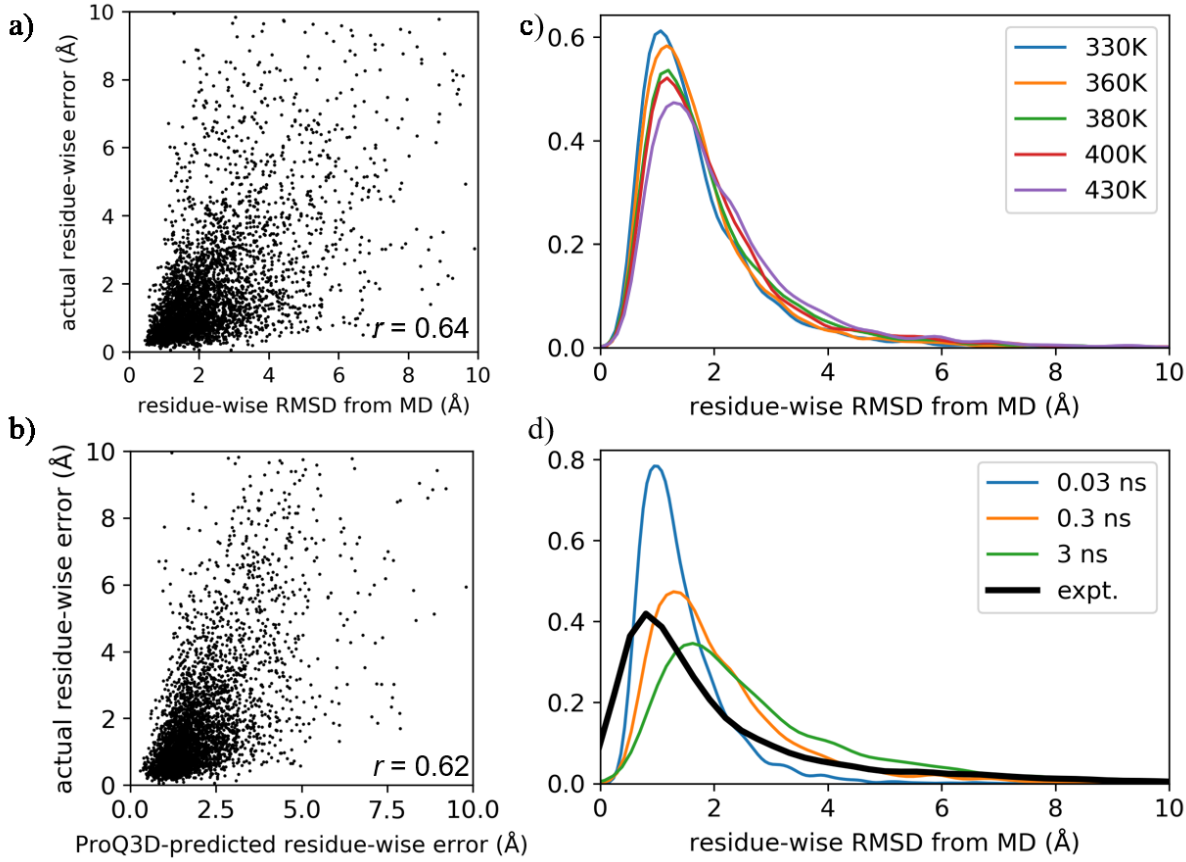

Figure 1. a) Actual residue-wise errors (defined in eq. 2) plotted against residue-wise RMSD (defined eq. 1) during 0.3 ns MD simulation (380K). Each dot presents a residue in the training set (31 proteins). b) Actual residue-wise error against the ProQ3D predictions of the training set. c) and d) Residue-wise RMSD distribution at different simulation temperatures (c) and

simulated times (d). The bold black line shows the distribution of actual residue-wise errors.

Convert distance into absolute local MAE predictor

Here, we use a simple linear transformation:

$$e_i^{\mathrm{MD}} = k * rmsd_i^{\mathrm{MD}} + b \quad (3)$$

The coefficients k and b were calculated based on the following statistics of simulated RMSD ($rmsd_i^{\mathrm{MD}}$, eq. 1) and actual errors (e_i^{actual}, eq. 2) in our training dataset:

$$\begin{cases} b = \min_i\left(e_i^{\mathrm{actual}}\right) - \min_i\left(rmsd_i^{\mathrm{MD}}\right) \\ k = \mathrm{median}_i\left(e_i^{\mathrm{actual}}\right) / \mathrm{median}_i\left(rmsd_i^{\mathrm{MD}}\right) \end{cases} \quad (4)$$

Obviously, k and b depend on the simulation time (Figure 2a). As shown in Figure 2b, the normalized residue-wise RMSD (e_i^{MD}) have similar distributions under different simulated time, and are in good agreement with the target distribution (e_i^{actual}).

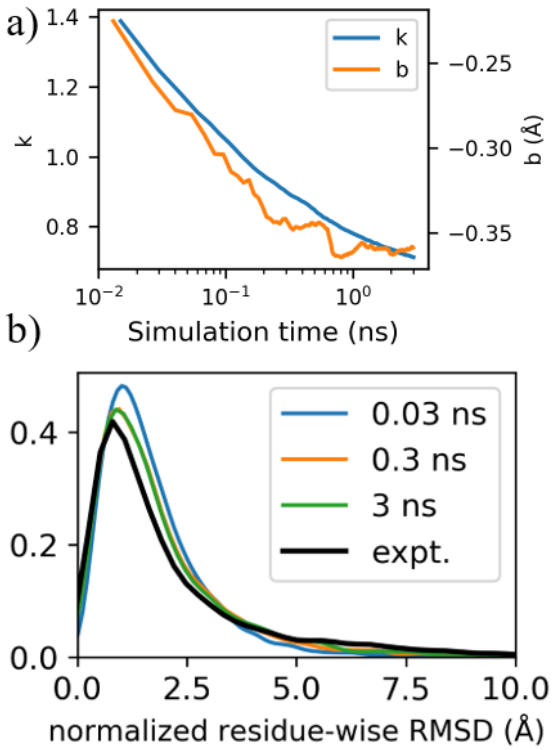

Figure 2. a) Normalization parameters k and b (see eq. 3) over simulation time (in logarithmic coordinates), from MD simulations at 380K. b) Distributions of the normalized residue-wise RMSD under different simulated time. The bold black line shows the actual residue-wise errors.

To better evaluate the performance of a MAE predictor, we converted the distance measures e_i^{actual} and e_i^{MD} to the s-score in the [0, 1] range:[57]

$$s_i = 1/[1 + \left(\frac{e_i}{d_0}\right)^2] \quad (5)$$

where e_i is the residue-wise position error in terms of distance, d_0 is set to 3 Å in consistent with MAE methods of ProQ series. Higher s-score indicates higher accuracy. Understandably, using s-score for MAE is better than direct use of distances. For those poor quality parts in a protein model, the RMSD can change in a fairly large range (e.g. 5 ~ 15 Å) without too much difference in accuracy, whereas a smaller change of RMSD from 1.5 Å to 3 Å can largely affect

applicability of a model. Figure 3 shows the distributions before and after distances were converted into s-scores, and clearly most residues have relatively high s-scores.

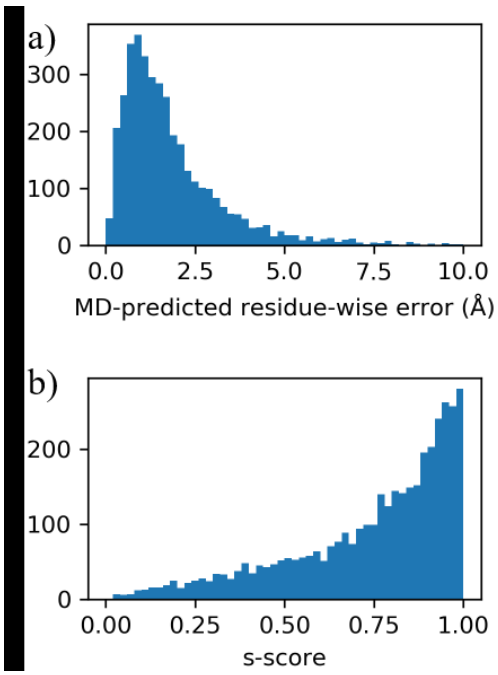

Figure 3. Distribution of MD-predicted residue-wise errors in histograms: a) before and b) after distances are converted into s-scores. Results from 380K MD simulations (0.3 ns) were shown.

A comparison with state-of-the-art MAE methods.

During the first 0.3 ns MD simulations of proteins in training set, the accuracy of residue-specific MAE predictions (as measured by the Pearson's correlation coefficient *r* between MD-predicted s-scores s_i^{MD} and real s-scores s_i^{actual}) increase quickly to be higher than those predicted by ProQ3 (Figure 4). After 0.5 ns of MD simulations, the results from the five parallel MD trajectories did not change significantly, which are better than ProQ3 and slighter worse than ProQ3D. However, when predictions from five individual trajectories were averaged, the performance better than ProQ3D can be achieved. Pearson's correlation coefficients (*r*) from

MD-based predictor reaches 0.66 within 0.3 ns and finally reaches 0.69 within 3 ns, whereas ProQ3D gives *r* of 0.64.

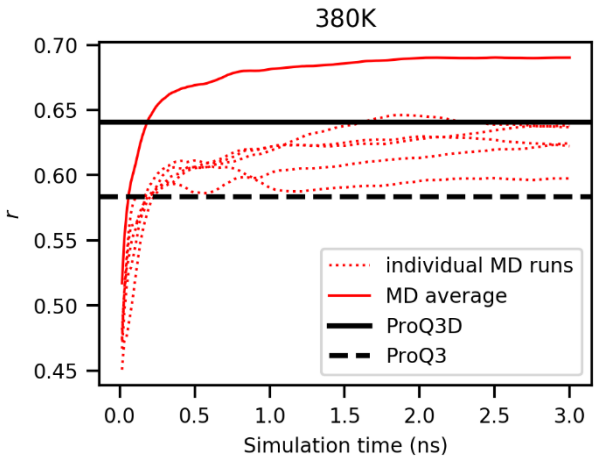

Figure 4. The correlation efficient (*r*) between MD predictions and actual errors (both in s-scores) in the training dataset over 3 ns simulation time at 380K. The dash red lines are the predictions from five individual MD runs, and the solid red line is the average of the five MD runs' predictions. The results of ProQ3 and ProQ3D are shown in the horizontal black dash line and black solid line, respectively.

For the local MAE (MQA) category in the CASP competition, the Accuracy of Self-Estimates (ASE) is the most popular criterion,[41, 58] which assesses the difference between the predicted and actual errors in a percentage scale:

$$\mathrm{ASE} = 100 \times \left(1 - \frac{1}{\mathrm{N}} \sum_{i=1}^{\mathrm{N}} \left| s_i^{\mathrm{pred}} - s_i^{\mathrm{actual}} \right| \right) \quad (6)$$

Where s_i^{pred} and s_i^{actual} are the s-score (see Eq. 5) of predicted and actual Cα position errors on i^{th} residue, respectively. The ASE score ranges from 0 to 100, with higher score indicating

more accurate MAE predictor.

As shown in Figure 5, the ASE scores increase rapidly at the beginning of the simulations, and the ASE scores of averaging trajectories (MD average) are significantly better than the results of all individual runs. When the temperature is lower than 380K, the higher the temperature is, the faster ASE scores converge during the simulation. But when the temperature is 380K or higher, the difference is not obvious. At all temperatures, both ASE scores from MD average and individual MD runs can exceed ProQ3 in a very short time, and the ASE scores of MD average can eventually exceed ProQ3D. Specifically, at 380K, the ASE scores of MD average can exceed ProQ3D within 0.2 ns and reach near convergence within 0.4 ns. These results indicate that the performance of our MD-based method can reach or even exceed the state-of-the-art single-model MAE method within a short simulation time.

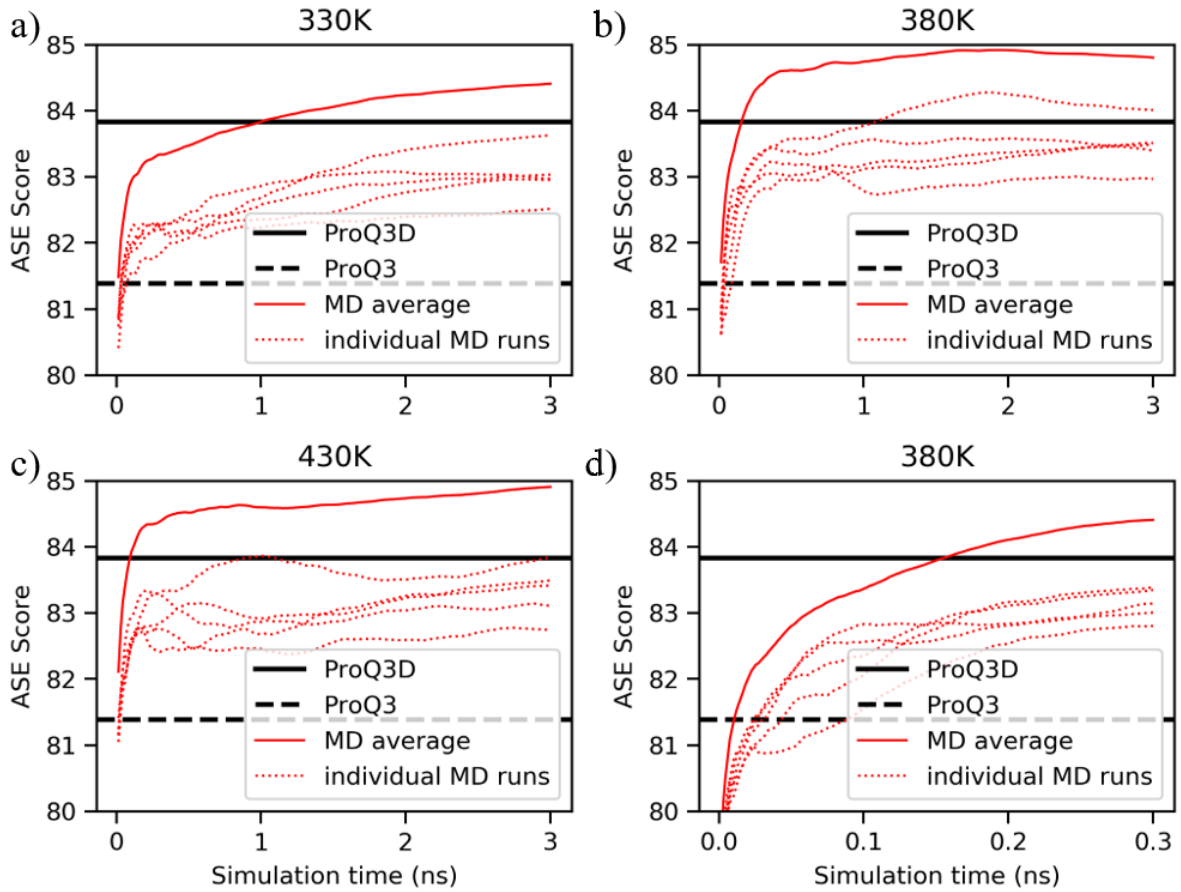

Figure 5. Average ASE scores (eq. 6) of MD predictions in the training dataset at a) 330K b) 380K c) 430K, over 3 ns simulation time. ASE scores of ProQ3 and ProQ3D were also shown for comparison. d) An enlarged view of the results from first 0.3 ns simulation in Figure 5b.

Combining MD-based MAE predictor with ProQ3D.

The ASE scores for each protein models are shown in Figure 6a. The ASE scores > 80 from MD average are in 27 out of 31 cases, and are > 90 in 6 cases. For ProQ3D, only one model scores more than 90. In 17 out of 31 cases, MD average can give better performance than ProQ3D. Only in two of the models (TR568, TR624), the scores of both methods are below 80. However, larger fluctuations of performances among different proteins can be observed for the MD-based

method, with standard deviation of 6.34 compared with 5.08 from ProQ3D. We also found that, the correlation coefficient between MD average and ProQ3D is 0.64, which is much smaller than the similarities among other methods.[41] Considering that our MD approaches are fundamentally different from previous MAE methods including ProQ3D, we speculated that there might be complementarity between them and, therefore, combined these two methods for more accurate estimation.

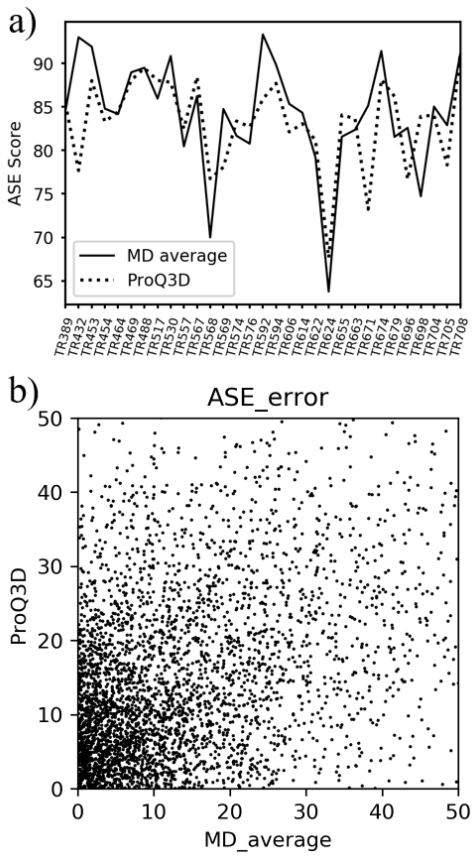

Figure 6. a) Average ASE scores of MD-based prediction (average of five parallel trajectories) and ProQ3D prediction for each protein model. b) Scatter plot of residue-wise ASE error (|predicted ASE score – actual ASE score|) of MD-based predictor and ProQ3D.

Consider the L1-norm penalty function in the ASE formula, we used least absolute

deviations (LAD) regression (eq. 5) to model the relationship between s-scores of ProQ3D and MD-based predicted error and s-score of actual error.

$$\hat{s}_i^{\text{comb}} = k_1 s_i^{\text{MD}} + k_2 s_i^{\text{ProQ3D}}$$

$$(k_1^*, k_2^*) = \underset{k_1,k_2}{\arg\min} \sum_{i=1}^{n} \left| s_i^{\text{actual}} - \hat{s}_i^{\text{comb}} \right| \quad (7)$$

Figure 7 shows the optimized parameters k_1 and k_2 as a function of simulation time t. When $t < 0.2$ ns, the contribution of MD-based method (k_1) increases with simulation time, but keeps less than ProQ3D. When $t > 0.2$ ns, the contribution of MD-based method becomes slightly larger than ProQ3D and keeps near invariant with increasing t.

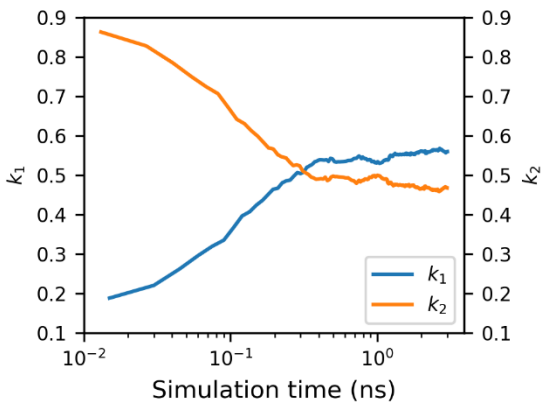

Figure 7. Combination coefficients k_1 and k_2 as a function of simulation time t (in logarithmic coordinates) at 380K.

As shown in Figure 8. The combined method performs significantly better than two parent methods. The combination with one individual run within 15 ps can even produce larger ASE scores than those from ProQ3D by 0.7, which almost reaches the best results of MD average. When combining ProQ3D with MD average, ASE score can reach up to 85.5, which increase the ASE score of ProQ3D and MD average by 1.7 and 0.9, respectively. This performance show

that a simple linear combination of MD-based method and ProQ3D can achieve better performance than any of the constitute methods. Even a very short time individual MD simulation can bring a big improvement to the combined methods.

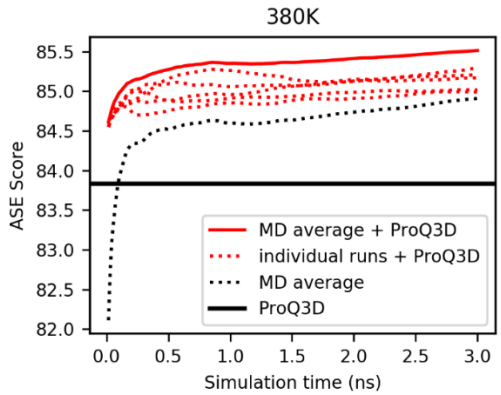

Figure 8. Average ASE scores of the combination (red lines) of MD-based method and ProQ3D in the training dataset compare with pure MD-based method (average, black dotted line) and ProQ3D (black solid line) over 3 ns simulation time at 380K.

Verification the MD-based methods on the test set.

Although the MD-based method is physics-based without using any optimizable parameter, there are two parameters (k and b) in eq. 3 based on statistics on a set of protein models (training set). For the combined method (MD + ProQ3D), there are two additional parameters (k_1, k_2) optimized on the training set. To validate the performance of the new methods on unseen data, 24 models of different proteins are chosen for testing. Based on the previous results, for each model, five parallel simulations were carried out at 380K for 0.3 ns. All the parameters are the same as above. On the test set, MD prediction performs no worse than the training set. The ASE scores of MD-based method can exceed ProQ3D within 0.03 ns and reach 86 within 0.3 ns

(Figure 9). In 17 out of 24 cases, MD average can give better performance than ProQ3D, and the ASE score of the combined methods reach 86.7, which exceed over the results from ProQ3D by 2.4. These results demonstrate that our method has good transferability on different proteins.

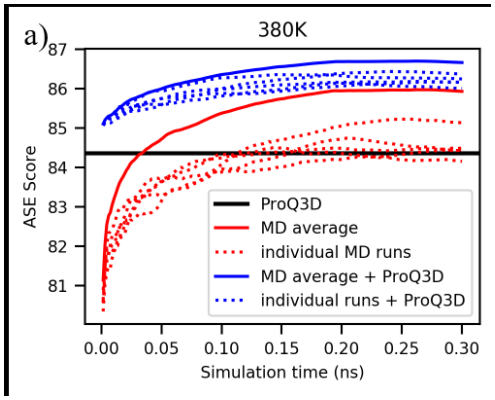

Figure 9. Average ASE scores of ProQ3D, MD-based predictor and the combination method from the test dataset. MD simulations were carried out at 380K for 0.3 ns.

Analysis on the MD trajectories

As shown in Figure 10a, we decomposed the movement of each residue in a MD trajectory into two directions, using a local (residue-specific) coordinates system after globally optimal superposition of the simulated model and the corresponding experimental structure. The direction from the residue's position in the initial model to that in the experimental structure is defined as "refinement axis" (X axis), because this should be the moving direction of an ideal model refinement. In addition to the refinement axis, residues are also moving in a direction (Y-axis) perpendicular to the refinement axis. As shown in Figure 10b, most of the residues are located on the right half of the X-axis, indicating that most of the residues move toward the refined direction, but this tendency is not strong. Interestingly, when the initial model is further

away from the experimental structure, the motions on both the X and Y axes tend to be larger (Figure 10c). Therefore, we can still estimate the quality of the initial model from the trajectory without any experimental structure. Noticeably, when residues in the model are very far from the experimental ones, there is *not* a clear tendency for moving in the direction of better model quality (refinement). Thus, using short MD simulations, we can estimate the model quality even if good refinement is not achieved. In this work, we assumed that the relationship between residue-wise actual errors and dynamic motions in MD simulation is linear. However, as we can see in Figure 10d, this relationship is actually weaker than linear. For modelled residues with larger position errors, their movements in MD (both X and Y directions) related to the actual errors ($|\boldsymbol{v_1}|$) are significantly smaller than residues with smaller modelling errors.

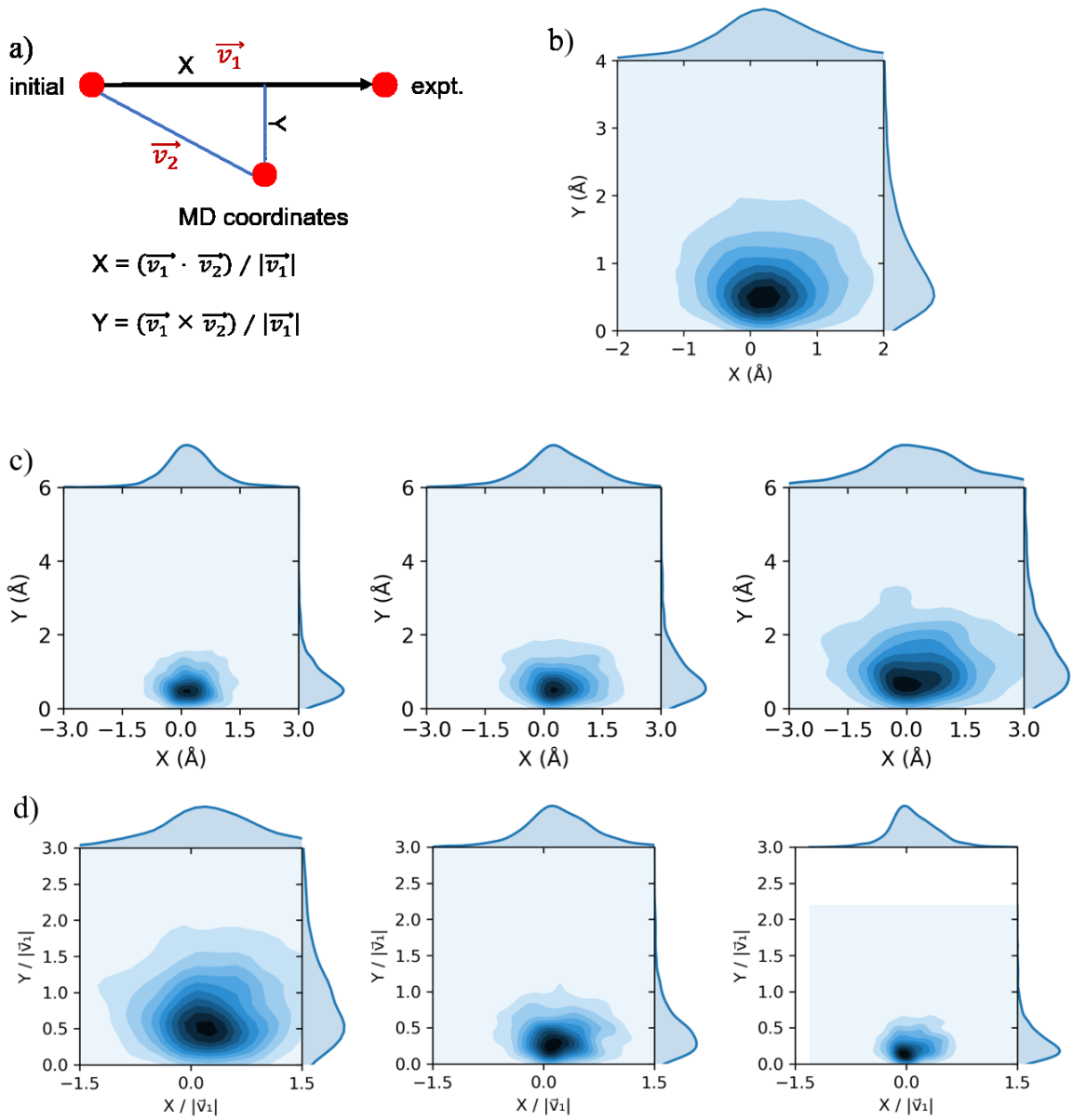

Figure 10. a) Local coordinates decomposition of residue-wise movements during MD simulations into X (refinement) axis and Y axis. b) X-Y distribution of all residues in the training dataset. c) X-Y distributions of residues in different ranges of actual residue-wise errors: small (0, 1.39 Å], medium (1.39 Å, 2.65 Å], and large (2.65 Å, 29.9 Å] from left to right. Each group has the same number of residues. d) X-Y distributions in (c) normalized by actual residue-wise errors. All MD simulations were carried out at 380K for 0.3 ns.

Discussion

In principle, the relationship between the sequence of a protein and its native structure is governed by physical chemistry.[59] Native structure of a protein corresponds to the global minimum of its free energy. Indeed, experimental structures of some small proteins can be predicted by simulating their spontaneous folding using molecular dynamics (MD).[60-61] However, the most successful protein structure prediction methods are usually based on data mining of known protein sequences and structures.[3, 62-64] These knowledge-based methods are much more efficient than physics-based simulations. However, they are less physically realistic, especially neglecting the solvent molecules which can play very important roles.

Recently, with rapid developments of computational hardware and software, MD simulation has been attracting increasing applications in studying conformational behaviors of peptides and proteins. This has been motivating continuous improvements of protein force fields, along with increasing predicting power. Feig and coworkers showed that restrained MD simulations (a few hundred nanoseconds or less) in explicit water can result in improvements of structure models.[46-47] We have shown that long-time replica-exchange MD simulations (microseconds per replica) sometimes can result in significant model improvements.[48] In recent CASP competitions, most top-ranking groups in the refinement category use certain type of MD simulation.[49] However, consistent improvement of structure models can only be achieved by applying restraints to keep structures around the initial model. This conservative strategy limits structural improvements that can be achieved. On the other hand, more extensive sampling, as used by our group,[22, 48] can give more significantly improved structures sometime,

but may also lead to much worse structures.[48] It is still unknown the problem is due to force field inaccuracy or insufficient simulation time (conformational sampling). If we can put more restraints on the part in a model that already close to native structure while let less accurate part more fully to move, more efficient conformational sampling and better refinement may be achieved.

In this work, we found that MD simulations can also be used to estimate model accuracy at residue level, with much less computational cost compared with structure refinement. During MD simulation, significant deviations from the initial position can be observed in very short time (tens of picoseconds), and residues with larger deviations from experimental structures indeed tend to move further away from their initial positions. We also found that averaging parallel trajectories can achieve better predictions, this can be understood because a MD trajectory is a stochastic search in the conformational space. Very encouragingly, further increasing the number of trajectories in our MD-based method may achieve better performance and more stable predictions.

Even the performance of current MD-based MAE method is not much better than top single-model MAE methods, developing a new type of methods is also important. Sometimes, a consensus-based prediction method can achieve better performance than any of the constituent methods,[39, 42-43, 45, 65] similar to the concept of ensemble learning, especially when considering the fundamental difference between our MD approach and those knowledge-based MAE methods.

Our previous study has shown that long-term MD simulations (~150ns) with enhanced

conformational sampling are required to refine protein structures. We analyzed the MD trajectories to see if the model trajectories are moving toward the direction of refinement. Many residues indeed move in the direction of refinement, but most of them are not closer than the initial structure to the experimental structure. Thus, it is difficult to use these trajectories to optimize the protein structures. This may be because these residues may take a more complicated transition path and time scale much longer than a few nanoseconds to be refined. Therefore, they do not move in the right direction initially. However, for MAE, it seems that a series of short MD simulations can achieve good results.

In the recent CASP experiments, best consensus-based methods and hybrid methods perform better than best single-model methods for both global and local MAE competitions. Still, CASP has been continuously emphasizing the importance of single-model methods.[41] First, a large number of predicted models which are required by consensus-based methods are not easily available in every-day practice, thus single-model methods are more useful. Moreover, single-model methods appear to be better than consensus-based methods in picking the best model from decoy sets, according to the results of recent CASP12.[41] Finally, single-model methods can be combined with consensus-based methods to achieve a better performance.[65] Therefore, the improvement of the single-model method still play a decisive role in current development of MAE method.[29, 36] We limit the scope of this work to "a proof of concept" study. However, in the future, more sophisticated methods will be developed, including those utilize the machine learning to combine various structural features from MD simulation, as well as traditional features for model quality assessment.

AUTHOR INFORMATION

Corresponding Author

*E-mail: jiangfan@pku.edu.cn (F.J.)

Notes

The authors declare no competing financial interest.

ACKNOWLEDGEMENTS

We thank the financial supports from the Shenzhen Science and Technology Innovation Committee (JCYJ20170412150507046), and the National Natural Science Foundation of China (21573009).

REFERENCES

1. Mardis, E. R., DNA sequencing technologies: 2006–2016. *Nature Protocols* **2017,** *12*, 213.
2. Shendure, J.; Balasubramanian, S.; Church, G. M.; Gilbert, W.; Rogers, J.; Schloss, J. A.; Waterston, R. H., DNA sequencing at 40: past, present and future. *Nature* **2017,** *550*, 345.
3. Roy, A.; Kucukural, A.; Zhang, Y., I-TASSER: a unified platform for automated protein structure and function prediction. *Nature protocols* **2010,** *5* (4), 725.
4. Bradley, P.; Misura, K. M.; Baker, D., Toward high-resolution de novo structure prediction for small proteins. *Science* **2005,** *309* (5742), 1868-1871.
5. Šali, A.; Blundell, T. L., Comparative protein modelling by satisfaction of spatial restraints. *Journal of molecular biology* **1993,** *234* (3), 779-815.
6. Laskowski, R. A.; Watson, J. D.; Thornton, J. M., ProFunc: a server for predicting protein function from 3D structure. *Nucleic acids research* **2005,** *33* (suppl_2), W89-W93.
7. Zhang, Y., Protein structure prediction: when is it useful? *Current opinion in structural biology* **2009,** *19* (2), 145-155.
8. Ekins, S.; Mestres, J.; Testa, B., In silico pharmacology for drug discovery: methods for virtual ligand screening and profiling. *British journal of pharmacology* **2007,** *152* (1), 9-20.
9. Zhang, Y., Progress and challenges in protein structure prediction. *Current opinion in*

structural biology **2008,** *18* (3), 342-348.

10. Li, J.; Cao, R.; Cheng, J., A large-scale conformation sampling and evaluation server for protein tertiary structure prediction and its assessment in CASP11. *BMC bioinformatics* **2015,** *16* (1), 337.

11. Lundström, J.; Rychlewski, L.; Bujnicki, J.; Elofsson, A., Pcons: A neural-network–based consensus predictor that improves fold recognition. *Protein Science* **2001,** *10* (11), 2354-2362.

12. Li, J.; Adhikari, B.; Cheng, J., An improved integration of template-based and template-free protein structure modeling methods and its assessment in CASP11. *Protein and peptide letters* **2015,** *22* (7), 586-593.

13. Li, J.; Bhattacharya, D.; Cao, R.; Adhikari, B.; Deng, X.; Eickholt, J.; Cheng, J., The MULTICOM protein tertiary structure prediction system. In *Protein Structure Prediction*, Springer: 2014; pp 29-41.

14. Jones, D. T., GenTHREADER: an efficient and reliable protein fold recognition method for genomic sequences1. *Journal of molecular biology* **1999,** *287* (4), 797-815.

15. Zhou, H.; Zhou, Y., Fold recognition by combining sequence profiles derived from evolution and from depth-dependent structural alignment of fragments. *Proteins: Structure, Function, and Bioinformatics* **2005,** *58* (2), 321-328.

16. McGuffin, L. J., The ModFOLD server for the quality assessment of protein structural models. *Bioinformatics* **2008,** *24* (4), 586-587.

17. Siezen, R. J.; de Vos, W. M.; Leunissen, J. A.; Dijkstra, B. W., Homology modelling and protein engineering strategy of subtilases, the family of subtilisin-like serine proteinases. *Protein Engineering, Design and Selection* **1991,** *4* (7), 719-737.

18. Lemer, C. M. R.; Rooman, M. J.; Wodak, S. J., Protein structure prediction by threading methods: evaluation of current techniques. *Proteins: Structure, Function, and Bioinformatics* **1995,** *23* (3), 337-355.

19. Myers, E. W., Toward simplifying and accurately formulating fragment assembly. *Journal of Computational Biology* **1995,** *2* (2), 275-290.

20. Pevzner, P. A.; Tang, H.; Waterman, M. S., An Eulerian path approach to DNA fragment assembly. *Proceedings of the National Academy of Sciences* **2001,** *98* (17), 9748-9753.

21. Jiang, F.; Wu, Y.-D., Folding of fourteen small proteins with a residue-specific force field and replica-exchange molecular dynamics. *Journal of the American Chemical Society* **2014,** *136* (27), 9536-9539.

22. Geng, H.; Jiang, F.; Wu, Y.-D., Accurate structure prediction and conformational analysis of cyclic peptides with residue-specific force fields. *The journal of physical chemistry letters* **2016,** *7* (10), 1805-1810.

23. Giorgetti, A.; Raimondo, D.; Miele, A. E.; Tramontano, A., Evaluating the usefulness of protein structure models for molecular replacement. *Bioinformatics* **2005,** *21* (suppl_2), ii72-ii76.

24. Brylinski, M.; Skolnick, J., A threading-based method (FINDSITE) for ligand-binding site prediction and functional annotation. *Proceedings of the National Academy of sciences* **2008,** *105* (1), 129-134.

25. Brylinski, M.; Skolnick, J., Q-Dock: Low-resolution flexible ligand docking with

pocket-specific threading restraints. *Journal of computational chemistry* **2008,** *29* (10), 1574-1588.

26. Becker, O. M.; Dhanoa, D. S.; Marantz, Y.; Chen, D.; Shacham, S.; Cheruku, S.; Heifetz, A.; Mohanty, P.; Fichman, M.; Sharadendu, A., An integrated in silico 3D model-driven discovery of a novel, potent, and selective amidosulfonamide 5-HT1A agonist (PRX-00023) for the treatment of anxiety and depression. *Journal of medicinal chemistry* **2006,** *49* (11), 3116-3135.

27. Zhang, J.; Wang, Q.; Vantasin, K.; Zhang, J.; He, Z.; Kosztin, I.; Shang, Y.; Xu, D., A multilayer evaluation approach for protein structure prediction and model quality assessment. *Proteins: Structure, Function, and Bioinformatics* **2011,** *79* (S10), 172-184.

28. Ray, A.; Lindahl, E.; Wallner, B., Improved model quality assessment using ProQ2. *BMC bioinformatics* **2012,** *13* (1), 224.

29. Liu, T.; Wang, Y.; Eickholt, J.; Wang, Z., Benchmarking deep networks for predicting residue-specific quality of individual protein models in CASP11. *Scientific reports* **2016,** *6*, 19301.

30. Hooft, R. W.; Vriend, G.; Sander, C.; Abola, E. E., Errors in protein structures. *Nature* **1996,** *381* (6580), 272.

31. Laskowski, R. A.; MacArthur, M. W.; Moss, D. S.; Thornton, J. M., PROCHECK: a program to check the stereochemical quality of protein structures. *Journal of applied crystallography* **1993,** *26* (2), 283-291.

32. Chen, V. B.; Arendall, W. B.; Headd, J. J.; Keedy, D. A.; Immormino, R. M.; Kapral, G. J.; Murray, L. W.; Richardson, J. S.; Richardson, D. C., MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallographica Section D: Biological Crystallography* **2010,** *66* (1), 12-21.

33. Zhang, C.; Liu, S.; Zhou, Y., Accurate and efficient loop selections by the DFIRE-based all-atom statistical potential. *Protein science* **2004,** *13* (2), 391-399.

34. Rohl, C. A.; Strauss, C. E.; Misura, K. M.; Baker, D., Protein structure prediction using Rosetta. In *Methods in enzymology*, Elsevier: 2004; Vol. 383, pp 66-93.

35. Uziela, K.; Shu, N.; Wallner, B.; Elofsson, A., ProQ3: Improved model quality assessments using Rosetta energy terms. *Scientific reports* **2016,** *6*, 33509.

36. Wallner, B.; Elofsson, A., Prediction of global and local model quality in CASP7 using Pcons and ProQ. *Proteins: Structure, Function, and Bioinformatics* **2007,** *69* (S8), 184-193.

37. Uziela, K.; Menéndez Hurtado, D.; Shu, N.; Wallner, B.; Elofsson, A., ProQ3D: improved model quality assessments using deep learning. *Bioinformatics* **2017,** *33* (10), 1578-1580.

38. Maghrabi, A. H.; McGuffin, L. J., ModFOLD6: an accurate web server for the global and local quality estimation of 3D protein models. *Nucleic acids research* **2017,** *45* (W1), W416-W421.

39. Cozzetto, D.; Kryshtafovych, A.; Tramontano, A., Evaluation of CASP8 model quality predictions. *Proteins: Structure, Function, and Bioinformatics* **2009,** *77* (S9), 157-166.

40. Cozzetto, D.; Kryshtafovych, A.; Ceriani, M.; Tramontano, A., Assessment of predictions in the model quality assessment category. *Proteins: Structure, Function, and Bioinformatics* **2007,** *69* (S8), 175-183.

41. Kryshtafovych, A.; Monastyrskyy, B.; Fidelis, K.; Schwede, T.; Tramontano, A., Assessment of model accuracy estimations in CASP12. *Proteins: Structure, Function, and Bioinformatics* **2018,** *86*, 345-360.

42. Kryshtafovych, A.; Fidelis, K.; Tramontano, A., Evaluation of model quality predictions in CASP9. *Proteins: Structure, Function, and Bioinformatics* **2011,** *79* (S10), 91-106.

43. Kryshtafovych, A.; Barbato, A.; Fidelis, K.; Monastyrskyy, B.; Schwede, T.; Tramontano, A., Assessment of the assessment: evaluation of the model quality estimates in CASP10. *Proteins: Structure, Function, and Bioinformatics* **2014,** *82* (S2), 112-126.

44. Kryshtafovych, A.; Barbato, A.; Monastyrskyy, B.; Fidelis, K.; Schwede, T.; Tramontano, A., Methods of model accuracy estimation can help selecting the best models from decoy sets: assessment of model accuracy estimations in CASP11. *Proteins: Structure, Function, and Bioinformatics* **2016,** *84* (S1), 349-369.

45. Elofsson, A.; Joo, K.; Keasar, C.; Lee, J.; Maghrabi, A. H.; Manavalan, B.; McGuffin, L. J.; Ménendez Hurtado, D.; Mirabello, C.; Pilstål, R., Methods for estimation of model accuracy in CASP12. *Proteins: Structure, Function, and Bioinformatics* **2018,** *86*, 361-373.

46. Mirjalili, V.; Feig, M., Protein structure refinement through structure selection and averaging from molecular dynamics ensembles. *Journal of chemical theory and computation* **2013,** *9* (2), 1294-1303.

47. Mirjalili, V.; Noyes, K.; Feig, M., Physics-based protein structure refinement through multiple molecular dynamics trajectories and structure averaging. *Proteins: Structure, Function, and Bioinformatics* **2014,** *82*, 196-207.

48. Xun, S.; Jiang, F.; Wu, Y.-D., Significant refinement of protein structure models using a residue-specific force field. *Journal of chemical theory and computation* **2015,** *11* (4), 1949-1956.

49. Hovan, L.; Oleinikovas, V.; Yalinca, H.; Kryshtafovych, A.; Saladino, G.; Gervasio, F. L., Assessment of the model refinement category in CASP12. *Proteins: Structure, Function, and Bioinformatics* **2018,** *86*, 152-167.

50. Zhou, C.-Y.; Jiang, F.; Wu, Y.-D., Residue-specific force field based on protein coil library. RSFF2: modification of AMBER ff99SB. *The journal of physical chemistry B* **2014,** *119* (3), 1035-1047.

51. Showalter, S. A.; Brüschweiler, R., Validation of molecular dynamics simulations of biomolecules using NMR spin relaxation as benchmarks: application to the AMBER99SB force field. *Journal of chemical theory and computation* **2007,** *3* (3), 961-975.

52. Jiang, F.; Han, W.; Wu, Y.-D., Influence of side chain conformations on local conformational features of amino acids and implication for force field development. *The Journal of Physical Chemistry B* **2010,** *114* (17), 5840-5850.

53. Adhikari, A. N.; Freed, K. F.; Sosnick, T. R., De novo prediction of protein folding pathways and structure using the principle of sequential stabilization. *Proceedings of the National Academy of Sciences* **2012**.

54. Jiang, F.; Han, W.; Wu, Y.-D., The intrinsic conformational features of amino acids from a protein coil library and their applications in force field development. *Physical Chemistry Chemical Physics* **2013,** *15* (10), 3413-3428.

55. Li, S.; Elcock, A. H., Residue-Specific Force Field (RSFF2) Improves the Modeling of Conformational Behavior of Peptides and Proteins. *The journal of physical chemistry letters* **2015,** *6* (11), 2127-2133.

56. Darden, T.; Perera, L.; Li, L.; Pedersen, L., New tricks for modelers from the crystallography toolkit: the particle mesh Ewald algorithm and its use in nucleic acid simulations. *Structure* **1999,** *7* (3), R55-R60.

57. Gerstein, M.; Levitt, M. In *Using iterative dynamic programming to obtain accurate pairwise and multiple alignments of protein structures*, Ismb, 1996; pp 59-66.

58. Kryshtafovych, A.; Monastyrskyy, B.; Fidelis, K., CASP 11 statistics and the prediction center evaluation system. *Proteins: Structure, Function, and Bioinformatics* **2016,** *84*, 15-19.

59. Levinthal, C., Are there pathways for protein folding? *Journal de chimie physique* **1968,** *65*, 44-45.

60. Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Shaw, D. E., How fast-folding proteins fold. *Science* **2011,** *334* (6055), 517-520.

61. Shaw, D. E.; Maragakis, P.; Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Eastwood, M. P.; Bank, J. A.; Jumper, J. M.; Salmon, J. K.; Shan, Y., Atomic-level characterization of the structural dynamics of proteins. *Science* **2010,** *330* (6002), 341-346.

62. Ovchinnikov, S.; Park, H.; Varghese, N.; Huang, P.-S.; Pavlopoulos, G. A.; Kim, D. E.; Kamisetty, H.; Kyrpides, N. C.; Baker, D., Protein structure determination using metagenome sequence data. *Science* **2017,** *355* (6322), 294-298.

63. Kortemme, T.; Morozov, A. V.; Baker, D., An orientation-dependent hydrogen bonding potential improves prediction of specificity and structure for proteins and protein–protein complexes. *Journal of molecular biology* **2003,** *326* (4), 1239-1259.

64. John, B.; Sali, A., Comparative protein structure modeling by iterative alignment, model building and model assessment. *Nucleic acids research* **2003,** *31* (14), 3982-3992.

65. Larsson, P.; Skwark, M. J.; Wallner, B.; Elofsson, A., Assessment of global and local model quality in CASP8 using Pcons and ProQ. *Proteins: Structure, Function, and Bioinformatics* **2009,** *77* (S9), 167-172.