Extracting representative structures from protein conformational ensembles

Alberto Perez, Arijit Roy, Koushik Kasavajhala, Amy Wagaman, Ken A. Dill, and Justin L. MacCallum

ABSTRACT

A large number of methods generate conformational ensembles of biomolecules. Often one structure is selected to be representative of the whole ensemble, usually by clustering and selecting the structure closest to the center of the most populated cluster. We find that this structure is not necessarily the best representation of the cluster and present here two computationally inexpensive averaging protocols that can systematically provide better representations of the system, which can be more directly compared with structures from X-ray crystallography. In practice, systematic errors in the generated conformational ensembles appear to limit the maximum improvement of averaging methods.

INTRODUCTION

Many methods generate distributions of protein conformations. Examples include molecular dynamics simulations, structure determination by refinement of nuclear magnetic resonance (NMR) data, and computational tools for protein structure prediction. We often want to compare these distributions with the ensemble average structure from X-ray crystallography. The question we address in this article is: How do we choose or generate a structure that is representative of the conformational ensemble, which can be directly compared with the average structure from X-ray crystallography?

We have previously explored the use of elastic network models to compare protein structures taking into account the flexibility of different regions of the protein. In this work, we look at a related problem: selecting one structure to represent an ensemble of conformations. We look at ways to average conformational flexibility while maintaining the correct physical nature of the structure.
the most likely position for each atom is the mean position and the most likely single structure is the mean structure. Why does this happen? The force at work is conformational entropy. There are simply far more conformations with high root mean square deviation (RMSD) than with low RMSD and thermal motions are constantly pushing atoms away from their average positions. Even though each atom is most likely to be at its mean position, the probability of having all atoms simultaneously close to their mean position is negligible.

This example illustrates the difficulty of comparing any structure generated from an ensemble with the average structure from X-ray crystallography. The Gaussian fluctuations in the crystal structure model translate into a quadratic energy function restraining each atom to its mean position. In this case, we have a perfect energy function and can generate an almost arbitrarily large sample of conformations, but the chance of generating a structure very close to the average is vanishingly small. This effect becomes more pronounced the more flexible the protein is and the broader the conformational ensemble becomes, because the average RMSD is directly related to the size of the atomic fluctuations (see caption of Fig. 1).

There is, however, enough information in the ensemble shown in Figure 1 to accurately estimate the average structure. However, nonlinearities and correlations between atoms (described in the next two sections) can complicate matters and lead to CartAvg to produce suboptimal estimates, which motivated us to develop an alternative graph theoretic method called NetAvg (see Supporting Information and Methods section).

In this article, we compare CartAvg and NetAvg with the common practice of choosing the structure closest to the center of the ensemble, called the medoid. We examine two types of datasets: (1) cases with ideal distributions, which have been constructed such that the ensemble is guaranteed to fluctuate around the experimental structure and (2) real-world distributions where the energy function is not perfect and thus the ensemble might not be centered around the experimental structure.

**Some atoms have nonlinear distributions**

In the previous example, the distribution of each atom was an isotropic Gaussian centered on the mean position. However, not all atomic distributions are so simple. Some atoms have more complex nonlinear distributions—particularly atoms in surface sidechains and in loops. For example, in a surface lysine residue, the length between the α-carbon and the terminal nitrogen is relatively constant, but the angular degrees of freedom are much more flexible. This leads to a dome-shaped distribution of the nitrogen atom [Fig. 2(A)]. Taking the mean of such a distribution produces an average position that does not lie within the cloud of positions that were actually sampled [Fig. 2(B)]. The mean position is artificially “compressed” toward the point of rotation.

A better “average” position—which is not really an average, but rather a “representative” position—would be one that lies near the most densely populated region of the point cloud. We have developed a graph theoretic method called NetAvg (described in Methods section and Supporting Information) that attempts to capture better representative structures from such nonlinear distributions. We defer description of the algorithm of NetAvg until the Methods section.

**Ignoring correlations between atoms leads to unphysical structures**

With NetAvg and CartAvg, we consider the position of one atom at a time and try to “project” toward a more central structure that is not in the ensemble. When determining the average position with either method, one of the major approximations we make is to ignore correlations between atom positions. We assume that the average position of each atom can be determined independently, ignoring the position of other atoms.

This assumption often turns out to be a bad approximation. Physical forces like bonds and steric repulsion...
lead to strong correlations between the positions of atoms that are located close together in space. By ignoring these correlations when computing the average, we often produce models that appear to violate these physical forces and have unrealistic bond geometry and steric clashes. For example, the rigid chemical structure of a phenylalanine ring introduces coupling between atoms on opposite sides of the ring. The averaging methods ignore these correlations and can lead to unphysical structures (Fig. 3, unphysical average).

For NetAvg and CartAvg, we do not deal with this problem of correlations producing unphysical structures directly. Instead, we compute the raw average structure ignoring correlations and then use a targeted energy minimization procedure (see Methods section) to produce a physically plausible structure that is close to the raw average structure (Fig. 3, physical minimized average).

METHODS

Scripts for performing the following calculations are available with an open source license (http://www.github.com/laufercenter/NetAvg).

Iterative superposition

Translation and rotation of the system is independent of the individual conformation and thus should be removed. We have superposed all the conformations to a reference frame (mean conformation in this work). We use an iterative procedure of superposing and recalculating the mean conformation to ensure the optimum reference frame.

Medoid: Representing the ensemble with the member closest to the Cartesian average

The medoid structure is the structure from the ensemble that is closest to the mean structure. The medoid is commonly used as a representative structure from a
We compute the medoid structure by first finding the average structure of the cluster by iterative superposition (described above) and then selecting the member of the ensemble with the smallest RMSD to the average.

**CartAvg: Representing the ensemble with a physically plausible estimate based on the Cartesian average**

The CartAvg structure is obtained by first iteratively superposing all structures from the ensemble onto the average structure until convergence (described above). Next, the new average structure is computed by considering each atom independently. The \( x \), \( y \), and \( z \)-coordinates are simply the mean \( x \), \( y \), and \( z \)-coordinates of that atom in the ensemble. Thus, CartAvg not only ignores correlations between atoms, it also ignores correlations between the \( x \), \( y \), and \( z \)-coordinates for each atom. The coordinates of the mean structure are then used as a target for energy minimization (described below) to produce a physically meaningful structure that is close to the Cartesian average.

**NetAvg: Representing the ensemble using a physically plausible estimate based on a nonlinear network-based average**

Physical forces acting on each atom can introduce correlations not only between atoms but also between the \( x \), \( y \), and \( z \)-coordinates of each atom. As shown in Figure 2, if an atom has a nonlinear distribution, the mean position will often lie in a region of space that is never sampled during the simulation. The NetAvg attempts to correct this by using a nonlinear graph-based approach that tries to choose an average position that lies within the cloud of positions sampled by each atom.

NetAvg still ignores correlations between atoms. However, unlike CartAvg, it explicitly includes correlations between the \( x \), \( y \), and \( z \)-coordinates of each atom, which allows it to capture non-linearities in the atomic distributions. We defer the details to the Supporting Information, but the main features of the algorithm are described below.

Each atom is considered individually. The point cloud for each atom is turned into a graph, where points that are close together are connected by edges. Next, the atom with the highest eigenvector centrality is identified, and the representative position for the atom is computed as the mean of the position with the highest eigenvector centrality and all other positions that are directly connected to it.

The eigenvector centrality is a measure of how important a node is and is related to Google’s PageRank algorithm. In the present application, NetAvg, eigenvector centrality is largely sensitive to density and tends to pick points in regions of space that are densely sampled. If there is no clear region that is more densely sampled, the eigenvector centrality will tend to choose points that are near the topological center of the point cloud. As with CartAvg, the structure obtained using this procedure is then used as a target during energy minimization (next section).

**Targeted Energy Minimization: Obtaining physically reasonable representative structures**

To produce physically reasonable average models using NetAvg or CartAvg, we combine these averaging methods with molecular dynamics-based energy minimization. Instead of reporting the average structure directly, we first select the structure from the ensemble that is closest to the average from NetAvg or CartAvg. Next, we perform an energy minimization to pull the selected structure towards the average structure in the presence of a physical force field, as described below.

All minimizations were carried out using the GROMACS software package, version 4.5.5. Proteins were modeled using the FF99SB-ILDN force field. Water was modeled implicitly using the generalized Born model of Onufriev, Bashford, and Case. The hydrophobic effect was approximated using the atomic contact energies (ACE) approximation. The selected model from the ensemble was pulled toward the average model using Cartesian position restraints on all heavy atoms with a force constant of 1000 kJ/mol/nm\(^2\). The energy was minimized with 1000 steps of steepest descent, followed by 1000 steps of l-BFGS.

**Assessing Success**

We use four measures to determine if the averaging protocols have been successful: heavy-atom RMSD, GDT-HA, CAD-score, and MolProbity (all defined below). These measures have been typically used in works of structure prediction such as CASP and represent a balance between global and local geometry of the backbone, as well as ensuring correct physical contacts and absence of steric clashes.

The heavy-atom RMSD measures the average deviation between the positions of the non-hydrogen atoms in the model with those in the experimental structure after rigid body superposition to minimize the deviation. Some residues contain chemically identical atoms related by internal rotations—Phe: CD1-CD2 and CE1-CE2; Tyr: CD1-CD2 and CE1-CE2; Asp: OD1-OD2; Glu: OE1-OE2; Leu: CD1-CD2; Val: CG1-CG2. These atoms were excluded from the all-atom RMSD calculation.

The high accuracy version of the global distance test (GDT-HA) measures how accurately the C\(^\alpha\) atoms are positioned. GDT-HA ranges from 0 (bad) to 100 (good) and can be understood approximately as the percentage of C\(^\alpha\) atoms that are positioned correctly.

The contact area difference (CAD) score is a recently developed all-atom measure of model quality, which is based on comparing contact areas between residues in the
Choosing Accurate Structures from Protein Ensembles

experimental and model structures. CAD-score is an all-atom measure that is sensitive to backbone and sidechain positioning. CAD-score ranges from 0 (bad) to 1.0 (good).

MolProbity\textsuperscript{26} is the only non-native-centric score used in our assessment of success. Rather than comparing the model with the experimental structure, MolProbity assesses how well the model matches physical constraints derived from examining high-resolution X-ray crystal structures. Lower scores indicate better, more physically plausible models.

We assessed if the differences between the different averaging methods were significant by performing the two-sided sign test. We consider $P$ values $<$ 0.01 to be significant. The significance results are detailed in Supporting Information Table S1.

Generation of ideal ensembles

To test the ability of the different algorithms to compute average models close to the crystal structure, we used two idealized simple models that we call \textit{Gauss} and \textit{Gö}.\textsuperscript{27} By construction, these models have the crystal structure at the “center” and we are able to compare performance of the different averaging schemes without worrying about issues of force field accuracy and sampling quality. We used three simple proteins with a mixture of secondary structures as test systems: Protein A (pdb: 1bdd), Protein G (pdb: 2gb1), and Ubiquitin (pdb: 1ubq).

For the Gauss model, we perturbed each atom in a random direction by a random amount drawn from a Gaussian distribution with zero mean. We generated different datasets having an RMSD standard deviation of 0.5, 1.0, 2.0, or 4.0 Å. We also varied the number of structures with 50, 100, 200, 400, or 800 structures in each dataset. In total, we generated (3 proteins) $\times$ (4 standard deviations) $\times$ (5 sizes) = 60 Gaussian datasets.

For the Gö model, we used the SMOG server\textsuperscript{28} to generate an all-atom Gö model using the default parameters. By construction, the crystal structure is the energetically most favorable structure. Each system was simulated for five million time steps using the parameters recommended by the SMOG server at a temperature of 50, 70, 90, or 110 reduced units. For each protein at each temperature, we then generated samples of size 50, 100, 200, 400, or 800 by randomly selecting frames from the 5 million frame full trajectory. In total, we generated (3 proteins) $\times$ (4 temperatures) $\times$ (5 sizes) = 60 Gö datasets.

Generation of real-world ensembles

The purpose of these ensembles is to illustrate how the averaging methods work in ensembles derived from physical models. We examine three cases, which we call Native MD, Refinement, and NMR.

To generate the Native MD dataset, we took a collection of 14 experimental structures from small proteins (see Supporting Information Table S2 for list of structures) using the FF99SB-ILDN parameters,\textsuperscript{18,19} neutralized the proteins with sodium or chloride standard ions parameters and solvated them in TIP3P water.\textsuperscript{29} We simulated them with the pmemd program from Amber\textsuperscript{30} with a Langevin thermostat at 300 K and constant pressure for 10 ns (2 fs time step). We used the averaging protocol on the last 5 ns of simulation sampling every 25 ps.

For the Refinement dataset, we used ensembles provided by Mirjalili and Feig, who had the best performing refinement method\textsuperscript{32} in CASP10.\textsuperscript{33} Briefly, they performed restrained molecular dynamics simulations starting from a non-native structure provided by the CASP organizers. The restraints were either based on information given as “hints” from the CASP organizers or in cases with no hints, based on the assumption that core secondary structure elements are correct. The $\alpha$-carbons of the selected residues were restrained harmonically to their initial coordinates with a force constant of 1 kcal/mol/Å$^2$. They then selected a subset of conformations based on their DFiRE scores\textsuperscript{34,35} and the RMSD from the initial template model.

In the NMR dataset, we strive to identify a representative structure from an NMR ensemble. Assessment is performed by looking at the agreement with independently solved X-ray structures of the same sequence. We have tested this approach in three systems (ubiquitin, lysozyme, and protein G) using solution and solid state NMR data (see Supporting Information Table S3 for pdb codes).

RESULTS AND DISCUSSION

\textbf{NetAvg and CartAvg produce physically reasonable models}

Table I shows the MolProbity scores and various related statistics for the experimental structure, the experimental structure after energy minimization, and

\begin{table}[h]
\centering
\caption{MolProbity Scores and Related Statistics for Structures from the Gö and Gauss Models}
\begin{tabular}{lccccc}
\hline
 & Experimental$^a$ & Minimized experimental$^b$ & After CartAvg & After NetAvg \\
\hline
Clash score$^c$ & 23.0 & 5.5 & 5.6 & 5.5 \\
Bad rotamers & 15.6\% & 1.1\% & 6.6\% & 6.5\% \\
Ramachandran outliers & 2.9\% & 1.1\% & 1.9\% & 1.5\% \\
CB deviation & 0.4\% & 0.6\% & 0.6\% & 0.8\% \\
Bad bond lengths & 0\% & 0\% & 0\% & 0\% \\
Bad bond angles & 1.3\% & 0\% & 1.4\% & 4.2\% \\
MPscore$^d$ & 3.1 & 1.6 & 2.2 & 2.1 \\
\hline
\end{tabular}
\end{table}

\footnotesize{$^a$Average scores for PDB entries 1bdd, 2gb1, and 1ubq. $^b$Average scores for PDB entries 1bdd, 2gb1, 1ubq after energy minimization with the same protocol used for NetAvg and CartAvg. $^c$Bad steric clashes per 1000 atoms. $^d$MPscore indicates the model has similar characteristics to a crystal structure of this resolution. Lower is better.}
for NetAvg and CartAvg. Both CartAvg and NetAvg produce physically reasonable models with MolProbity scores of 2.2 ± 0.5 and 2.1 ± 0.5, respectively. These values are actually lower than for the experimental structures from the PDB. The relatively poor scores for the structures from the PDB can be attributed to the fact that these structures are quite old (published 1987—1992) and significant advances have been made in structure refinement and validation since then.

Minimization of the experimental structure significantly reduces the number of clashes, bad rotameric states, Ramachandran outliers, and overall MolProbity score, which indicates that the force field provides a physical description of the system that is reasonably consistent with the statistics from high-resolution crystal structures that are used in MolProbity.

The average structures have worse MolProbity scores than the minimized experimental structure. In particular, there are more bad rotamers and bad bond angles, which indicates that the guided minimization at the end of the averaging protocol has introduced some conformational strain into the protein. The structures from CartAvg and NetAvg are not quite at the level of a high-resolution crystal structure, but they display features comparable to crystal structures of approximately 2.2 Å resolution.

Both CartAvg and NetAvg dramatically improve the ideal Gaussian models

The Gauss model is primarily intended as a control. There are no couplings between atoms and the distribution of each atom is a simple isotropic Gaussian, so we expect the averaging protocols to work well. Indeed, both CartAvg and NetAvg perform well on this test (Fig. 4) and give dramatic improvements in RMSD compared to the medoid model of the ensemble. On average, CartAvg performs better than NetAvg. This difference is statistically significant (Supporting Information Table S1), but small (0.02 Å).

The final quality of the model is independent of the quality of the medoid. This is expected as even when the cloud of points for each atom is very broad (leading to a high medoid RMSD), we can still accurately estimate the center of the ensemble provided we have a modest number of samples.

NetAvg outperforms the other methods on samples from the Gō model

In almost all cases, both NetAvg and CartAvg improve the all-atom RMSD compared to the medoid structure (Fig. 5, upper). GDT-TS and CAD-score show similar improvements (Supporting Information Fig. S1). Table II summarizes the degree of improvement for each score with both averaging methods. For all measures, the NetAvg and CartAvg models are better than the medoid and the differences are statistically significant (Supporting Information Table S1). In all cases, NetAvg is significantly better than CartAvg.

The Gō model can produce complex, curved distributions of the type shown in Figure 2. NetAvg was designed explicitly to deal with such distributions and, as expected, performs better than CartAvg on this dataset.

NetAvg and CartAvg give modest improvements on the Native MD dataset

On average both NetAvg and CartAvg produce models that are closer to the crystal structure than the medoid of the ensemble for the Native MD dataset (Fig. 5, middle; Supporting Information Fig. S2). Table II shows that the mean improvements range from 4.0 to 8.5%, depending on the method and the score being assessed. Both NetAvg and CartAvg are significantly better than the medoid, but the difference between NetAvg and CartAvg is not significant for any of the three scores (RMSD, GDT-HA, CAD-score; Supporting Information Table S1). It is obvious that the performance on the Native MD dataset is not nearly as good as either of the ideal datasets (Gō, Gauss). We will return to this point later.

NetAvg and CartAvg give small improvements on the refinement dataset

This dataset is based on restrained molecular dynamics simulations with a relatively stiff force constant of 1 kcal/mol/Å² enforced on the backbone, which results in very “tight” ensembles, with small conformational fluctuations. Both NetAvg and CartAvg make small
improvements to the RMSD, GDT-HA, and CAD-score over the medoid (Fig. 5, lower; Supporting Information Fig. S3; Table II), although some of the results are not statistically significant: CAD-score for CartAvg, RMSD for NetAvg, and the differences between NetAvg and CartAvg for all three scores (Supporting Information Table S1). The averaging methods display their worst overall performance on this dataset. We will return to this point later.

The performance of averaging is related to the width of the conformational ensemble

When the conformational ensemble is very “tight,” there is relatively little that averaging can do to improve models. All of the structures are very similar and the average—which lies within the cloud of structures—must also be similar. In contrast, averaging can improve more when the ensemble is broad, because there is more to gain by choosing the correct position for each atom. This is evident in all of the datasets, particularly the Gauss (Fig. 4) and G/C22o (Fig. 5, upper) datasets.

This also helps to explain why the performance on the Refinement dataset was worse than other cases. The refinement simulations were conducted with relatively stiff position restraints, which prevent major fluctuations in conformation, which is evident in Figure 6. With such tight ensembles, it is difficult for any of the averaging methods to improve much over the medoid.

We also wondered about the possibility of outliers in the dataset to influence the average structure. These are structures that are very different from the rest of the ensemble and produce a significant deviation in the average structure. We devised a “pruning algorithm,” described in Supporting Information to reduce the effect of these outliers. In our experience with our datasets pruning has little effect on the resulting average structure (see Supporting Information Fig. S1).

For averaging to be successful, the atom cloud must overlap with the atom position in the native structure

The idea of averaging, regardless of the method used, is to select the most representative position for each atom from within the space defined by the ensemble. These methods are not capable of recovering an average position that lies outside of the cloud of positions sampled for each atom (Fig. 6). The implication of this is that if there is no overlap between the ensemble and the native state, averaging will not be able to correct this.

The results on the two ideal datasets are substantially better than on the two real-world datasets (Figs. 4 and 5). The ideal datasets have been constructed so that the experimental structure is at the center of the ensemble, while the real-world datasets are subject to bias and errors in the force field. Additionally, for refinement, the

Figure 5
NetAvg and CartAvg produce lower all-atom RMSD models than the medoid. The results are best for the Gō dataset and worst for the Refinement dataset.
simulations are not started from the experimental structure.

We can quantify the degree of systematic bias by using our knowledge of the experimental structure to construct a hybrid best model, where, for each atom, we take the position from the sampled ensemble that is closest to correct. If the RMSD of this hybrid best model is low, it means that each atom samples a position close to the correct position at least some of the time. If the RMSD is high, it means that there are atoms that never sample positions close to the correct position.

The average RMSD of the hybrid best models is much higher for the two real-world datasets (Fig. 7), which indicates that some atoms in the real-world dataset are systematically shifted away from the correct position. There is no way that any of the averaging schemes can recover from this.

**NMR ensembles**

Two things are apparent: (1) NMR ensembles are very tight and (2) the small number of structures does not guarantee a uniform sampling around an average state. We have used both solution NMR and solid-state NMR ensembles. For most applications involving computational structural biology the first model is used as representative of the ensemble. We have compared this first model as well as the average to what we would obtain from CartAvg and NetAvg (see Supporting Information Table S3). We observe that, with a single exception, CartAvg and NetAvg produce structures that are closer to

| Gō | Native MD | Refinement |
|----|-----------|------------|
|    | NetAvg    | CartAvg    | NetAvg    | CartAvg    | NetAvg    | CartAvg    |
| Average RMSD improvement (Å) | 0.42       | 0.25       | 0.10      | 0.18       | 0.10      | 0.12       |
| Average percent RMSD improvement | 31         | 20         | 4.2       | 8.5        | 2.3       | 3.0        |
| Average GDT-HA improvement | 0.08       | 0.05       | 0.04      | 0.05       | 0.01      | 0.02       |
| Average percent GDT-HA improvement | 11         | 8          | 5.5       | 6.5        | 2.5       | 2.6        |
| Average CAD-score improvement | 0.10       | 0.08       | 0.03      | 0.03       | 0.01      | 0.02       |
| Average percent CAD-score improvement | 14         | 11         | 4.0       | 4.0        | 1.5       | 2.5        |

**Figure 6**

Successful averaging requires overlap between the sampled ensemble and the experimental structure. This panel shows the lack of overlap between the simulated ensemble and experimental structure for one of the proteins (PDB ID: 2K5R) from the Refinement dataset. Positions for the N, C, CA atoms are shown as spheres for the experimental structure and clouds of points for the refinement ensemble. The coloring goes from blue (N-termini) to red (C-termini). Highlighted are a loop region (green) and the C-termini (red) for which the overlap between the ensemble and the experimental structure is poor.

**Figure 7**

The Native MD and Refinement datasets contain atoms that do not sample positions near the experimentally observed positions. This panel shows the average heavy atom RMSD for the hybrid best models for each dataset. For each atom in the hybrid best model, we select the position that was closest to that observed in the experimental structure.
the crystal structure (and always better than the raw ensemble average). However, the improvement is quite small due to the tightness of the ensembles. Ensembles coming from solid-state NMR are less tight and hence the improvement is larger. In most cases, atoms in the NMR ensemble are systematically displaced from their position in the crystal structure and just as we observed in the case of MD ensembles, NetAvg and CartAvg cannot compensate for the lack of overlapping conformations with crystallography.

Rules of thumb for selecting average structures from ensembles

We now give a few rules of thumb for selecting average structures from conformational ensembles.

- One should routinely use some kind of averaging procedure. This usually improves the structures by some amount over using centroid structures as representative, produces physically realistic structures, and is computationally cheap.
- An argument can be made for only using structures that were actually sampled (such as the medoid). However, as Figure 1 shows, entropy will always push toward higher RMSD, even though there are other structures that have lower energy and are a better representation of the ensemble.
- If the ensemble is “tight,” neither CartAvg or NetAvg can improve much over the medoid.
- For ensembles where some atoms have curved distributions, prefer NetAvg to CartAvg. For other ensembles, there is little difference between the methods.
- For large datasets (over 1000 structures), prefer CartAvg as the eigenvector centrality calculation in NetAvg becomes expensive.
- Average structures have better agreement with experimental data but have a nonphysical, purely geometrical, nature. In this case, it is essential to use some kind of minimization procedure to ensure the models are physically correct.

CONCLUSIONS

We have presented here two very computationally inexpensive methods (CartAvg and NetAvg) that can provide structures that are better representations of the ensemble than the standard practice of using the medoid. NetAvg is designed to deal with curved distributions of atoms and performs well on an ideal dataset. In practice, the effectiveness of both methods is often hindered by errors in the underlying atom distributions. We expect these protocols to be easily adapted in to pipelines in which large ensembles of structures are produced (e.g., structure prediction or docking studies).

REFERENCES

1. McCammon JA, Gelin BR, Karplus M. Dynamics of folded proteins. Nature 1977;267:385–390.
2. Levitt M, Sharon R. Accurate simulation of protein dynamics in solution. Proc Natl Acad Sci USA 1989;85:7577–7581.
3. Piana S, Dror RO, Shaw DE. How fast-folding proteins fold. Science 2011;334:517–520.
4. Wüthrich K. Protein structure determination in solution by NMR spectroscopy. J Biol Chem 1990.
5. Brünger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS, Read RJ, Rice LM, Simonson T, Warren GL. Crystallography & NMR system: a new software suite for macromolecular structure determination. Acta Crystallogr D Biol Crystallogr 1998;54:905–921.
6. Simons KT, Bonneau R, Ruczinski I, Baker D. Ab initio protein structure prediction of CASP III targets using ROSETTA. Proteins 1999 (Suppl 3):171–176.
7. Sali A, Potterton L, Yuan F, van Vlijmen H, Karplus M. Evaluation of comparative protein modeling by MODELLER. Proteins 1985:23:318–326.
8. Roy A, Kucukural A, Zhang Y. I-TASSER: a unified platform for automated protein structure and function prediction. Nat Protoc 2010;5:725–738.
9. Perez A, Yang Z, Bahar I, Dill KA. FlexE: using elastic network models to compare models of protein structure. J Chem Theory Comput 2012;8:3985–3991.
10. Nugent T, Cozzetto D, Jones DT. Evaluation of predictions in the CASP10 model refinement category. Proteins 2014;82:98–111.
11. Sullivan DC, Kuntz ID. Conformation spaces of proteins. Proteins 2001;42:495–511.
12. Shao J, Tanner SW, Thompson N, Cheatham TE. Clustering molecular dynamics trajectories: I. Characterizing the performance of different clustering algorithms. J Chem Theory Comput 2007;3:2312–2334.
13. Keller B, Daura X, van Gunsteren WF. Comparing geometric and kinetic cluster algorithms for molecular simulation data. J Chem Phys 2010;132:074110.
14. Newman M. The mathematics of networks. In: Blume LE, Durlauf SN, editors. The New Palgrave Encyclopedia of Economics; Palgrave Macmillan, Basingstoke; 2008.
15. Newman M. Networks: an introduction, Google Books, Oxford University Press, New York, NY (2010).
16. Wills RS. Google’s pagerank. Math Intelligencer 2006;28:6–11.
17. Hess B, Kutzner C, van der Spoel D, Lindahl E. GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. J Chem Theory Comput 2008;4:435–447.
18. Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, Simmerling C. Comparison of multiple Amber force fields and development of improved protein backbone parameters. Proteins 2006;65:712–725.
19. Lindorff-Larsen K, Piana S, Palmo K, Maragakis P, Klepeis JL, Dror RO, et al. Improved side-chain torsion potentials for the Amber ff99SB protein force field. Proteins 2010;78:1950–1958.
20. Onufriev A, Bashford D, Case DA. Exploring protein native states and large-scale conformational changes with a modified generalized born model. Proteins 2004;55:383–394.
21. Zhang C, Vasmatzis G, Cornette JL, DeLisi C. Determination of atomic desolvation energies from the structures of crystallized proteins. J Mol Biol 1997;267:707–726.
22. Liu DC, Nocedal J. On the limited memory BFGS method for large scale optimization. Math Program 1989;45:503–528.
23. Coutias EA, Seok C, Dill KA. Using quaternions to calculate RMSD. J Comput Chem 2004;25:1849–1857.
24. Zemla A. LGA: a method for finding 3D similarities in protein structures. Nucleic Acids Res 2003;31:3370–3374.
25. Olechnovič K, Kulberkytė E, Venclovas C. CAD-score: a new contact area difference-based function for evaluation of protein structural models. Proteins 2013;81:149–162.
26. Davis JW, Murray IW, Richardson JS, Richardson DC. MolProbity: structure validation and all-atom contact analysis for nucleic acids and their complexes. Nucleic Acids 2004;32:W615–W619.
27. Gö N, Scheraga HA. On the use of classical statistical mechanics in the treatment of polymer chain conformation. Macromolecules 1976;9:535–542.
28. Noel JK, Whitford PC, Sanbonmatsu KY, Onuchic JN. SMOG@ctbp: simplified deployment of structure-based models in GROMACS. Nucleic Acids Res 2010;38:W657–W661.
29. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. J Chem Phys 1983;79:926.
30. Case DA, Darden TA, Cheatham TE, Simmerling CL, Wang J, Duke RE, Luo R, Walker RC, Zhang W, Merz KM, Roberts B, Hayik S, Roitberg A, Seabra G, Swails J, Goetz AW, Kolossváry I, Wong KF, Paesani F, Vanicek J, Wolf RM, Liu J, Wu X, Brozell SR, Steinbrecher T, Gohlke H, Cai Q, Ye X, Wang J, Hsieh MJ, Cui G, Roe DR, Matthews DH, Seetin MG, Salomon-Ferrer R, Sagui C, Babin V, Luchko T, Gusarov S, Kovalenko A, Kollman PA. Amber12. University of California San Francisco; 2012.
31. Darden T, York D, Pedersen L. Particle mesh Ewald: an N-log(N) method for Ewald sums in large systems. J Chem Phys 1993;98:10089–10092.
32. Mirjalili V, Feig M. Protein structure refinement through structure selection and averaging from molecular dynamics ensembles. J Chem Theory Comput 2013;9:1294–1303.
33. Moult J, Fidelis K, Kryshtafovych A, Schwede T, Tramontano A. Critical assessment of methods of protein structure prediction (CASP)—round X. Proteins 2014;82(Suppl 2):1–6.
34. Yang Y, Zhou Y. Ab initio folding of terminal segments with secondary structures reveals the fine difference between two closely related all-atom statistical energy functions. Protein Sci 2008;17:1212–1219.
35. Yang Y, Zhou Y. Specific interactions for ab initio folding of protein terminal regions with secondary structures. Proteins 2008;72:793–803.