Protein loops with multiple meta-stable conformations: a challenge for sampling and scoring methods

Amélie Barozet, Marc Bianciotto, Marc Vaisset, Thierry Simeon, Hervé Minoux, Juan Cortés

ORIGINAL ARTICLE

Protein loops with multiple meta-stable conformations: a challenge for sampling and scoring methods

(Short title: Protein loops with multiple meta-stable conformations)

Amélie Barozet[1,2] | Marc Bianciotto[2] | Marc Vaisset[1] | Thierry Siméon[1] | Hervé Minoux[2] | Juan Cortés[1]

[1]LAAS-CNRS, Université de Toulouse, CNRS, Toulouse, France

[2]Sanofi recherche & développement, Integrated Drug Discovery, Molecular Design Sciences, , 13 quai Jules Guesde, BP 14, 94403 Vitry-sur-Seine Cedex, France

Correspondence
Amélie Barozet, LAAS-CNRS, Université de Toulouse, CNRS, Toulouse, France.
Juan Cortés, LAAS-CNRS, Université de Toulouse, CNRS, Toulouse, France
Email: abarozet@laas.fr, juan.cortes@laas.fr

Funding information
French National Association of Research and Technology (ANRT), contract number 2016/0239. This work used the HPC resources of the CALMIP supercomputing center under the allocation 2016-P16032.

Flexible regions in proteins, such as loops, cannot be represented by a single conformation. Instead, conformational ensembles are needed to provide a more global picture. In this context, identifying statistically meaningful conformations within an ensemble generated by loop sampling techniques remains an open problem. The difficulty is primarily related to the lack of structural data about these flexible regions. With the majority of structural data coming from X-ray crystallography and ignoring plasticity, the conception and evaluation of loop scoring methods is challenging. In this work, we compare the performance of various scoring methods on a set of 8 protein loops that are known to be flexible. The ability of each method to identify and select all of the known conformations is assessed, and the underlying energy landscapes are produced and projected to visualize the qualitative differences obtained when using the methods.

Statistical potentials are found to provide considerable reliability despite their being designed to tradeoff accuracy

Abbreviations: PCA, Principal Component Analysis; PDB, Protein Data Bank; RMSD, Root-Mean-Squared Deviation.

for lower computational cost. On a large pool of loop models, they are capable of filtering out statistically improbable states while retaining those that resemble known (and thus likely) conformations. However, computationally expensive methods are still required for more precise assessment and structural refinement. The results also highlight the importance of employing several scaffolds for the protein, due to the high influence of small structural rearrangements in the rest of the protein over the modeled energy landscape for the loop.

KEYWORDS
Protein loop modeling, Flexible protein loops, Conformational ensembles, Scoring functions

1 | INTRODUCTION

1.1 | Motivation

Faced with increasingly complex modeling and prediction challenges, the bioinformatics community is constantly developing new methods to better apprehend structural and dynamic properties of proteins. The existence of international competitions involving numerous teams in the fields of protein folding [1], docking [2] or antibody modeling [3] reveals the interest in evaluating and comparing available methods.

For the particular problem of protein loop modeling, methods are often evaluated on their ability to identify the "native" conformation among a pool of decoys. Such evaluation inherently comes from the nature of available data, largely extracted from X-ray crystallography and thus mostly limited to providing a single snapshot of the structure. However, due to the intrinsic flexibility of coil fragments, such a representation is only partial, and so are the performed evaluation and comparison of loop modeling methods. Furthermore, a recent work by Marks and co-workers [4] shows that existing methods struggle to correctly model flexible loops.

Yet, an accurate evaluation of the quality of a protein loop structural model is important for many applications. It is primarily useful in the context of loop structure prediction [5, 6, 7, 8, 9, 10, 11], to determine the stable conformation(s) that the loop is most likely to adopt. It may also be used as a filter to eliminate high-energy conformations before costly downstream steps, for instance in ensemble docking [12]. Loop structure evaluation is furthermore employed in loop design, either to verify that the proposed loop will preferentially adopt the desired conformation [13], or, on the contrary, that it will not adopt an undesirable one (negative design) [14, 15, 16].

Most loop modeling approaches involve sampling and scoring methods that are independent from each other (i.e. they can be mixed and matched within different protocols). Various methods have been proposed to generate loop conformational ensembles, aiming to capture the whole diversity of the states potentially adopted by the loop (see [17] for a review). Then, various scoring functions can be applied to assess the likelihood of each sampled state. These scoring functions greatly differ both from a conceptual and from a computational point of view. Among them, atomistic physics-based methods aim at estimating an energy based on the calculation of forces involved in the structure [18, 19, 20]. They are, in general, computationally demanding and highly sensitive to slight conformational

changes. Conversely, knowledge-based methods [21, 22, 23, 24, 25] are a tempting option due to their low computational cost. These methods exploit available experimental data to assess the quality of a given structure. However, most of the published data come from X-ray crystallography, which captures only a snapshot of a protein's structure, with potential artifacts due to crystal packing interactions. In that context, the behavior of knowledge-based methods is unpredictable when dealing with a very flexible protein or region.

This paper compares the performance of several scoring methods on several systems comprising a loop known to be flexible. After exhaustively sampling the conformational space of the different loops using our recent loop sampling method MoMA-LoopSampler [26], the scoring methods are employed to evaluate the sampled loop states. The ability of the scoring methods to identify one or several of the known conformations is assessed. The implicitly modeled conformational landscape is then analyzed and visualized, using an appropriate 2D projection for the sampled loop states. Consistency of this landscape with the known conformations of the loop is verified, and the influence of the surrounding protein conformation is detailed.

We also discuss possible effects of experimental conditions on the performance of the scoring methods at identifying experimentally observed loop conformations. Indeed, while the analyzed landscape is meant to represent the conformational space of the free loop, the experimental X-ray loop conformations might be influenced by interactions with ligands, ions, cosolvents, or contacts with crystal mates, depending on the strength and nature of these intermolecular interactions. For example, the interaction between the loop and a ligand might select a loop conformation that is already accessible in the free energy landscape, or stabilize a loop conformation that has a low probability in the free protein by an induced fit effect. In this latter case, one cannot expect the liganded conformation to be identified among the top-scoring ones if the ligand is not included in the model. Nevertheless, suitable scoring methods should be able to associate it with an accessible region in the energy landscape.

By analyzing the produced landscapes in addition to the agreement between known structures and top-scoring loop states, this work aims at identifying the qualitative differences between the results obtained by the various scoring methods. In turn, by examining these differences, we aim at providing guidelines as to which methods to employ depending on the problem at hand, and what conditions should be gathered to expect accurate results. In particular, such a comparison is intended to verify whether the tradeoff between computational cost and accuracy offered by faster statistics-based potentials remains appealing. We do not extract an overall ranking from our analyses, since this would make little sense in our opinion, and because different scoring methods may have different strengths and weaknesses, as discussed along this paper.

The remainder of this section lists the loop systems used in this work, and describes the scoring methods that will be compared. Section 2 presents the *in silico* protocol employed. Section 3 details the results obtained by the different scoring methods on the different systems, while Section 4 attempts to summarize the various results and to draw more general trends about the behavior of individual scoring functions.

Throughout the manuscript, we make the distinction between a *loop*, set of residues and atoms forming a flexible protein fragment; a loop *state*, fully determined by the values of its internal degrees of freedom; and a loop *conformation*, defined as a consensus state, or a limited set of similar states.

1.2 | Loop systems

Eight flexible protein loops that have been crystallized in at least two different conformations were gathered for this work. The list, together with relevant information, is provided in Table 1. A more detailed list is available in the Supplementary Material (Table S1). The atomic models of the loops have satisfying fit to electron density, as shown by the list of RSRZ (real-space R-value Z-score [27, 28]) outliers provided in Table S2, and by Figure S1. Information

about crystallization conditions, including crystal contacts and the presence of ligands, is provided in Table S3, which highlights potential sources of conformational change for the loops. Visualization of the different known conformations, along with distances between them are provided in Figure S2 and Table S5, respectively. Two loops in the same conformation are maximum 0.81 Å distant (backbone-RMSD-wise), while two loops in different conformations are at least 2.16 Å distant. All systems, except #1 and #7, were also used in a related publication that motivated our work [4].

These systems were chosen so as to provide a variety of loop lengths and protein sizes. The set includes loops with two or three known conformations. Systems #1 (streptavidin) and #7 (triosephosphate isomerase, TPI) were included in this study because they are well-known proteins in which the flexibility of the loop has been shown to play a functional role.

1.3 | Scoring methods

Different scoring methods were tested in this work (Table 2). The first method employs the AMBER force field ff14SBonlysc [20] to score loop states. It is a physics-based method that works on an all-atom protein model, including side-chains. The second method uses the ref2015 ROSETTA scoring function [29]. It is a hybrid method that combines physics-based terms such as those employed by AMBER force fields with other statistical terms. This scoring function also uses an all-atom structure with side-chains. Although ROSETTA includes an option to perform relaxations after replacing side-chains with their centroid, this was not tested in this work. Both AMBER force fields and ROSETTA scoring functions are very sensitive to slight divergences from the 'ideal' geometry or to minor steric clashes. Consequently, the associated methods model very rough molecular energy landscapes where low-energy states border on high-energy ones. In order to better assess the stability of a structural model, a relaxation has to be performed so as to allow the modeled state to fall into the closest basin. Those relaxations can turn out to be prohibitively computationally expensive, especially when the number of loop states to score increases.

All the other tested methods are statistical potentials, which do not require relaxation. Two of them, DFIRE2 [21] and SOAP-Loop [23] are all-atom potentials that require the side-chains to be placed in the structure to score. DFIRE2 is a simple distance-dependent potential. It uses a single descriptor for each heavy atoms pair: the triplet a_1, a_2, r where a_1 and a_2 are the residue-specific atom types of the first and second atoms in the pair, respectively, and r is the distance between them. SOAP-Loop also employs inter-atomic distances as descriptors, but includes terms related to the orientation between pairs of covalent bonds and the relative atomic surface accessibility.

The last three methods are KORP [25], SBROD [24] and a very simple statistical potential solely based on the loop's ϕ and ψ dihedrals [22], which will subsequently be called *Torsions-only*. They are coarse-grained potentials and none of them requires the side-chains to be modeled. KORP only needs the positions of the N, C_α and C backbone atoms. It uses one distance and five angular features to describe the relative position and orientation of each amino-acid pair. Torsions-only uses the amino-acid type, along with the ϕ and ψ angles of the residues in the loop as the only descriptors. SBROD uses many features gathered into four groups: residue-residue pairwise features, backbone atom-atom pairwise features, hydrogen bonding features and solvent-solvate features. While DFIRE2, SOAP-Loop, KORP and Torsions-only are all Bayesian-based, SBROD uses the *Ridge Regression* machine learning technique to optimize the weights in the linear model. Another major difference between SBROD and the other statistical potentials compared in this work is the training dataset. DFIRE2, SOAP-Loop, Torsions-only and KORP use non redundant sets of protein structural data to derive observed frequencies of the different features. Conversely, SBROD is trained on sets of decoy models and is designed to discriminate between well-folded and misfolded structures, which makes it fundamentally different from the other knowledge-based potentials.

2 | MATERIAL AND METHODS

2.1 | Preprocessing of structure files

Structural data corresponding to the IDs listed in Table 1 were extracted from the Protein Data Bank (PDB) [30]. Ligands, ions, water molecules and other non-protein elements were removed. The protein chains listed in Table 1 were isolated and kept as individual structures. These structures were then idealized using AMBER (see Section S2 for the detailed protocol), with a resulting median RMSD of 0.2 Å on all heavy backbone atoms upon relaxation. The structures thus obtained were used as scaffolds for loop sampling.

2.2 | Sampling loop states

Loop sampling was performed using MoMA-LoopSampler [26], a recent method performing a global exploration of the conformational space. This method generates collision-free states for the loop backbone by concatenating tripeptides from a dedicated database and employing a semi-analytical inverse kinematics method to close the loop. Exhaustiveness of the sampling was showcased on benchmark sets of 9-, 12- and 15-residue loops [26].

Once a state is sampled for the backbone, side-chains are placed with an in-house procedure detailed in Section S3. Briefly, the method samples χ angles following the continuous rotamers from BASILISK [31]. Slight collisions are solved using random perturbations of the dihedral angles in the side-chains of the loop or in its surroundings. If the method fails to place side-chains without collisions, the sampled backbone state is rejected.

Each scaffold was employed to sample 5,000 states with side-chains using MoMA-LoopSampler, except for the 8 streptavidin scaffolds, for which the loop length and the constrained environment allowed an exhaustive brute force sampling, concatenating all possible fragments from the tripeptide library. This represents 5,338 states from scaffold 2F01(A), 3,825 from 2F01(B), 4,042 from 3RY1(A), 1,304 from 3RY1(B), 702 from 3RY1(C), 2,820 from 3RY1(D), 5,320 from 3RY2(A) and 3,611 from 3RY2(B).

2.3 | Scoring loop states

For every scoring method except AMBER and ROSETTA, the binaries to score loop states were either downloaded from the dedicated websites, or provided by the authors. The binaries were used to score each individual loop state. For AMBER and ROSETTA, sampled states were first relaxed before being scored. Methods versions are given in Table S4, while execution details (including relaxation protocols for AMBER and ROSETTA) are provided in Section S4.

2.4 | Landscape reconstruction

The combination of the sampled states and their associated scores can be used to represent the energy landscape modeled by a given scoring method. Each sampled state corresponds to a point in an n-dimensional space where n corresponds to the number of degrees of freedom of the sampled loop. In order to visualize these points, they were projected in 2D space. This projection was given by the two principal components of a Principal Component Analysis (PCA) run on the aggregated states sampled from all scaffolds, for each loop system. The Cartesian coordinates of the C_α atoms of the loop were used as variables for the PCA.

The 2D space was then discretized using a grid (using 40 bins on each axis), where each cell was colored according to the best score of all the states projected within it. The presence of a few cells with a score far above the mean tends

to make the landscape appear flat. To circumvent this, cells whose associated score was more than three standard deviations above the mean were considered empty (as if no sampled state was projected there). Empty cells were considered to have the maximum score observed among the populated cells. A bicubic interpolation method was used to smooth the landscapes.

3 | RESULTS

Results are presented from several perspectives. The quality of the sampled conformational ensembles is first analyzed (Section 3.1), because sampling is the first bottleneck of landscape modeling. Indeed, if statistically likely conformations are missing in the generated ensemble, the energy landscape will be inaccurate, whatever scoring method is subsequently used. Next, since they are a determining criterion in the choice of a scoring method, the running times observed for the different functions are detailed (Section 3.2). The relationship between known conformations and scores is then examined. Such an analysis is complex for two reasons. The first reason is that known conformations are not necessarily the most stable conformations when the protein is alone in solution: instead, they may be stabilized by ligands, crystal contacts or metallic ions. However, the fact that they are observed does indicate that they are statistically likely conformations, which should not be eliminated by a filtering method. The second reason why this analysis is challenging is that the methods we compare provide scores and not binary good/bad classification, so that eliminating unlikely conformations requires setting an arbitrary threshold, either on the score itself or on the rank. Given these considerations, we analyze how known conformations are scored from two different points of view. We first report the ranks measured for the the sampled states that are closest to known conformations (Section 3.3), and then provide the distances between top-scoring states and known conformations (Section 3.4). Finally, the landscapes implicitly modeled by the different scoring functions are presented, aiming to provide a more exhaustive comparison of the results they provide (Section 3.5). In order to determine whether the different functions can agree despite their fundamental differences, Section S7 analyzes the correlations between the scoring methods, both overall (Figure S4) and depending on the predicted scores (Figure S5).

3.1 | Sampling known conformations

Figure 1 shows the distance of the closest sampled state to each experimentally-determined conformation, from each scaffold. Note that throughout the results section, distances between two loop states are given as the root-mean-square deviation (RMSD) of the heavy atoms of the backbone. To better understand what the figure shows, let us illustrate with the example of scaffold 3AHW(A) from RNU2. The scaffold comes from a crystal structure with the loop in conformation 3 (as indicated by the color of the outer circle). From this scaffold, a state within 1 Å of conformation 3 was sampled. However, no state was sampled closer than 2.7 Å from conformation 1 or 3.2 Å from conformation 2. The data used for the radar chart representations in Figure 1 is provided in a table in Supplementary Material (Table S6).

Overall, the results suggest that the sampled ensemble contains states close to each known conformation. One can however notice a limitation: the crystallographic structure corresponding to the employed scaffold is almost always sampled more closely than the other experimentally-determined conformations. Although the presence of such a bias is not surprising, the RMSD difference can be substantial in some cases (e.g. conformation 1 for streptavidin, or conformation 2 for PTPN9). A closer observation of the concerned structures reveals large rearrangements in the loop environment that accompany the loop conformational change and thus hinder the sampling of some regions

of the loop's conformational space, where other known conformations could be found. For example in streptavidin, the position of the backbone of GLU-51 in conformation 1 coincides with the position of the side-chain of ARG-84 in conformations 2 and 3. As a consequence, loop conformation 1 cannot be sampled very closely from scaffolds originating from conformations 2 or 3. Similarly in RNU2, there are major backbone collisions between the loop in conformation 2 (GLY-33 - ASP-34) and the scaffold in conformation 3 (SER-74 - ARG-75).

This problem illustrates a limitation common to all loop sampling methods that consider the rest of the protein as a rigid body, and underlines the necessity to better account for flexibility outside the loop. A possibility could be to remove all the side-chains in a large surrounding of the loop anchors before sampling, although that may unnecessarily broaden the space accessible to the loop and make an exhaustive sampling harder. Another alternative could be to employ a scaffold ensemble instead of a single one. This is further discussed in Section 4.

3.2 | Running times

Running times differ by several order of magnitude from one scoring method to another, and obviously depend on the system's size. Figure 2 reports the average time required to score one sampled state for three systems of different sizes. ROSETTA and AMBER are by far the most costly methods, with comparable running times using the relaxation protocol adopted in this work. SOAP-Loop is the slowest statistical method, possibly due to its evaluation of the atomic surface accessibility. SBROD is the next method in terms of running time, followed by DFIRE2. Torsions-only and KORP are the least expensive methods, presumably because the former only employs the provided ϕ and ψ angles of the loop and the latter offers a convenient batch mode, where the structure of the whole protein is only provided once.

3.3 | Ability to rank near-native conformations

Figure 3 gives the ranks of the five closest sampled states to each known conformation. As an example to read this figure, let us consider the sixth line corresponding to known conformation 1 of NTPase loop. The upper parts of the disks relate to the five closest states to conformation 1 sampled from the first scaffold (4KFR(B)). The ranks obtained using AMBER place the third closest state to conformation 1 among the ten states with the lowest energies, while the other four closest states have a rank above 100. None of the other scoring method places any of the five states closest to conformation 1 among the ten best scoring states. This data is also given as tables in Supplementary Material (Tables S7-S13).

There are global trends regarding how well the states in the vicinity of the different known conformations are scored. These trends are unsurprisingly system-dependent: sampled states similar to known conformations of PTPN9 are well identified, while states similar to known conformations of NTPase are not. The accuracy with which known conformations are sampled can partially explain this trend, but other factors are needed to explain these inter-system differences, such as the intrinsic flexibility of the loop, or how favorable its surroundings are to the creation of attractive or repulsive contacts. Disparities are also observed between the different conformations of a single system. The nature of the conformation (stabilizing contacts possibly involving ligands or crystal mates that are not included in the model, "canonical" shape, …) may be a determining factor.

Surprisingly, results do not show a clear correlation between the performance of the scoring methods and the expected difficulty to identify near-native conformations due to interactions with ligands and crystals mates (see Table S3) that are excluded from the model. This can be illustrated on Streptavidin: due to the presence of a ligand and crystal contacts, conformation 1 should be more difficult to identify than conformations 2 and 3, but, in general,

scoring methods do not perform better for the *a priori* easier cases. Note also that conformation 1 of Streptavidin has been experimentally observed in different conditions. A similar behaviour can be observed for PTPN9: the prediction of the loop conformation corresponding to scaffold 4ICZ(A) should be a relatively difficult case due to interactions with the ligand that, in principle, could stabilize this loop conformation. On the other hand, predicting the other loop conformation from scaffolds 2PA5(A), 4GE2(A) and 4GE6(B) should be easier since there are neither interactions with ligands nor crystal contacts. However, the performance of scoring methods in identifying the two conformations is very similar. In particular, DFIRE2 and KORP provide very good results in both cases. One can also observe counter-intuitive results. For instance, the two conformations of the Pot1pC loop should correspond to a difficult case due to crystal contacts and interactions with the ligand, but all the scoring methods perform relatively well. On the contrary, the performance is significantly worse for MR-MLE (for the two conformations), which should be an easier case, in principle. This point, regarding the expected difficulty of the predictions, is further discussed in the next section.

Due to the variability of the performance depending on the system, and independently from possible effects of experimental conditions, it is hard to pinpoint the best scoring method with respect to this evaluation criterion. Nevertheless, one can observe an overall good performance of KORP and DFIRE2 in identifying states near known conformations. KORP is the method that gives the best scores to sampled states around conformation 1 of streptavidin. It is rather consistent across the different conformations to identify, but fails to detect any state close to conformations 2 and 3 of streptavidin, conformation 1 of NTPase, conformation 2 of MR-MLE and conformation 3 of RNU2 (but these last two conformations are not well identified by any of the tested scoring methods). DFIRE2 performs slightly better than KORP on conformation 3 of streptavidin but otherwise misses the same conformations and conformation 2 of NTPase. Despite their similar overall success, KORP and DFIRE2 do not identify the same near-native states, perform differently depending on the sampling scaffold and can be of different precision when identifying a conformation.

AMBER rarely fails completely for a system, but it is less robust than KORP or DFIRE2, with ranks varying substantially from one of the 5 closest states to another. This may come from a lack of convergence of the relaxations, or from the roughness of the conformational landscape modeled by this function. ROSETTA has a similar performance and the same shortcoming concerning robustness. This is not surprising since ROSETTA also models a rather rough landscape and heavily depends on the prior relaxation. SOAP-Loop, despite not needing a relaxation, like AMBER and ROSETTA do, obtains similar results to these two methods. SBROD rarely scores near-native states better than other methods and performs badly overall. This may be due to the major differences in the way this method was designed compared to other statistical methods. Finally, Torsions-only is the method with the least satisfying results if compared to other scoring functions. However, taking into account its extreme simplicity, results are still remarkable, and it turns out to be as good as other methods at identifying states similar to known conformations for Pot1pC or NTPase.

3.4 | Top scoring loop states

The presence of known conformations among the top scoring states was then analyzed. Figure 4 shows the number of states within different distance thresholds of known conformations among the top 1% states identified by each scoring method from each scaffold. The distances between the five top-scoring states (for each method-scaffold combination) and their closest known conformation were also calculated and are provided in Figure S3 and Tables S14-S20.

As mentioned above, the difficulty to identify experimentally observed conformations may depend on the existence of stabilizing interactions with ligands and/or crystal mates (which are not included in the model), or on other experimental conditions. Based on the data presented in Table S3 and on a visual analysis of the structures, we man-

ually assigned each scaffold a predicted difficulty among three levels defined as follows. Easy cases correspond to scaffolds for which (at least) the corresponding loop conformation should be easy to identify due to the absence of interactions with ligands and crystal contacts. For intermediate cases, these interactions may exist, but are relatively weak. Difficult cases are those for which interactions with ligands and/or crystal mates are clearly present in the X-ray structures. Results presented in Figure 4 do not show a clear correlation between the difficulty level and the performance of scoring methods, contrary to what could be expected *a priori*.

The difficulty to identify “native" conformations can also be due to a sampling problem, as explained in Section 3.1. In the present analysis, in general, this is not an issue when the loop is sampled from the scaffold corresponding to a given conformation. There are only a few cases for which states below 1.5 Å to the known conformation were not sampled: 3AGO(A), 3QPE(D), 3VCC(A) and 4KFR(B) (see Figure 1 and Table S6). However, sampling can be a clear limitation in some cases when trying to identify loop conformations from a different scaffold. This happens for some of the conformations of streptavidin, PTPN9 and RNU2.

Regarding the performance of the different scoring methods, Figures 4 and S3 show that KORP performs well at identifying known conformations. Although it performs heterogeneously on the different systems (like the other methods do), the number of states close to known conformation among its identified top 1% states is comparable to or larger than that of other methods. Unsurprinsingly, KORP shows a tendency to better identify the conformation corresponding to the scaffold the state was sampled from, but this tendency is not as strong as it is for DFIRE2, for example, and KORP is capable of identifying different known conformations from one scaffold. It is the case for Pot1pC, TPI or MR-MLE, for instance.

ROSETTA also identifies many states close to known conformations (with some variety in the identified conformations) although it does not perform as well as KORP or DFIRE2 for a few systems, for instance PTPN9. SOAP-Loop performs well, and although it does not identify as many states close to known conformations as KORP does overall, it identifies many of the very close states (within 1.5 Å of known conformations). DFIRE2 identifies many states close to known conformations, but is rarely capable of identifying several known conformations from one single scaffold.

The other methods are not as good at identifying known conformations, missing all known conformations for at least two systems out of eight: AMBER identifies known conformations for streptavidin, Pot1pC and TPI among its top 1% and top 5 states, but almost completely misses known conformations for MR-MLE and UTB. SBROD identifies very few states corresponding to known conformations for Streptavidin, MR-MLE, NTPase and PTPN9, but identifies many more states close to known conformations of UTB than any other method, including within its top 5 states. Torsions-only identifies few states close to known conformations for all systems, except for Pot1pC, for which its results are comparable to those of other methods.

It should be noted that there may exist statistically probable conformations different from those observed in the available crystallography structures. Indeed, the fact that no known conformation appear among the top-ranked states may be due to the identification of other locally stable conformations. However, this is impossible to confirm in the absence of additional structural data. The current accessible information designates KORP, but also ROSETTA, SOAP-Loop and DFIRE2 to a lesser extent as the most reliable methods from the top-scoring-states point of view.

Figure S7 completes this analysis by providing the plots of RMSD to known conformations vs predicted score. This representation shows the presence of a “folding funnel" for individual scoring method-scaffold-target combinations. For instance, the folding funnel is predicted by KORP from scaffold 2F01(A) with respect to conformation 1, or by ROSETTA from scaffold 3AGN(A) with respect to conformation 1. These plots confirm the good performance of DFIRE2, KORP, but also ROSETTA in identifying the known conformations of many systems. Also interesting to mention, this figure demonstrates the quality of the sampling, by showing that MoMA-LoopSampler generated many states close to known conformations.

3.5 | Modeled energy landscapes

Directly analyzing the energy landscapes modeled by the different scoring methods in combination with MoMA-LoopSampler allows to gain a more global insight into the topography induced by these methods. Only the most informative landscapes are depicted here but the full landscapes of all systems are provided in Supplementary Material (Figure S6).

For streptavidin (Figure S6(a)), AMBER manages to identify the different basins with higher precision than the other methods. From the first scaffold, it even identifies two main basins, one around conformation 1 and one close to conformations 2 and 3. It consistently places a basin around the conformation corresponding to the employed scaffold. The only other method that does so for all scaffolds is KORP. Other methods sometimes identify the basin around conformation 1 or the one around conformations 2 and 3, but not both.

In MR-MLE modeled landscapes (Figures 5(a) and S6(b)), the basin identified by DFIRE2 (using any scaffold) is located around conformation 1, although it sometimes overlaps with conformation 2. KORP, however, identifies a basin that is around known conformation 2 for all scaffolds. Even though both conformations can be identified as stable conformations for this system, none of the methods clearly identifies both basins. Both ROSETTA and AMBER model a rather flat landscape.

Landscapes modeled for NTPase (Figure S6(c)) are not in accordance with the known conformations. This may be due to these conformations being artifactual, and not likely to exist in solution. However, one cannot rule out a deficiency of the scoring methods in this case either.

Landscapes obtained for Pot1pC (Figure S6(d)) are consistent with the known conformations for this system. Depending on the method, a basin is identified in the area around conformations 1 or 2 or between them.

The case of PTPN9 (Figures 5(b) and S6(e)) is a very good illustration of the power of statistical methods. One of the two known conformations is in an area sampled by the method using some but not all of the scaffolds, while the other is located on the edge of the projected landscape. KORP and DFIRE2 very clearly identify that area as a deep basin whereas landscapes obtained by other methods (AMBER and ROSETTA in particular) are much less consistent with the crystallographic structures of the loop. The landscapes modeled by KORP or DFIRE2 could guide a more thorough sampling around that basin.

Among the three known conformations for the loop in RNU2, only conformation 1 is identified as stable by the scoring methods, and only from the first scaffold (Figures 5(c) and S6(f)). For all methods except SBROD, the identified area is fuzzy, although the basin gets deeper around the projected position of conformation 1 for DFIRE2, ROSETTA, or SOAP-loop. SBROD identifies a very narrow and deep basin around conformation 1 from the first scaffold, with a precision that is much higher than that of other methods.

KORP is the method that produces the most consistent and precise landscapes for TPI (Figures 5(d) and S6(g)). While most methods place both stable conformations in a very vast basin, KORP models one relatively narrow basin, deeper around the crystallographic conformation originally present in the scaffold. The example of TPI clearly shows the influence of the starting scaffold. While for some systems the landscapes look similar whatever the scaffold, landscapes produced for TPI from scaffolds originally in conformation 1 (1YPI(A), 1YPI(B)) are clearly different from those produced from scaffolds in conformation 2 (2YPI(A), 2YPI(B)), with a clear displacement of the main basin towards one or the other known conformation.

Landscapes modeled for UTB by AMBER and ROSETTA are very rough (Figure S6(h)), making it difficult to draw any conclusion from them. Landscapes generated by statistical methods (especially DFIRE2, KORP and SBROD) are again more consistent with experimental data.

The different landscapes overall confirm that statistical methods produce smoother landscapes. The main pitfall

of this is that they lack precision and are fuzzy. While they usually manage to identify a main basin containing the known conformations, they are rarely capable of differentiating them. Overall, DFIRE2 and KORP produce the most consistent landscapes. Considering their extreme simplicity and speed compared to AMBER and ROSETTA, they constitute a very good choice for a first analysis of the sampled structures, to filter out very improbable conformations or to select an area to sample more thoroughly.

4 | DISCUSSION

The results presented in the previous section show that, both from the sampling and scoring points of view, the scaffold structures play a crucial role. Although loop modeling protocols usually include a refinement post-processing stage for a selected subset of the sampled loop states that locally adapts the conformation of the scaffold, some flexibility of the loop environment may already be necessary at the global sampling stage to guarantee an exhaustive exploration of the conformational space. For instance, if some surrounding side-chains are not modeled as flexible elements, they may prevent relevant alternative loop conformations from being sampled. Besides, even when the loop conformational space is properly covered during the sampling phase, the scaffold structure still greatly influences the predicted topography of the landscape, whichever scoring method is used. These observations underline the need of careful scaffold structure preparation for loop modeling. Specifically, due to the high sensitivity of the scoring functions to minor changes in the scaffold structures, the need to use several starting structures becomes clear (obtained e.g. by applying slight perturbations to a modeled or known conformation, or by gathering several known structures if those exist). From this perspective, employing fast sampling and scoring methods is appealing since it allows the exploration of several starting structure candidates using fewer computational resources. MoMA-LoopSampler was previously shown to provide exhaustive ensembles while excluding highly statistically improbable conformations, thus generating better-filtered ensembles. This makes MoMA-LoopSampler a suitable sampling method to use on multiple starting scaffolds. Fast statistical scoring methods then constitute a natural complement to such a loop modeling process. Note that the importance of the scaffold was already identified by Marks *et al.* in their analysis of flexible loops [4]. However, due to the coupling of the sampling and scoring steps in their work, the exact influence of the scaffold on each one of these two stages was not further analyzed.

Estimating the exact contribution of an accurate side-chain placement in flexible loop modeling is not a straightforward task. However, one can expect that correctly modeling side-chains improves the quality assessment performed by scoring methods employing all-atom models. For this purpose, we could have used side-chain prediction methods such as SCWRL4 [32]. Although SCWRL4 is a good and popular technique, it does not enforce strict rules concerning steric clashes, thus producing infeasible side-chain placements when the environment is very constrained. Given the importance of steric clashes and side-chain rotamers in most all-atom methods, and for consistency in sampling, we decided to implement our own side-chain placement method in MoMA-LoopSampler. It follows similar ideas to those applied in the backbone sampling phase. It uses the same model with the dihedral angles as sole degrees of freedom, forbids major collisions and allows for some deviations from the rotamers while still sampling around them. Due to steric constraints, this side-chain placement process has a relatively high failure rate and rejects many backbone sampled states. In addition, it returns the first placement that respects these constraints, without evaluating its quality. To improve results, we are currently working on a method integrating energy minimization for side-chain positioning. Building upon the results obtained in this work, DFIRE2 may be a good option to guide this minimization since it represents a good trade-off between accuracy and rapidity while considering an all-atom model.

Results presented in this paper do not allow to establish a clear correlation between the quality of the predictions

and the expected difficulty to identify “native" loop conformations due to experimental conditions: scoring methods perform well in some *a priori* difficult cases, whereas they fail to predict easier cases. Nevertheless, we believe that an accurate consideration of the loop environment, particularly regarding the presence of ligands or ions, is essential for a meaningful and complete analysis of the conformational energy landscape.

Unsurprisingly, statistical methods that do not require structural relaxation were found to be much faster than AMBER and ROSETTA. KORP and DFIRE2 are among the fastest methods while yielding remarkably satisfying results. They model smoother and more consistent landscapes than other methods. A major downfall is that they are rarely able to model landscapes with several basins, but the use of several starting scaffold structures may circumvent this shortcoming. KORP does not even consider the side-chains in the structures. Given that their placement is delicate and time-consuming, this is a considerable advantage over DFIRE2.

Concerning the energy landscapes, known conformations are often found adjacent to a high energy barrier. We hypothesize that these conformations are stabilized by a certain number of atomic contacts, within the loop or with surrounding residues. Areas of high energy would then correspond to conformations having these contacting atoms in steric collision. Note that using a 2D representation that only displays the energy of the best-scored state at a projected position prevents some energy barriers from appearing in the projected landscape. Therefore, areas projected in the middle of basins may still be forming numerous contacts, even though the energy barrier corresponding to bringing these contacting atoms closer together until they overlap is not apparent.

While all other statistics-based methods are built using a Bayesian framework on structural data from proteins with non redundant sequences, SBROD is built with a regression method and is trained to distinguish the native fold among several decoys. This design enables SBROD to perform well for *ab initio* structure prediction, as demonstrated in the latest CASP13 experiment (`www.predictioncenter.org/casp13/`), but yields disappointing results on the flexible loops studied in this work.

SOAP-Loop is the method that best agrees with other methods, including ROSETTA. It still takes a considerable amount of time to score the different states and does not provide overall results as satisfying as those of KORP. Thus, it is less suited to landscape reconstruction and to the analysis of numerous sampled states as performed in this work.

The landscapes modeled by AMBER and ROSETTA are, as expected, too rough to enable a satisfying analysis of a loop system. It is likely that the relaxations performed in this work were insufficient: for AMBER, the number of cycles may have been too low, or the convergence criterion too high; for ROSETTA, unrestricting the backbone dihedral angles of the rest of the protein may have yielded better results. However, longer relaxations would have been extremely costly in terms of computational resources. Although those scoring methods are inappropriate for the global/ensemble modeling of flexible loops, the results they show in this work suggest they may perform well for the refinement of stable structures. Statistical methods have their precision limited by design. AMBER and ROSETTA, conversely, may be used on a more limited number of states, e.g. to discriminate among similar models. Indeed, the strengths of different types of scoring methods can be exploited within multi-stage modeling approaches that first globally explore the conformational space using a fast and coarse scoring method, and then refine some regions using a more accurate one. It can also be interesting to combine several scoring functions within multi-objective optimization methods [33].

Although the present work only exploits experimental data obtained through X-ray crystallography, data obtained via nuclear magnetic resonance (NMR) could provide additional insight into the reliability of scoring methods on flexible loops. Comparison of experimental measurements from NMR such as residual dipolar couplings or chemical shifts with simulated measurements from sampled ensembles weighted using different scoring functions would constitute an alternative way to evaluate scoring methods. A major advantage of this approach is that it does not require the generation of all-atom models from NMR data.

5 | CONCLUSION

In this work, we have investigated the capability of state-of-the-art sampling and scoring methods to model flexible protein loops, which may adopt different (meta-)stable conformations. Our analysis shows that, despite the promising results obtained during both sampling and scoring steps, substantial methodological work is still required to identify the most probable conformations in an accurate and reliable manner.

To begin with, the success of loop sampling methods is limited by the difficulty to efficiently account for flexibility outside the loop and to correctly place side-chains. Indeed, whatever the methods employed for scoring multiple states, the structural scaffolds over which the loops were modeled proved decisive for the topography of the implicit landscapes. The integration of a flexible component in loop sampling methods thus constitutes an important direction for future work. Regarding side-chains, DFIRE2, that considers the position of all atoms and is among the fastest methods, could be used to optimize side-chain placement before scoring. Although structural relaxation is not needed in theory for this method, local optimization of side-chains generated by a global search strategy (as applied in this work) could improve the results.

Concerning loop scoring methods, some of them can reliably identify unfeasible states and are capable of providing valuable insight into the global topography of a loop's energy landscape. However, the modeled landscapes are often too fuzzy to allow a precise modeling of the loop's conformational space. In addition, most scoring methods provide erratic results from one loop to another, making their performance on a fully unknown system too unpredictable. In practice, such observations suggest that scoring methods can be reliably employed for applications requiring to coarsely filter loop states, but that their results are not accurate-enough for applications such as protein design. More precisely, the qualitative comparison of scoring methods for loop modeling presented in this paper validates the use of fast statistical potentials such as KORP or DFIRE2 as primary filters or as overall quality assessment methods for large pools of loop structures. Indeed, these methods can identify states close to statistically-probable conformations, regardless of poor local geometry or inner collisions. However, their low sensitivity to small conformational changes prevents them from providing a more precise evaluation. For such cases, physics-based or hybrid methods would be more appropriate, provided that the necessary structural relaxations are carefully performed.

data availability

The conformational ensembles used for our analysis are freely available at: `http://moma.laas.fr/static/data/loop-scoring_scaffolds_and_states.zip` For each of the 8 proteins used in our study, and for each of the scaffolds, we provide the structure of the entire protein chain and the set of loop states sampled by MoMA-LoopSampler. Each set of loop states is provided in a single PDB file, using the MODEL record type as separator. Sampled states before and after relaxation (using the AMBER and ROSETTA relaxation protocols explained in Section S4.3) are provided in different files.

acknowledgements

We thank P. Chacón, S. Grudinin, and Y. Li for help with obtaining and using KORP, SBROD and Torsions-only scoring methods, respectively, and P. Chacón and S. Grudinin for helpful discussion about the results. We thank K. Molloy for his help with relaxations using ROSETTA. The French National Association of Research and Technology (ANRT) is gratefully acknowledged for supporting A.B. (contract 2016/0239). This work used the HPC resources of the CALMIP supercomputing center under the allocation 2016-P16032.

conflict of interest

The authors declare no conflict of interest.

references

[1] Moult J, Fidelis K, Kryshtafovych A, Schwede T, Tramontano A. Critical assessment of methods of protein structure prediction (CASP)–Round XII. Proteins 2018;86(S1):7–15.

[2] Lensink MF, Velankar S, Wodak SJ. Modeling protein-protein and protein-peptide complexes: CAPRI 6th edition. Proteins 2017;85(3):359–377.

[3] Almagro JC, Teplyakov A, Luo J, Sweet RW, Kodangattil S, Hernandez-Guzman F, et al. Second antibody modeling assessment (AMA-II). Proteins 2014;82(8):1553–1562.

[4] Marks C, Shi J, Deane CM, Valencia A. Predicting loop conformational ensembles. Bioinformatics 2018;34(6):949–956.

[5] Fiser A, Do RKG, Šali A. Modeling of loops in protein structures. Protein Sci 2000;9(9):1753–1773.

[6] Jacobson MP, Pincus DL, Rapp CS, Day TJF, Honig B, Shaw DE, et al. A hierarchical approach to all-atom protein loop prediction. Proteins 2004;55(2):351–367.

[7] Soto CS, Fasnacht M, Zhu J, Forrest L, Honig B. Loop modeling: Sampling, filtering, and scoring. Proteins 2008;70(3):834–843.

[8] Mandell DJ, Coutsias EA, Kortemme T. Sub-angstrom accuracy in protein loop reconstruction by robotics-inspired conformational sampling. Nat Methods 2009;6(8):551–552.

[9] Lee J, Lee D, Park H, Coutsias EA, Seok C. Protein loop modeling by using fragment assembly and analytical loop closure. Proteins 2010;78(16):3428–3436.

[10] Zhao S, Zhu K, Li J, Friesner RA. Progress in super long loop prediction. Proteins 2011;79(10):2920–2935.

[11] López-Blanco JR, Canosa-Valls AJ, Li Y, Chacón P. RCD+: Fast loop modeling server. Nucleic Acids Res 2016;44(W1):W395–W400.

[12] Amaro RE, Baudry J, Chodera J, Demir O, McCammon JA, Miao Y, et al. Ensemble Docking in Drug Discovery. Biophys J 2018;114(10):2271–2278.

[13] Kundert K, Kortemme T. Computational design of structured loops for new protein functions. Biol Chem 2019;400(3):275–288.

[14] Jin W, Kambara O, Sasakawa H, Tamura A, Takada S. De Novo Design of Foldable Proteins with Smooth Folding Funnel: Automated Negative Design and Experimental Verification. Structure 2003;11(5):581–590.

[15] Hu X, Wang H, Ke H, Kuhlman B. High-resolution design of a protein loop. PNAS 2007;104(45):17668–17673.

[16] Koga N, Tatsumi-Koga R, Liu G, Xiao R, Acton TB, Montelione GT, et al. Principles for designing ideal protein structures. Nature 2012;491(7423):222–227.

[17] Shehu A, Kavraki LE. Modeling structures and motions of loops in protein molecules. Entropy 2012;14(12):252–290.

[18] Ponder JW, Case DA. Force fields for protein simulations. Adv Protein Chem 2003;66:27–85.

[19] Vanommeslaeghe K, Hatcher E, Acharya C, Kundu S, Zhong S, Shim J, et al. CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. J Comput Chem 2010;31(4):671–690.

[20] Maier JA, Martinez C, Kasavajhala K, Wickstrom L, Hauser KE, Simmerling C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. J Chem Theory Comput 2015;11(8):3696–3713.

[21] 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(7):1212–1219.

[22] Rata IA, Li Y, Jakobsson E. Backbone Statistical Potential from Local Sequence-Structure Interactions in Protein Loops. J Phys Chem B 2010;114(5):1859–1869.

[23] Dong GQ, Fan H, Schneidman-Duhovny D, Webb B, Sali A. Optimized atomic statistical potentials: assessment of protein interfaces and loops. Bioinformatics 2013;29(24):3158–3166.

[24] Karasikov M, Pagès G, Grudinin S. Smooth orientation-dependent scoring function for coarse-grained protein quality assessment. Bioinformatics 2018;.

[25] López-Blanco JR, Chacón P. KORP: knowledge-based 6D potential for fast protein and loop modeling. Bioinformatics 2019;.

[26] Barozet A, Molloy K, Vaisset M, Siméon T, Cortés J. A Reinforcement-Learning-Based Approach to Enhance Exhaustive Protein Loop Sampling. Bioinformatics 2020;36(4):1099–1106.

[27] Jones TA, Zou JY, Cowan SW, Kjeldgaard M. Improved methods for building protein models in electron density maps and the location of errors in these models. Acta Crystallographica Section A 1991 Mar;47(2):110–119.

[28] Kleywegt GJ, Harris MR, Zou Jy, Taylor TC, Wählby A, Jones TA. The Uppsala Electron-Density Server. Acta Crystallographica Section D 2004 Dec;60(12 Part 1):2240–2249.

[29] Alford RF, Leaver-Fay A, Jeliazkov JR, O'Meara MJ, DiMaio FP, Park H, et al. The Rosetta All-Atom Energy Function for Macromolecular Modeling and Design. J Chem Theory Comput 2017;13:3031–3048.

[30] Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The Protein Data Bank. Nucleic Acids Res 2000;28(1):235–242.

[31] Harder T, Boomsma W, Paluszewski M, Frellsen J, Johansson KE, Hamelryck T. Beyond Rotamers: a Generative, Probabilistic Model of Side Chains in Proteins. BMC Bioinf 2010;11(1):306.

[32] Krivov GG, Shapovalov MV, Dunbrack RL. Improved prediction of protein side-chain conformations with SCWRL4. Proteins 2009;77(4):778–795.

[33] Li Y, Rata I, Jakobsson E. Sampling multiple scoring functions can improve protein loop structure prediction accuracy. J Chem Inf Model 2011;51(7):1656–1666.

TABLE 1 Protein loops studied in this work.

#	System name	Loop residues	Loop length	Scaffolds[1] Conformation 1	Conformation 2	Conformation 3
1	**Streptavidin**	44-52	9	2F01(A), 2F01(B), 3RY1(A), 3RY2(A), 3RY2(B),	3RY1(B), 3RY1(D)	3RY1(C)
2	**MR-MLE**	115-125	11	3N4F(A),	3QPE(D), 3VCC(A)	
3	**NTPase**	41-50	10	4KFR(B)	4KFU(A)	
4	**Pot1pC**	109-118	10	4HID(A), 4HIK(A), 4HIM(A), 4HIO(A), 4HJ9(A)	4HJ7(A)	
5	**PTPN9**	466-477	11	2PA5(A), 4GE2(A), 4GE6(B)	4ICZ(A)	
6	**RNU2**	29-40	12	3AGN(A)	3AGO(A)	3AHW(A)
7	**TPI**	165-179	15	1YPI(A), 1YPI(B)	2YPI(A), 2YPI(B)	
8	**UTB**	66-76	11	3IRS(C)	3K4W(L)	

[1] List of PDB-IDs of the structures used as scaffolds for loop sampling, classified according to the conformation of the loop in the corresponding X-ray structure.

TABLE 2 Scoring methods compared in this work.

Method	Type	Relaxation	Side-chains
AMBER	Physics-based	Needed	Needed
ROSETTA	Hybrid	Needed	Needed
DFIRE2	Statistical	-	Needed
SOAP-Loop	Statistical	-	Needed
KORP	Statistical	-	-
SBROD	Statistical	-	-
Torsions-only	Statistical	-	-

FIGURE 1 Lowest RMSD to each known stable conformation among sampled states, for each scaffold. RMSDs are calculated on the heavy atoms of the backbone. Scaffolds are distributed around the disk and their names indicated outside the disk. The outer circle is colored according to the conformation present in the X-Ray structure the scaffold originated from. The outer circle corresponds to a RMSD of 0 Å, while the center corresponds to RMSDs of 4 Å and above: a larger colored area indicates that the corresponding stable conformation is found with greater accuracy in the sampled ensemble. The data are also provided in Table S6.

FIGURE 2 Running times per scored sampled state measured on three different systems. These were obtained on a single core of a 2GHz Intel® Xeon® processor. Note that the scale on the y-axis is logarithmic. RNU2, TPI and UTB scaffolds contain 114, 247, and 843 atoms each, respectively, which makes them representative of the different systems in terms of size.

FIGURE 3 Ranks of the five closest sampled states to each known conformation, from each scaffold. Scaffolds are distributed around the disk and separated by thicker dark grey lines. Their names have been omitted for clarity but they are arranged as in Figure 1. The outer circle is colored according to the conformation present in the X-Ray structure the scaffold originated from. Each line corresponds to a given known conformation. The ranks are represented by the proximity to the outer circle or to the center. The outer circle corresponds to the best scoring ranks, while the center corresponds to the worst rank. Note that the radial axis has a logarithmic scale. With such a representation, points far from the center of the circle correspond to well-scored sampled states.

FIGURE 4 Number of states close to known conformations among the top 1% states identified by the different scoring methods, from each scaffold. Distances correspond to backbone RMSD. The difficulty of each case *a priori* (estimated from the presence of a ligand, crystal contacts, or experimental conditions of the known crystal conformation), is indicated by the number of '*' in the scaffold name (* : easy, ** : intermediate, *** : hard).

FIGURE 5 Energy landscapes obtained by the different scoring methods. (a) MR-MLE loop, (b) PTPN9 loop, (c) RNU2 loop, (d) TPI loop. Each row correspond to a starting scaffold used for the conformational sampling and each column to a scoring method. The x- and y- axes correspond to the first and second principal components of the PCA run on the Cartesian coordinates of C_α atoms, respectively (see Section 2.4). The projected positions of the crystal conformations of the loops are indicated in the landscapes.

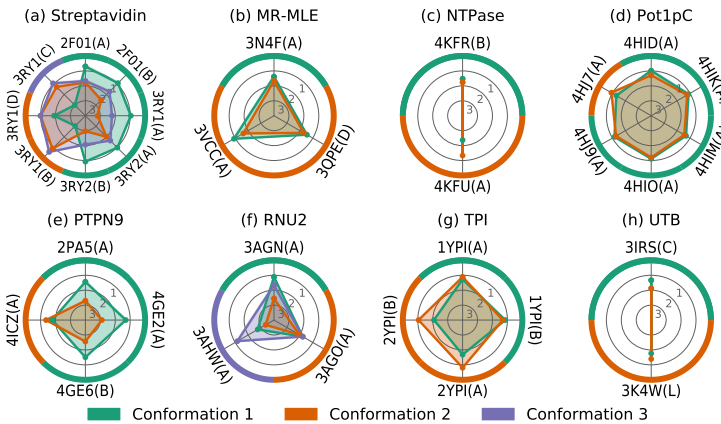

FIGURE 1

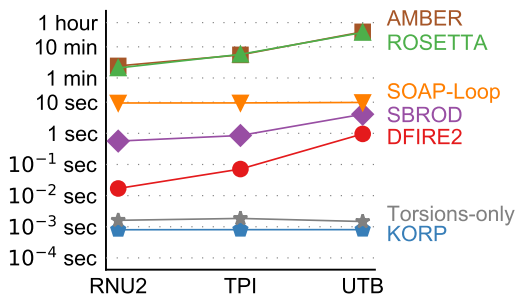

FIGURE 2

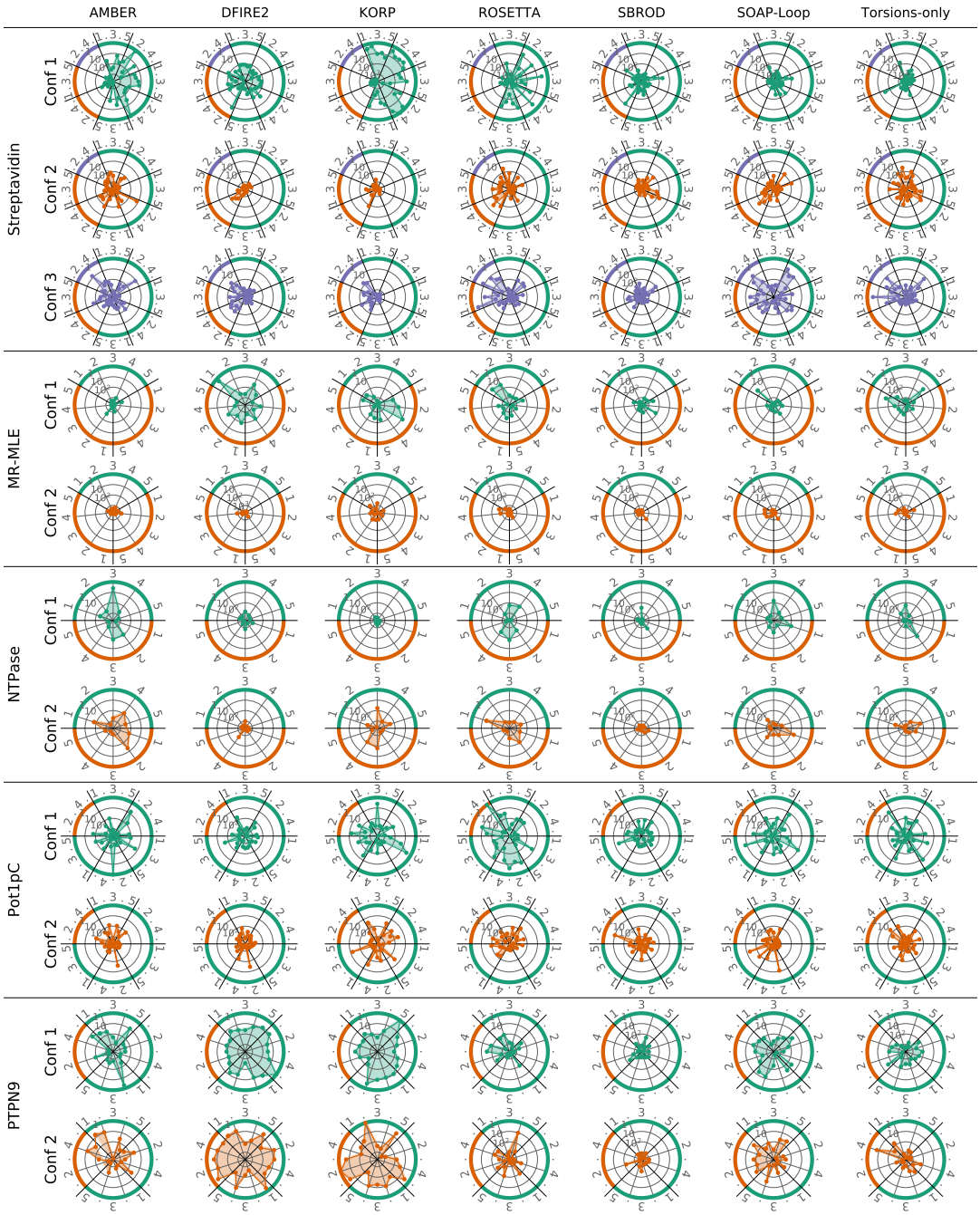

FIGURE 3 (First part)

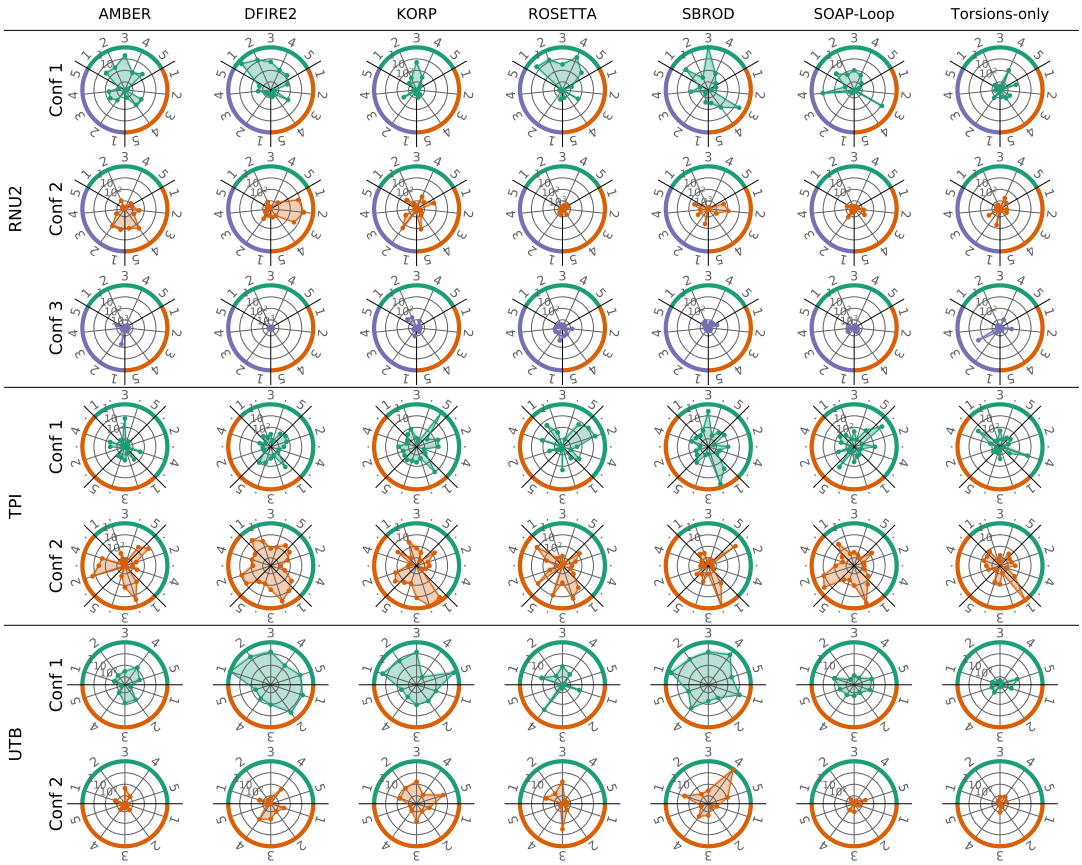

FIGURE 3 (Continued)


Streptavidin
MR-MLE
NTPase
Pot1pC
PTPN9
RNU2
TPI
UTB
AMBER
DFIRE2
KORP
ROSETTA
SBROD
SOAP-Loop
Torsions-only
States within 1.5 Å of several conformations
States within 1.5 Å of conformation 1
States within 1.5 Å of conformation 2
States within 1.5 Å of conformation 3
States within 2.5 Å of several conformations
States within 2.5 Å of conformation 1
States within 2.5 Å of conformation 2
States within 2.5 Å of conformation 3

FIGURE 4

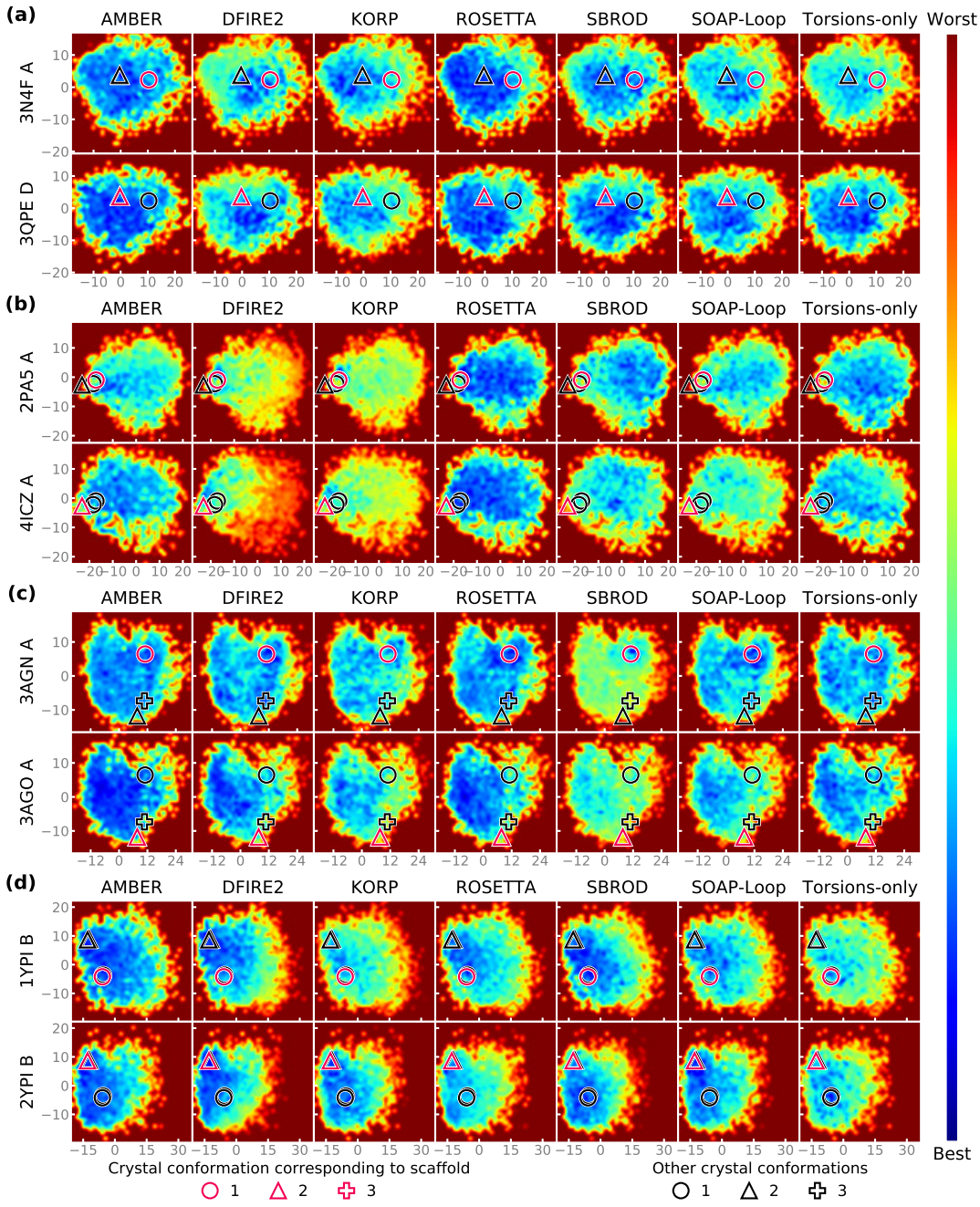
(a)
AMBER
DFIRE2
KORP
ROSETTA
SBROD
SOAP-Loop
Torsions-only
Worst
3N4F A
3QPE D
(b)
2PA5 A
4ICZ A
(c)
3AGN A
3AGO A
(d)
1YPI B
2YPI B
Best
Crystal conformation corresponding to scaffold
Other crystal conformations
1
2
3

FIGURE 5