

Subject Section

Predicting loop conformational ensembles

Claire Marks^{1,*}, Jiye Shi² and Charlotte M. Deane¹

¹Department of Statistics, University of Oxford, Oxford, OX1 3LB, United Kingdom

²Department of Chemistry, UCB Pharma, Slough, SL1 3WE, United Kingdom.

*To whom correspondence should be addressed.

Associate Editor: XXXXXXX

Received on XXXXX; revised on XXXXX; accepted on XXXXX

Abstract

Motivation: Protein function is often facilitated by the existence of multiple stable conformations. Structure prediction algorithms need to be able to model these different conformations accurately, and produce an ensemble of structures that represent a target's conformational diversity rather than just a single state. Here, we investigate whether current loop prediction algorithms are capable of this. We use the algorithms to predict the structures of loops with multiple experimentally-determined conformations, and the structures of loops with only one conformation, and assess their ability to generate and select decoys that are close to any, or all, of the observed structures.

Results: We find that while loops with only one known conformation are predicted well, conformationally diverse loops are modelled poorly, and in most cases the predictions returned by the methods do not resemble any of the known conformers. Our results contradict the often-held assumption that multiple native conformations will be present in the decoy set, making the production of accurate conformational ensembles impossible, and hence indicating that current methodologies are not well suited to prediction of conformationally diverse, often functionally important protein regions.

Contact: marks@stats.ox.ac.uk

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

For a protein to carry out its function effectively, a change in its structure is often necessary (Henzler-Wildman and Kern, 2007). These conformational changes can occur on a wide range of scales, from small changes in sidechain conformation (Ruvinsky *et al.*, 2011) to motions of whole domains (Qi and Hayway, 2009), and have been shown to be important for many diverse biological processes, such as enzyme catalysis (Gutteridge and Thornton, 2004), transport (Roberts *et al.*, 2013), signalling (Grant *et al.*, 2010), and molecular recognition (Armstrong *et al.*, 2008; Wang *et al.*, 2013). A classic example is haemoglobin, whose conformation changes upon binding oxygen such that its affinity for binding further oxygen molecules increases (Kavanaugh *et al.*, 2005). It has even been proposed that the diverse range of functions displayed by proteins is itself a result of conformational diversity - a protein that can perform different functions, because it is present in a number of different conformations *in vivo*, may be the starting point for the evolution of new proteins that become specialised in their specific functions (James and Tawfik, 2003).

As the number of experimentally-determined structures has grown, there are a number of proteins whose structures have been solved several times. From this data, it has become possible to infer and investigate conformational changes from analyses of multiple structures (e.g. Chothia and Lesk (1986); Kosloff and Kolodny (2008); Palopoli *et al.* (2016); Monzon *et al.* (2017)). Several databases have been created that link such structures together, and provide information about the experimental factors that may have led to the conformational differences (Echols *et al.*, 2003; Monzon *et al.*, 2013; Amemiya *et al.*, 2012; Juritz *et al.*, 2011; Hrabe *et al.*, 2016; Chang *et al.*, 2016). Molecular dynamics simulations are also often used to study flexibility - the calculated interactions between atoms are used to determine the atomic trajectories in an attempt to mimic the natural motions of the protein (e.g. Hummer *et al.* (2004); Benson and Daggett (2008)). Alternative means of investigating or predicting conformational changes include normal mode analysis (Dobbins *et al.*, 2008), statistical approaches (Pandey *et al.*, 2005), machine learning (Yaseen *et al.*, 2016), and network analysis (Jacobs *et al.*, 2001; Sarkar, 2017).

Even though the flexible nature of proteins is well known, the experimentally determined structures most commonly used by researchers can give a false impression of rigidity. Proteins can therefore often be treated as static objects whose structures do not vary. In the structure

prediction community, the standard approach is to create models that match the static X-ray structure (e.g. Moult *et al.* (1995); Teplyakov *et al.* (2014)), treating the X-ray determined structure as the ‘correct’ conformation, and measuring predictive accuracy based on that single conformation. However, there may be a set of conformations that a protein is able to assume, not just one. Models that are structurally distant from the designated correct structure, and hence thought to be inaccurate, may in fact be close to another native conformation of the protein.

In this paper, we investigate this issue in the context of loop structure prediction. The ability of loops to adopt different conformations is often vital for protein function (Gu *et al.*, 2015; Papaleo *et al.*, 2016). As is the case when predicting entire protein structures, in loop prediction the aim is usually to produce a single conformation. The quality of a prediction is assessed by calculating its distance to a single, experimentally-determined structure; the ability of some loops to adopt different conformations is not normally considered. However, the ability to predict all the possible conformations of a loop would be beneficial for many applications, such as the prediction of protein-protein interactions (Kuzu *et al.*, 2013), virtual screening of drug candidates (Osguthorpe *et al.*, 2012), and docking (Greenidge *et al.*, 2014), and would increase our understanding of protein function (e.g. ligand interactions and catalytic mechanisms). In addition, although differences in conformation may arise from the presence of binding partners, information about this binding, for example the position and orientation of ligands, is generally not known when modelling a novel structure. It is therefore also desirable for multiple conformations to be predicted without any binding partners being present.

In this study, we identify loops for which there are multiple structures deposited in the PDB, and investigate the ability of several loop modelling algorithms to produce accurate loop conformations for these targets. We also find a set of loops for which only one conformation has been observed, and compare the prediction accuracy achieved for the two sets. By taking into account all conformations that have been observed experimentally when assessing the accuracy of prediction, and not just a single conformation, we are able to determine the suitability of current loop modelling software for gaining insights into the functions of conformationally variable loops.

2 Methods

2.1 Target Selection

Two main target sets are used in this research; one containing loops with multiple known conformations, and another comprising loops that have been observed multiple times, but with only a single conformation. We do not consider loops whose structures have been solved only once. Using PISCES (Wang and Dunbrack, 2003), we extracted from the PDB (Berman *et al.*, 2000) all structures solved using X-ray crystallography, with resolution below 2 Å, and placed them into ensembles of identical sequence. Any proteins that did not have any sequence-identical matches were ignored. Only a single copy of a protein was used from any PDB entry. Loop regions were then identified using DSSP (Joosten *et al.*, 2011). Here and throughout this paper, we consider loops to be sections of protein that connect helices or strands (assigned the labels H, I, G and E by DSSP), where the helices or strands are at least three residues in length. If the exact definitions of loop regions varied between sequence-identical structures, we extended the loop definition to include all possible loop residues. For example, if the loop of one structure in a set is defined to cover residues 10 to 16, but DSSP places the same loop in the other member of the ensemble between residues 11 to 17, our definition of the loop is residues 10 to 17.

The structural variation displayed by our loop ensembles was quantified by calculating the backbone RMSD between every pair of loops

in an ensemble after superimposing the backbone atoms of the anchor residues (the two residues on either side of the loop). We use the maximum pairwise RMSD of a loop ensemble to measure structural variation. We extracted two groups of possible targets: one containing loops displaying high levels of conformational variation (maximum pairwise RMSD over 2 Å), and one with structurally invariable loops (with a maximum pairwise RMSD below 0.3 Å; this is a strict cutoff that allows very little variation between structures, meaning the loops in each ensemble are unambiguously of the same conformation). From these two groups, we selected five loop ensembles at random for each loop length between 10 and 15 residues, inclusive. Each target set therefore contains 30 loops, and each has the same distribution of loop lengths. For brevity, we refer to the two target sets as the ‘high variation set’ and the ‘low variation set’; full lists of PDB targets for each loop are given in the Supplementary Information (SI).

The loops in the individual PDB entries for the high variation set were grouped into structurally distinct subgroups (which we refer to as conformations), based on clusters formed using the UPGMA algorithm (with a cutoff of 1.5 Å, as used by Nowak *et al.* (2016)). This gave clusters in which the maximum pairwise RMSD between members was always less than 1 Å (below this value we consider loops to have the same conformation), but was often lower (the average maximum pairwise RMSD across all clusters was 0.37 Å). The number of conformations present for a loop ranges from two to four; in total there are 66 conformations across the 30 target loops.

We also use a third set of 20 target loops (the ‘inflexible loop set’) which have been identified as highly inflexible, based on the observation that they move only minimally during molecular dynamics simulations (their C α atoms moved less than 0.5 Å) (Benson and Daggett, 2008). One loop (PDB entry 1G61, residues 12-19) was omitted from the original set (as reported in (Benson and Daggett, 2008)), because the residue numbers reported do not correspond to a loop in the protein structure.

2.2 Loop Structure Prediction

Four loop structure prediction algorithms were used: Sphinx, FREAD, Rosetta, and LEAP. These algorithms use a variety of approaches to generate loop conformations. FREAD (Deane and Blundell, 2001; Choi and Deane, 2011) is a knowledge-based method that uses a database of experimentally-determined protein structures to predict loop conformations. LEAP (Liang *et al.*, 2014) and the loop modelling program within the Rosetta software (Stein and Kortemme, 2013) are *ab initio* methods, that generate loop structures computationally, without knowledge of previously observed conformations. Sphinx (Marks *et al.*, 2017) is a hybrid method, that uses known protein fragments of a different length to the target loop to model part of the structure (according to a sequence alignment), then completes it using *ab initio* approaches, producing decoys of the correct length.

For the 30 loop ensembles in each of the ‘high’ and ‘low’ variation sets, all four prediction algorithms were used to predict the loop structures. We used every PDB entry associated with each loop ensemble as a ‘scaffold’ on which decoys were built, meaning prediction was performed several times for each loop ensemble in the target sets. There were 3 scaffolds for which Rosetta produced an error, meaning no predictions were made using them: 2RBXX and 2YJDB from high variation set loops 26 and 28, and 1N5NA from low variation loop 1. All chains other than the one on which the target loop is found and any ligands were removed since it is unrealistic to assume the relative positions of these entities would be known in a true modelling situation.

For FREAD and Sphinx, the database used was one containing the loop structures of all PDB entries with non-identical sequences (extracted using PISCES (Wang and Dunbrack, 2003) and using the highest-resolution

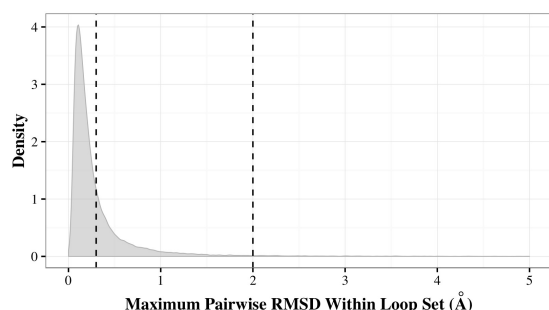


Fig. 1. The amount of structural variation observed for protein loops. The distribution shown is that of the maximum pairwise RMSDs calculated for ensembles of loop structures from sequence-identical proteins. The dashed vertical lines split the loop ensembles into three groups: low variation (RMSDs of 0 - 0.3 Å), medium variation (0.3 - 2 Å), and high variation (2 Å or above). Most loop ensembles demonstrate only small structural differences - there are 47,034 loops with low variation, 17,635 with medium variation, and 1028 with high variation.

structure for each sequence), split into all possible 3- to 30-residue fragments. Fragments from proteins with an identical sequence to the target protein were ignored. FREAD was run using the default parameters; Sphinx was implemented as described in (Marks *et al.*, 2017), but with the minimisation step omitted.

Using Rosetta, 500 decoys were generated for each target (as in (Stein and Kortemme, 2013); see SI for details of commands used). For LEAP (as in (Liang *et al.*, 2014)) we ran the software ten times for each target to give ten different predictions, which we then ranked using their energies as calculated internally by the algorithm.

3 Results

Structural Variation of Protein Loops

In January 2016, there were 87,180 protein structures in the PDB that were solved using X-ray crystallography, with a resolution of up to 2 Å. We extracted the 5,548 unique sequences whose structures had been solved multiple times, and from these we identified 65,697 loops. We refer to the collection of structures available for a loop as a loop ensemble (see Methods for details); on average there are 3.5 structures per ensemble. The majority of the identified loops are short (e.g. 11,656 are 4-residue loops), however there is still data available for longer loops (there are at least 180 loop ensembles for all lengths up to 25 residues).

A plot showing the structural variation observed is shown in Figure 1. We evaluate this structural variation using the maximum backbone RMSD calculated between loops within each ensemble. Most loops show very little variation between structures, with the majority having a maximum pairwise RMSD of between 0 and 0.3 Å. This indicates that loops are generally not flexible. Only 1.6% of loop ensembles show variation in excess of 2 Å (1,028 ensembles). Conformational variability increases with loop length, with longer loops having, in general, larger maximum pairwise RMSDs. For example, the proportion of loops demonstrating high levels of structural variation (over 2 Å) is 0.54%, 2.34%, 2.72%, 3.31% and 3.85% for 4-, 8-, 12-, 16- and 20-residue loops respectively (see SI); however, these proportions are still low even for long loops.

From the identified ensembles, we selected two target sets for use in the following investigations; one containing loops with high variation (having a maximum pairwise RMSD over 2 Å) and another with loops of low variation (maximum pairwise RMSD below 0.3 Å).

3.1 Loop Structure Prediction

We used four loop prediction algorithms to model the loops of the high and low variation target sets: Sphinx (Marks *et al.*, 2017), FREAD (Deane and Blundell, 2001; Choi and Deane, 2011), LEAP (Liang *et al.*, 2014) and Rosetta (Stein and Kortemme, 2013). The four algorithms encompass different methodologies; FREAD is a knowledge-based method that uses fragments of known structure as candidate loop conformations (decoys), Rosetta and LEAP are *ab initio* methods that produce decoys from scratch, and Sphinx is a hybrid method that combines aspects of both knowledge-based and *ab initio* approaches. We performed prediction using every protein chain associated with each loop ensemble as a basis on which to build decoys; we refer to these structures as scaffolds.

A summary of the results produced by the algorithms for the high and low variation sets are shown in Figure 2 (see SI for full results for each scaffold). For each decoy generated, we calculated the backbone RMSD compared to all structures in that loop ensemble. The ability of the prediction algorithms to predict a loop conformation is assessed using the RMSD of the best (lowest-RMSD) decoy generated, the RMSD of the top-ranked prediction, and the RMSD of the best decoy in the top five. In Figure 2, we report Sphinx's performance using the best results achieved across all the scaffolds for a given loop ensemble. While not realistic, since in a real modelling situation the true RMSD would not be known and generally only one scaffold would be used, in this way we consider the potential (best-case) accuracy of the methods.

Intuitively, loops with different conformations would be considered easier to predict, since there are multiple correct answers. However, our results suggest that accurately predicting loops with high conformational variation is more difficult than predicting loops that only adopt a single structure. For Sphinx, the average RMSD of the best decoy generated across all 66 conformations in the high variation set is 1.77 Å (median

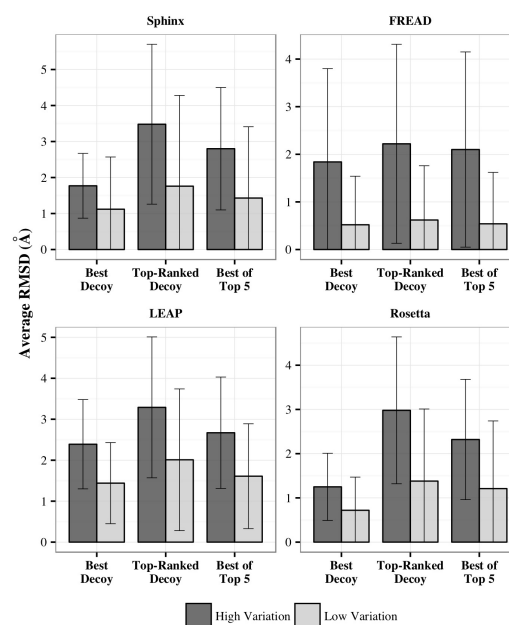


Fig. 2. Summary of prediction results for the high and low variation target sets. For each prediction method (Sphinx, FREAD, LEAP, and Rosetta), we present the RMSD of the best decoy generated, the RMSD of the top-ranked decoy, and the RMSD of the best decoy in the top five, averaged over all conformations in each target set. Error bars show the standard deviation. Structure prediction was carried out using all PDB entries of identical sequence as scaffolds - the results displayed here represent the best achieved across all runs. Full results for each individual scaffold are given in the SI. Loops of the high variation set are predicted more poorly than those of the low variation set.

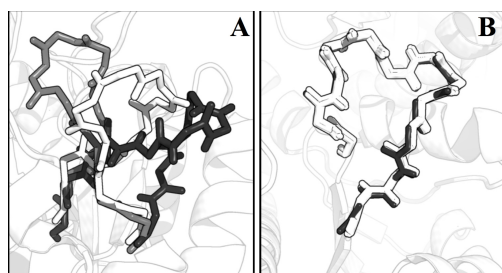


Fig. 3. Examples of predictions produced by Sphinx, for one loop of the high variation set (loop 15) and one of the low variation set (loop 10). For the high variation case (panel A), two conformations are observed (PDB entries 3IRSC and 3K4WL, shown in grey and white respectively). The best top-ranked decoy in this case (black) is not close to either conformation, with an RMSD of 6.99 Å to the loop from structure 3IRSC, and 5.28 Å to that from 3K4WL. For the low variation example (B), all five of the experimental structures display the same conformation (PDB entries 3TANA, 3TAPA, 3TAQA, 3TARA and 4DQSA; shown in white). The best top-ranked decoy produced by Sphinx (black) matches this conformation closely, with an RMSD of 0.23 Å.

1.92 Å); for low variation loops it is 1.12 Å (median 0.28 Å). In addition, for twenty of the loop targets with low variation, a decoy exists that has an RMSD to an experimentally observed conformation of less than 0.5 Å. This is only true for eight of the high variation loops, and never occurs for more than one conformation in a loop ensemble.

When the final ranking step is performed, the difference in accuracy between the two sets becomes even more stark. An example prediction from each of the high and low variation sets is shown in Figure 3. The average RMSD achieved by Sphinx for the top-ranked decoy (for all conformations) is 3.48 Å (median of 3.43 Å) for high variation loops and 1.76 Å (median 0.42 Å) for low variation loops. When only considering the best-predicted conformation for each loop, high variation loops are still modelled more poorly, with an average RMSD of 2.38 Å. For loops with multiple conformations, Sphinx could produce a prediction within 1 Å of a native conformation in only ten cases, while for low variation loops twenty sub-angstrom predictions were made. Considering the best of the five top-ranked decoys, the average RMSDs for high and low variation loops are 2.80 Å and 1.43 Å respectively.

For highly variable loops, on occasion the decoys selected by the ranking system are close to one of the conformations (for example loop ensemble 26, where conformation 1 is predicted with sub-angstrom accuracy), however the majority of decoys selected are not close to any experimentally observed conformation. This implies that it would be very difficult, if not impossible, to use Sphinx to produce a small ensemble of structures that represent the different possible conformations of a loop.

We examined the RMSD distributions at various stages of the Sphinx prediction algorithm in order to pinpoint where the difference in performance arises. Figure 4 shows the RMSD distribution of the decoy set generated by Sphinx at three different stages: after all decoys have been generated, after the selection of the top 500 decoys by the knowledge-based scoring function, and after the top 5 decoys have been selected using the SOAP-Loop potential (Dong *et al.*, 2013). RMSDs shown are the RMSDs of each decoy to its nearest experimentally-observed conformation. In each case, there are more near-native decoys present for the low variation loops. However, the difference between the two increases at each stage. This indicates that while it may be easier to generate near-native decoys for structurally invariable loops, the ranking step appears to be amplifying this effect.

To investigate this further, we calculated the Pearson correlation coefficient between the SOAP-Loop score and the RMSD to the nearest conformation, for the sets of 500 decoys produced by Sphinx. The average correlation coefficient observed for the low variation set was 0.54; for the

high variation loops this value was 0.29 (and in some cases was negative). SOAP-Loop, the scoring function used by Sphinx, is therefore better able to score loops that have only one conformation, and is less accurate when the loop has the ability to adopt multiple conformations.

For one of our high variation loop targets (loop 30), we artificially generated some near-native decoys by allowing Sphinx to use fragments from the native structures. We randomly selected 5 decoys that were close to each conformation (with RMSDs ranging between 0.5 and 1 Å), and introduced them to the decoy set - after ranking, the top five decoys were all from this near-native set, suggesting accurate decoys for both conformations would have been selected had we been able to generate them. However, some of our near-native decoys were not ranked accurately (sometimes outside the top 500). This also occurred for some of the high variation targets (e.g. loops 7 and 12), but never for the low variation set; for these targets if a high accuracy decoy was generated, then the top-ranked prediction was always very accurate. The low accuracy predictions for the high variation set therefore probably stem from a combination of inability to produce near-native decoys and poor scoring accuracy.

FREAD produces a similar result to Sphinx. The coverage (the percentage of loops for which FREAD was able to find appropriate fragments) for the high variation set is lower than for the low variation set (63.3% compared to 86.7%). The average RMSD of the best decoy across all conformations is 1.84 Å for the high variation set and 0.52 Å for the low variation set. After ranking the fragments that FREAD selects from the database by their anchor RMSD, the average RMSD of the top prediction is 2.22 Å for the conformations of the high variation set, and 0.62 Å for the low variation loops. In general, where FREAD is able to make a prediction, one conformation is predicted with higher accuracy than the others - only considering the best-predicted conformation from each ensemble, the average RMSD of the top-ranked decoy is 0.77 Å. The best of the top five decoys has an average RMSD of 2.10 Å and 0.54 Å for the high and low variation sets respectively. If only considering the best-predicted conformation for each loop ensemble in the high variation set, the average RMSD of the best in the top five is reduced to 0.63 Å. The low coverage achieved for the high variation set may effect the reported results - FREAD is designed to only return a prediction if there is an appropriate fragment in the database, making the RMSDs given here lower than they would be if FREAD was forced to predict in every case. This is likely to affect the high variation set more than the low variation set, since coverage for this set was lower.

The results achieved by LEAP and Rosetta also show a similar trend. For LEAP, the best decoy produced for each conformation has an average

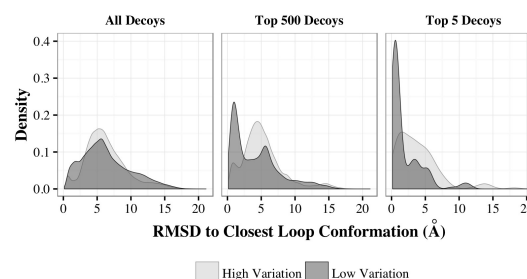


Fig. 4. RMSD distributions at different stages of the Sphinx algorithm. The stages shown are after all decoys have been generated, after our own knowledge-based scoring function has been used to select the top 500, and after the top 5 decoys are selected by SOAP-Loop. RMSDs used are for the decoys generated using one randomly-selected scaffold from each loop in the high and low variation sets (we observed very little difference in Sphinx's performance when run on all PDB entries associated with a loop). At each decoy selection stage, the difference between the RMSD distributions becomes larger, with the decoy sets for low variation loops containing more near-native structures.

RMSD of 1.44 Å for loops with only one conformation, and 2.39 Å for those that display conformational variability. Decoys were made with sub-angstrom accuracy for 14 loops in the low variation set, but only five in the high variation set. When ranked according to their calculated energies, the top-ranked decoy had an average RMSD of 2.01 Å for the low variation set and 3.29 Å for the high variation set, and the best of the top five had an average RMSD of 1.61 Å and 2.67 Å for the low and high variation sets respectively. In the case of Rosetta, the best decoy produced for each conformation had an average RMSD of 0.72 Å for the low variation set and 1.25 Å for the high variation set. For 18 loops in the high variation set, a decoy with an RMSD below 1 Å (to any conformation) was generated, whereas decoys with sub-angstrom accuracy were created for 25 of the low variation targets (and 18 of these had an RMSD below 0.5 Å, which was only true for 8 loops in the high variation set). After ranking, the pattern remains: the top prediction has an average RMSD of 2.98 Å and 1.38 Å, and an average best-of-top-five RMSD of 2.32 Å and 1.21 Å, for the high and low variation sets respectively.

We selected a single PDB scaffold for each target loop, and calculated the RMSDs between all pairs of decoys generated using that scaffold, to see how the decoys generated for each target set are spread throughout the conformational space. This indicates how structurally alike the decoys are to one another. For each method, the decoys generated for the loops with only one conformation show less structural variability than those targets with multiple conformations (however the trend is more extreme for the decoys obtained using Sphinx - see SI). This trend could mean that for low variation targets, loop prediction algorithms are able to explore the conformational space more efficiently. We have also used Sphinx to predict a set of disordered loops (where multiple copies of the same loop had at least half of their residues missing in the PDB structures) - they also follow this pattern, with the decoys displaying even more dissimilarity than those of the high variation set (see SI for more details).

We also investigated the effect of the scaffold on to which the loops were modelled - other regions of the protein may differ between conformation, and hence have an impact on prediction accuracy. For each loop conformation in the high variation set, we looked at the results achieved when using a scaffold structure that, in its native form, has the same loop conformation as the conformation in question, and compared these results to those obtained using a scaffold holding a different loop conformation (Figure 5). We found that, especially for FREAD and Rosetta, accuracy is lower when using a scaffold which holds the 'incorrect' loop conformation. This has implications for the overall objective of producing a set of structures that describes the conformational diversity of a loop - some conformations of a loop may be inaccessible using a particular scaffold, further indicating that predicting multiple conformations on to a single scaffold to form a conformational ensemble is currently impossible.

In addition to structure prediction, we investigated whether molecular dynamics simulations are capable of replicating the conformational diversity observed in crystal structures. We observed that MD was not able to reproduce the experimental loop conformations, however the amount of loop motion seen during the simulation seems to reflect the true diversity (i.e. loops with multiple conformations were seen to move more - see SI for more details).

3.2 'Inflexible Loop' Target Set

In order to test whether our results could be replicated on another set of loops, we used Sphinx to predict the structures of the 'inflexible loop' set (see Methods). These loops were identified as structurally invariable based on a series of molecular dynamics simulations (Benson and Daggett, 2008). In all but three cases, a decoy with sub-angstrom RMSD was generated; the average RMSD of the best decoy in the top 500 was 0.78 Å. After

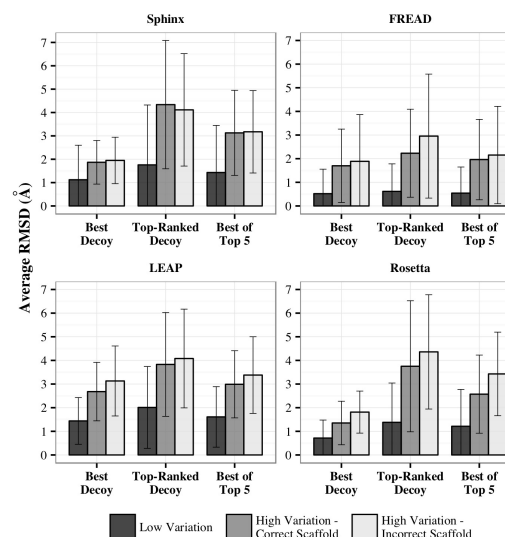


Fig. 5. The effect of scaffold on prediction accuracy. Here we show prediction accuracy when modelling onto a scaffold that (in its native form) has the same conformation as the target conformation, compared to the when a scaffold with a different conformation is used. Results shown are the RMSD of the best decoy, top-ranked decoy and best of top five decoys, averaged over all conformations, for each of the tested prediction algorithms (Sphinx, FREAD, Rosetta and LEAP). Error bars represent the standard deviation for these values. For comparison, the results from the low variation set are included (in this case, all scaffolds contain the same conformation). Prediction accuracy tends to be lower when using a scaffold of the 'incorrect' conformation.

ranking, the RMSD of the top prediction is also low at 1.15 Å. Full results are for each target are given in the SI. These findings are consistent with the results achieved previously; loops with low flexibility are generally modelled with high accuracy.

3.3 Are there any structural differences between high and low variation loops?

We looked for differences between high and low variation loops that may cause the observed disparity in prediction accuracy. If loops that display high variation are unusual in some way, this may lead to difficulties in generating and accurately scoring near-native decoys. To examine this hypothesis, we compared several features of the loops: the quality of the experimental structure (i.e. B factors and R-free values), amino acid frequencies, bond lengths, bond angles, ϕ/ψ distributions, ω angles, anchor geometry, and the presence of contacts. For most of the characteristics examined, the differences between the high and low variation sets were very small and hence these features are unlikely to be the cause of the observed accuracy differences (see SI for more details).

We observed that high variation loops tend to have smaller anchor separations (the distance between the C α atoms of the residues before and after the loop) than low variation loops (see SI). Intuitively, anchor residues that are far apart should reduce the number of possible conformations, more so than loops whose anchors are closer together (Choi *et al.*, 2013). This may affect the conformational search undertaken by loop modelling algorithms - a smaller number of possible structures limits the conformational space that must be searched, making the prediction of low variation loops easier. However, when we looked at members of the high and low variation target sets with similar anchor distances, the difference in prediction accuracy remained - the difference cannot therefore be attributed solely to anchor separation.

We also investigated the number of atoms in each protein that are in close proximity to the loop residues in each set. Loops with one observed

conformation tend to have a greater number of atoms in their vicinity than the loops that show high conformational variability (see SI). This could affect the ability of the scoring functions to select the best decoys. For example, the knowledge-based scoring function within Sphinx only considers atoms that are up to 8 Å away from the target loop - our tests indicate that high variation loops tend to have fewer neighbouring atoms within this range, and therefore less data is used in the score calculation. This may lead to greater ranking inaccuracies in these loops.

In addition to forming more contacts between atoms of the same chain, more loops of the high variation set have other molecules in their vicinity, such as small molecules, metal ions, DNA and other protein chains - 26 of the high variation loops have atoms of such molecules within 6 Å of the loop in at least one member of the ensemble, whereas this is true for only 19 of the low variation loops. Only the protein chain on which the target loop was located was used during prediction, with all other protein chains, ligands etc. removed, since in a realistic modelling situation their positions would not be known - however, since the presence of ligands may be a contributing factor to the observed changes in conformation, their absence may affect the accuracy of prediction. When divided into groups based on binding (a target was considered bound if at least one member of the loop ensemble had a binding partner within 6 Å), the low variation loops were predicted with an RMSD of 0.68 Å for the unbound group (11 target loop ensembles) and 2.38 Å for the bound group (19 targets). The absence of binding partners during prediction therefore seems to negatively affect model accuracy. However, this trend was not observed for high variation loops - the top-ranked decoy, considering the best-predicted conformation only, had an RMSD of 3.88 Å for the unbound set (4 targets) and 3.43 Å for the bound set (26 targets). Therefore accuracy is still poor for loops where the change in conformation is not a result of binding.

3.4 Prediction of Conformational Variation

We demonstrated previously that the conformational space explored by the loop modelling algorithms, represented by the distribution of RMSDs between decoys, is different for low variation loops compared to high variation loops. Furthermore, since these distributions can be obtained without prior knowledge of the loop structure, they may be used to predict the conformational diversity of the loop. This approach to flexibility prediction would offer an advantage over molecular dynamics simulations, in that loop prediction takes less time than MD (on average, Sphinx took 2 hours to produce a prediction, while the MD simulations carried out took several days - see SI). To test this hypothesis, we compared the distributions from each individual loop to two overall distributions, representing the high and low variation sets. These distributions were created by combining the distributions for all loops in each set, except the particular target being considered, using the Sphinx decoy distributions described earlier. The Kolmogorov-Smirnov test was used to establish which of the 'high' or 'low' distributions was closest to each loop's independent set of RMSDs, and thereby predict the conformational diversity of the desired loop. For 25 (83%) of the 30 high variation loops, their variation was predicted correctly (i.e. the closest RMSD distribution was that of the high variation targets). Of the low variation loops, 18 (60%) were predicted correctly. The structure prediction protocol itself may therefore be used to predict whether a given loop is capable of adopting multiple conformations, or conversely, whether it is structurally rigid. This knowledge would be useful for many applications, for example when carrying out procedures such as docking, since it would indicate where the incorporation of flexibility would or would not be suitable. In addition, given our earlier result that the structures of loops with multiple conformations are predicted poorly, the distribution of pairwise decoy RMSDs could also be used to estimate prediction accuracy, hence giving a measure of confidence in a model.

4 Conclusion

We have described our investigations into predicting the structures of loops with multiple experimentally-observed conformations. We identified a large number of loops for which multiple high-resolution X-ray structures exist in the PDB, and found that in the majority of cases there is very little variation between them. The RMSD between different structures of the same loop is normally approximately 0.1-0.2 Å. This would indicate that most loops may be relatively static, contrary to the popular view that they display large amounts of flexibility.

When modelling a loop capable of adopting multiple conformations, instead of producing just one prediction it would be more accurate to produce an ensemble of structures. However, we have shown that the accuracy of prediction is markedly lower for variable loops than it is for conformationally stable ones, indicating that this is impossible with current methodologies. We have found that this trend appears with several loop prediction algorithms, of knowledge-based, *ab initio* and hybrid types (FREAD, LEAP, Rosetta and Sphinx). Differences can be observed at all stages of loop modelling, but the effect is larger after decoy ranking. We found that the choice of scaffold on which to model the loop has an effect on accuracy; predictions for a given conformation are less accurate using a scaffold that does not have the same conformation, when compared to one that does. The production of an ensemble of conformations using a single scaffold is therefore not currently achievable.

We examined several properties of the loops to try and identify any differences between loops of single and multiple conformations that may cause this discrepancy. Only very small differences are seen in the local geometry of the loop structures (e.g. ϕ , ψ and ω angles); the main differences are seen in the distance between their anchors, and the proximity of the loop to other residues in the protein. The anchor residues of single-conformation loops are generally further apart than loops with multiple conformations. Since the number of possible conformations is greater for loops with smaller anchor distances (Choi *et al.*, 2013), this may affect the ability of structure prediction algorithms to search the conformational space effectively.

Loops with only one observed conformation form more contacts to the surrounding protein, meaning ranking functions, which commonly use interatomic distances to calculate scores, have more data available to them. In addition, high variation loops tend to have more binding partners situated close to them (e.g. small molecules or other protein chains). The observed changes in conformation may be a result of these binding interactions, however binding partners were removed during prediction since their positions would realistically be unknown. Hence, we wish to be able to produce conformational ensembles without this information. This may be another contributing factor to the observed difference in accuracy - for low variation targets, there is a clear contrast in prediction quality for targets that contain binding partners in their crystal structures compared to those that do not, with predictions for bound targets being on average 1.7 Å less accurate. However, for high variation loops, this trend was not observed; accuracy is poor even for loops whose change in conformation is not caused by a binding event.

Differences in conformation may also be a result of crystal packing. When a protein structure has been determined by crystallography multiple times, there is a reasonable chance that the loops have experienced a variety of forces coming from different types of crystal packing. Some examples in both the high and low variation sets have crystal contacts; some have none (see SI).

Both the binding partner and crystal packing observations suggest that loops belonging to the low variation set have a strong intrinsic tendency to maintain their conformations despite the changing environment (i.e., different packing), and that conformation is likely to be an energy minimum. Conversely, loops in the high variation set are expected to

be more susceptible to perturbations coming from the environment, and may be less stable than those in the low variation set. Since information about binding partners and crystal packing would not be known without a structure, these factors cannot be taken into account during modelling.

We find that the conformational space explored by a loop prediction algorithm, quantified by the distribution of RMSDs between decoys, is different for loops of multiple conformations compared to loops with only one conformation. High variation loops tend to have larger pairwise decoy RMSDs than their low variation counterparts. We have shown how this property can be used to predict the native ‘flexibility’ of a target loop, by comparing its pairwise RMSD distribution to those of loops of known variation. The incorporation of other features into a flexibility prediction protocol, such as the number of contacts present and the distance between anchor residues, may improve results.

One caveat to consider when interpreting these results is that it cannot definitively be proven using our methodology that loops truly exist in only one conformation. Even though we look at loops with multiple structures available, it may be the case that a loop is able to adopt different conformations, but only one has been observed so far experimentally. As a consequence, some of the loops in our ‘low variation’ set may actually be conformationally variable.

Our results have major implications for the field of protein structure prediction. It has been often assumed that multiple native structures for a conformationally diverse target will be present in the generated decoy sets, leading to the formation of ensembles that accurately reflect the true nature of the target. We have shown this to be false. Existing methods may in fact be incapable of predicting the structures of such targets, possibly because of our fundamental approach to their development - methods are evaluated based on their ability to replicate single experimental structures, and hence we may be inadvertently training algorithms only to do this, making them unable to predict structural ensembles. Instead, prediction methods could be evaluated by comparing their results to multiple known conformations, to increase our accuracy when predicting the structures of these often functionally important targets.

Funding

This work was supported by the Engineering and Physical Sciences Research Council (grant reference EP/G037280/1) and UCB Pharma Ltd.

References

- Amemiya, T. *et al.* (2012) PSCDB: a database for protein structural change upon ligand binding. *Nucleic Acids Res.*, **40**, D554-D558.
- Armstrong, K. M. *et al.* (2008) Conformational changes and flexibility in T-cell receptor recognition of peptide-MHC complexes. *Biochem J.*, **415**, 183-196.
- Benson, N. C. and Daggett, V. (2008) Dynameomics: Large-scale assessment of native protein flexibility. *Protein Sci.*, **17**, 2038-2050.
- Berman, H. M. *et al.* (2000) The Protein Data Bank. *Nucleic Acids Res.*, **28**, 235-242.
- Chang, C. W. *et al.* (2016) CCProf: exploring conformational change profile of proteins. *Database*, baw029.
- Choi, Y. and Deane, C. M. (2011) Predicting antibody complementarity determining region structures without classification. *Mol. Biosyst.*, **7**, 3327-34.
- Choi, Y. *et al.* (2013) How long is a piece of loop? *PeerJ.*, **1**, 1-15.
- Chothia, C. and Lesk, A. M. (1986) The relation between the divergence of sequence and structure in proteins. *EMBO J.*, **5**, 823-826.
- Deane, C. M. and Blundell, T. L. (2001) CODA: A combined algorithm for predicting the structurally variable regions of protein models. *Protein Sci.*, **10**, 599-612.
- Dobbins, S. E. *et al.* (2008) Insights into protein flexibility: the relationship between normal modes and conformational change upon protein-protein docking. *Proc. Natl. Acad. Sci. USA*, **105**, 10390-10395.
- Dong, G. Q. *et al.* (2013) Optimized atomic statistical potentials: assessment of protein interfaces and loops. *Bioinformatics*, **29**, 3158-66.
- Echols, N. *et al.* (2003) MolMovDB: analysis and visualization of conformational change and structural flexibility. *Nucleic Acids Res.*, **31**, 478-482.
- Grant, B. J. *et al.* (2010) Large conformational changes in proteins: signaling and other functions. *Curr. Opin. Struct. Biol.*, **20**, 142-147.
- Greenidge, P. A. *et al.* (2014) Improving docking results via reranking of ensembles of ligand poses in multiple X-ray protein conformations with MM-GBSA. *J. Chem. Inf. Model.*, **54**, 2697-2717.
- Gu, Y. *et al.* (2015) Decoding the mobility and time scales of protein loops. *J. Chem. Theory Comput.*, **11**, 1308-1314.
- Gutteridge, A. and Thornton, J. (2004) Conformational change in substrate binding, catalysis and product release: an open and shut case? *FEBS Lett.*, **567**, 67-73.
- Henzler-Wildman, K. and Kern, D. (2007) Dynamic personalities of proteins. *Nature*, **450**, 964-972.
- Hrabe, T. *et al.* (2016) PDBFlex: exploring flexibility in protein structures. *Nucleic Acids Res.*, **44**, D423-D428.
- Hummer, G. *et al.* (2004) Unveiling functional protein motions with picosecond x-ray crystallography and molecular dynamics simulations. *Proc. Natl. Acad. Sci. USA*, **101**, 15330-15334.
- Jacobs, D. J. *et al.* (2001) Protein flexibility predictions using graph theory. *Proteins*, **44**, 150-165.
- James, L. C. and Tawfik, D. S. (2003) Conformational diversity and protein evolution - A 60-year-old hypothesis revisited. *Trends Biochem. Sci.*, **28**, 361-368.
- Joosten, R. P. *et al.* (2011) A series of PDB related databases for everyday needs. *Nucleic Acids Res.*, **39**, D411-9.
- Juritz, E. I. *et al.* (2011) PCDB: a database of protein conformational diversity. *Nucleic Acids Res.*, **39**, D475-D479.
- Kavanaugh, J. S. *et al.* (2005) Crystallographic evidence for a new ensemble of ligand-induced allosteric transitions in hemoglobin: the T-to-T_{High} quaternary transitions. *Biochemistry*, **44**, 6101-6121.
- Kosloff, M. and Kolodny, R. (2008) Sequence-similar, structure-dissimilar protein pairs in the PDB. *Proteins*, **71**, 891-902.
- Kuzu, G. *et al.* (2013) Exploiting conformational ensembles in modeling protein-protein interactions on the proteome scale. *J. Proteome Res.*, **12**, 2641-2653.
- Liang, S. *et al.* (2014) LEAP: Highly accurate prediction of protein loop conformations by integrating coarse-grained sampling and optimized energy scores with all-atom refinement of backbone and side chains. *J. Comput. Chem.*, **35**, 335-341.
- Marks, C. (2017) Sphinx: Merging knowledge-based and *ab initio* approaches to improve protein loop prediction. *Bioinformatics*, btw823.
- Monzon, A. M. (2013) CoDNAs: a database of conformational diversity in the native state of proteins. *Bioinformatics*, **29**, 2512-2514.
- Monzon, A. M. *et al.* (2017) Conformational diversity analysis reveals three functional mechanisms in proteins. *PLoS Comput. Biol.*, **13**, e1005398.
- Moult, J. *et al.* (1995) A large-scale experiment to assess protein structure prediction methods. *Proteins*, **23**, ii-iv.
- Nowak, J. *et al.* (2016) Length-independent structural similarities enrich the antibody CDR canonical class model. *mAbs*, **8**, 751-60.
- Osguthorpe, D. J. *et al.* (2012) Generation of receptor structural ensembles for virtual screening using binding site shape analysis and clustering. *Chem. Biol. Drug Des.*, **80**, 182-193.
- Palopoli, N. *et al.* (2016) Addressing the role of conformational diversity in protein structure prediction. *PLoS One*, **11**, e0154923.
- Pandey, B. P. *et al.* (2005) Protein flexibility prediction by an all-atom mean-field statistical theory. *Protein Sci.*, **14**, 1772-1777.
- Papaleo, E. (2016) The role of protein loops and linkers in conformational dynamics and allostery. *Chem. Rev.*, **116**, 6391-6423.
- Qi, G. and Hayward, S. (2009) Database of ligand-induced domain movements in enzymes. *BMC Struct. Biol.*, **9**, 13.
- Roberts, A. J. *et al.* (2013) Functions and mechanics of dynein motor proteins. *Nat. Rev. Mol. Cell Biol.*, **14**, 713-726.
- Ruvinsky, A. M. *et al.* (2011) Side-chain conformational changes upon protein-protein association. *J. Mol. Biol.*, **408**, 356-365.
- Sarkar, R. (2017) Native flexibility of structurally homologous proteins: insights from anisotropic network model. *BMC Biophys.*, **10**, 1.
- Stein, A. and Kortemme, T. (2013) Improvements to robotics-inspired conformational sampling in Rosetta. *PLoS One*, **8**, e63090.
- Tepljakov, A. *et al.* (2014) Antibody modeling assessment II. Structures and models. *Proteins*, **82**, 1563-82.
- Wang, G. and Dunbrack, R. L. (2003) PISCES: a protein sequence culling server. *Bioinformatics*, **19**, 1589-1591.
- Wang, Q. *et al.* (2013) Protein recognition and selection through conformational and mutually induced fit. *Proc. Natl. Acad. Sci. USA*, **110**, 20545-20550.
- Yaseen, A. *et al.* (2016) FLEXc: protein flexibility prediction using context-based statistics, predicted structural features, and sequence information. *BMC Bioinform.*, **17**, 281.