JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

Validated Conformational Ensembles Are Key for the Successful Prediction of Protein Complexes

Carles Pons,[†,‡] R. Bryn Fenwick,[§] Santiago Esteban-Martín,[†,§] Xavier Salvatella,*[,§,∥] and Juan Fernandez-Recio*[,†]

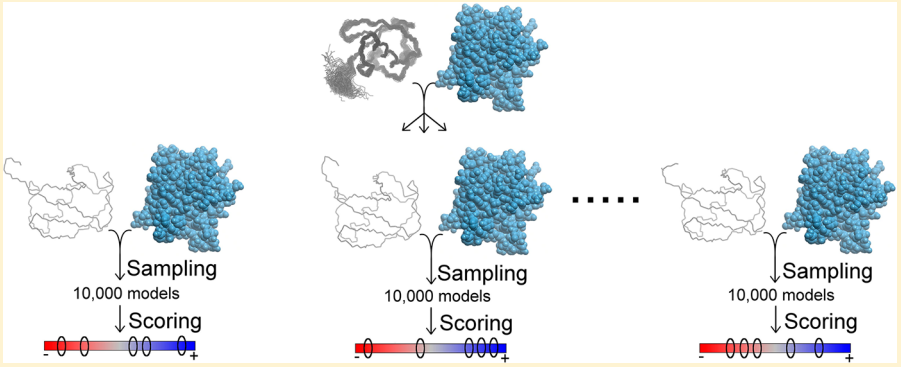
[†]Joint BSC-IRB research programme in Computational Biology, Barcelona Supercomputing Center (BSC), Jordi Girona 29, Barcelona 08034, Spain

[‡]Computational Bioinformatics, National Institute of Bioinformatics (INB), Jordi Girona 29, Barcelona 08034, Spain

[§]Joint BSC-IRB Research Programme in Computational Biology, Institute for Research in Biomedicine (IRB Barcelona), Baldiri Reixac 10, 08028 Barcelona, Spain

[∥]Institució Catalana de Recerca i Estudis Avançats (ICREA), Spain

Supporting Information



ABSTRACT: Conformational fluctuations in proteins play key roles in their functions and interactions. In this work, validated conformational ensembles for ubiquitin have been used in docking trials. The ensembles were used in a systematic predictive study of known ubiquitin complexes by applying a cross-docking strategy against the bound structure of each partner. The global docking predictions obtained with the complete ubiquitin ensembles were significantly better than those obtained with the crystallographic structure of free ubiquitin. Importantly, in all cases we identified an individual ensemble member that performed equally well, or even better, than the bound structure of ubiquitin. These results unequivocally demonstrate that, for proteins that recognize binding partners by conformational selection, the availability of conformational ensembles can greatly improve the performance of automatic docking predictions. Our results highlight the need for docking methodologies to capitalize on validated ensemble representations of biomacromolecules.

INTRODUCTION

The association of two or more proteins to form macromolecular complexes is a fundamental mechanism in cell function.[1] The characterization of the atomic details of such interactions is key to understand the underlying nature of the association process[2] and to modulate interactions of biomedical and biotechnological interest. This challenging task constitutes one of the major goals of structural biology, and interdisciplinary methodologies need to be implemented to overcome the technical limitations of current experimental techniques.[3,4]

In response to this, a variety of protein−protein docking tools have been reported to model the structure of protein complexes starting from the coordinates of the individual components.[5] Although highly successful models can be produced in cases where the interacting proteins behave as rigid bodies,[6] the main challenge for the general applicability of docking predictions is how to accurately model the flexibility of the interacting proteins.[7,8] This problem is related to our lack of detailed understanding of the protein−protein association mechanism where, except for a few rigid-body cases that appear to follow a "lock-and-key" mechanism,[9] most of these interactions involve conformational changes and are therefore best explained by the induced-fit[10] or conformational selection[11−16] mechanisms of molecular recognition. Therefore, over the last 10 years, different state-of-the-art docking algorithms have attempted to improve the description of the association process by including conformational flexibility upon binding.

Received: November 12, 2012
Published: February 19, 2013

The majority of flexible docking approaches implicitly mimic the induced-fit mechanism,[10] by following a two-step process involving an initial rigid-body search followed by a refinement step in which the molecules are allowed to change their conformation. Flexibility can be included explicitly, by using molecular mechanics force fields and/or rotamer libraries to model side-chain and/or backbone movements,[17−20] or implicitly, where a certain degree of interpenetration is allowed by using soft potentials implemented either in grids,[21,22] all-atom,[23,24] or coarse-grained[25−27] representations.

An alternative approach is to instead mimic conformational selection.[11,15,16] According to this model the free protein is best represented as an ensemble of conformations in equilibrium, containing every possible state that the protein may adopt, including the bound state. Upon binding, the bound conformation is selected and stabilized by the binding partner, due to its optimal geometry and energy complementarity, thus shifting the population of the ensemble in favor of the bound state.[13,14,28,29] In cases where recognition occurs according to this model, one strategy to account for flexibility in docking would therefore be to generate a conformational ensemble for the free protein, with the expectation that ensemble members resembling the bound state would lead to better docking predictions.[30] Several macromolecular docking algorithms have been reported to perform ensemble docking, typically using a few selected conformers applied to specific cases of interest.[31−33] However, it is not yet clear how to generate conformational ensembles or select subensembles that are adequate for docking.

The few systematic studies on the use of conformational ensembles in docking have relied on a relatively small number of representative conformations derived from theoretical models that do not necessarily provide a realistic representation of the motions occurring in the time scale of molecular association.[34−36] An alternative approach based on nuclear magnetic resonance (NMR) ensembles has been described,[36] but the structural variability of such ensembles originates from lack of precision in the average coordinates of the protein rather than from actual structural heterogeneity due to protein flexibility.[37] As a consequence, none of such ensembles showed a clear improvement in docking predictive rates with respect to the unbound state. Therefore, there is a need to understand better how to generate conformational ensembles that could be useful for docking in cases following the conformational selection mechanism. Ensembles that underestimate structural heterogeneity could lead to inaccurate results in docking due to the absence of bound conformers, while ensembles that overestimate it could lead to unrealistic population of high-energy states, with excessive number of false positives and higher computational costs in docking.

Recently, we and others have generated conformational ensembles that aim at representing the structural heterogeneity of biomolecules up to the millisecond time scale by using ensemble molecular dynamics (MD) simulations restrained by residual dipolar couplings (RDC) measured using NMR,[38−40] which provide information about the amplitude of the angular fluctuations of specific bond vectors of the protein. Importantly, these ensembles report on the structural heterogeneity[39] due to motions occurring in the time scale of biomolecular association, as recently shown for the protein ubiquitin.[41] These ensembles are known to cover the same conformational space as the X-ray structures of ubiquitin in complex with different protein partners, which shows that ubiquitin complexes are formed to a large extent by conformational selection.[28,39,40,42] However, the validity of such ensembles for predicting the structure of ubiquitin complexes by docking has never been proven. With this purpose, we have used here pyDock, an energy-based rigid-body docking and scoring scheme, to dock a conformational ensemble of ubiquitin to 11 different binding partners, and compared the docking results to those obtained with the bound and unbound crystallographic structures of ubiquitin. We show for the first time that such experimentally validated ensembles are optimal for docking and can improve the performance with respect to the use of unbound structure, giving the same predictive results as when using the bound structures.

METHODS

Generation of the Ubiquitin Conformational Ensemble. The ensemble was derived from a set of 1971 RDCs and 2663 nuclear Overhauser effects (NOEs) available in the literature[39] by using these NMR parameters to restrain ensemble MD simulations carried out using an in-house modified version of the molecular simulation package CHARMM (version c35). The restrained simulations, started from the X-ray structure of ubiquitin (PDB code 1UBQ) and composed of 10 simulated annealing cycles, were carried out, as previously described,[40,43,44] using an ensemble size of 32 and lead to a 320 membered ensemble. The RDCs corresponding to residues in the C-terminal disordered tail of ubiquitin were not used as restraints because their fluctuations correlate to those of the alignment tensor.[45] For comparison purposes, we also generated another ubiquitin ensemble using unrestrained MD simulations as described in the Supporting Information (SI).

Docking Sampling and Scoring. In this work we ran >7000 docking experiments between different structures of ubiquitin (bound, unbound, and conformers from unbound ensembles) and the corresponding partner protein in a set of ubiquitin complexes, as described below. For each docking experiment we followed the same protocol: First we ran FTDock[21] with 0.7 Å grid resolution and 12° angular resolution for the generation of 10 000 rigid-body poses per docking execution based on surface complementarity and an electrostatics term. These docking models were evaluated by pyDock,[46] a scoring function that calculates the binding energy between the interacting proteins, taking into account the desolvation, electrostatics, and van der Waals energy contributions. The desolvation term was based on the accessible surface area (ASA) with atomic solvation parameters previously optimized for rigid-body docking.[47] The electrostatics term was calculated as a Coulomb potential with a distance-dependent dielectric constant, and the van der Waals term was based on a 6-12 Lennard-Jones potential. Both terms used AMBER94 parameters and were implemented as soft potentials to allow a certain overlap between the structures. This scoring function yielded successful results in the Critical Assessment of PRediction of Interactions (CAPRI).[48,49] The quality of the docking models was assessed by calculating the ligand Cα-RMSD with respect to the crystal structure of the known complexes. A docking run was considered successful if a near-native solution, i.e., a docking pose with ligand Cα-RMSD < 10 Å, was ranked among the top 10 predictions according to the pyDock scoring.

Single-Structure and Ensemble Docking Benchmarks. We used a set of ubiquitin complexes with known crystallographic structures, as listed in a recent work,[39] after discarding

homocomplexes, redundant cases, and complexes involving proteins with <30 amino acids (see Table 1). For each one of these 11 ubiquitin complexes we ran docking experiments using different conformations for the ubiquitin: (i) the bound crystallographic structure in each one of the complexes; (ii) the unbound ubiquitin structure (PDB code 1UBQ); and (iii) a conformational ensemble derived from NMR data (see above). For comparison purposes, we also used an ensemble obtained using MD. In experiments using conformational ensembles, a single docking was run for each of the 320 conformers. However, the discrete sampling of FTDock might randomly benefit from a particular starting disposition of a conformer in the grid, and thus the better docking results when using an ensemble might be explained just by simple statistical reasons. To disregard this possibility, we ran docking simulations from 10 different random rotations of the initial structures for the bound and unbound ubiquitin (for comparison purposes, we also show the results of the docking with the single unbound or bound crystallographic structures; Table 2). In all docking experiments we used the bound conformation of the binding partner of ubiquitin, for technical reasons and because it gives a better assessment of the suitability of ubiquitin conformational ensembles for docking (see Results and Discussion). For the global docking analysis, all the docking poses resulting from the different conformers (or random rotations) were merged as a single set and sorted according to pyDock score. The global docking analysis included a filtering step to reduce redundancy, based on a single-pass "Leader" clustering algorithm. The process started with the top-ranked docking model according to pyDock, removed all duplicate docking poses with ligand Cα-RMSD < 5 Å and repeated this iteratively for the next remaining models. A total number of >7000 docking executions were run in the Mare Nostrum supercomputer for this work.

Table 1. List of Ubiquitin Complexes Used in This Work[a]

	ubiquitin			partner		complex
PDB	chain	no. res	RMSD (Å)	chain	no. res	I_{area} (Å^2)
1CMX	B	75	0.88	A	214	2330
1NBF	C	75	1.96	B	347	3446
1OTR	B	76	2.05	A	49	1061
1P3Q	V	73	0.85	Q, R	77	1842
1S1Q	D	76	1.57	C	139	1213
1WR1	A	76	2.22	B	58	814
1XD3	B	75	0.78	A	229	2232
1YD8	U	73	0.61	H	93	1119
2AYO	B	75	1.88	A	355	2898
2C7M	B	73	0.46	A	58	1358
2G45	B	76	1.48	A	117	981

[a]The chains used for docking and their lengths are reported. RMSD: root mean square deviation of the unbound ubiquitin (PDB ID 1UBQ) with respect to the bound ubiquitin in the complex. I_{area}: the surface area buried upon complex formation.

RESULTS AND DISCUSSION

Docking with Single Ubiquitin Structures. This docking study has been performed on a set of protein−protein complexes, in which ubiquitin is bound to a variety of partner proteins, which are shown in Figure 1. Most of these complexes can be defined as difficult docking cases as they are characterized by conformational changes >1 Å and small surface accessible areas.[50] Table 1 shows the list of complex structures and provides information on the structural changes experienced by ubiquitin upon binding in terms of RMSD as well as the intermolecular contact area (I_{area}). The aim of the current work is to find whether experimentally derived conformational ensembles for unbound ubiquitin can be useful for modeling the structure of these ubiquitin complexes by docking. We decided to use the bound conformation of the partner protein in each complex for several reasons: First, one might think that a more realistic scenario for docking purposes would be, in principle, to use conformational ensembles for the partner protein as well, however generating experimental ensembles for the unbound partner protein in each complex and running millions of docking executions are beyond the scope of current work. Second, for the purposes of this work, that is testing the ubiquitin ensembles for docking, the bound conformation of the partner is a better substitute for an ideal partner ensemble than the unbound conformation itself because, under the assumption of conformational selection, the bound conformation of the partner is present in the unbound ensemble. In this way we can test whether any conformer from the ubiquitin ensembles can be successfully

Table 2. Docking Results: Best Rank of a Near-Native Solution According to pyDock[a]

	bound ubiquitin			unbound ubiquitin		RDC ensemble		MD ensemble	
case	single	10 rot	native added	single	10 rot	global	best	global	best
1CMX	1	1	1	8	1	4	1	2	1
1NBF	1	1	1	3	21	9	1	1	1
1OTR	55	136	117	435	170	5	1	5	1
1P3Q	80	123	13	21	27	3	1	3	1
1S1Q	85	157	12	1031	4816	140	6	344	4
1WR1	167	541	541	337	2342	111	3	553	9
1XD3	1	1	1	1	2	4	1	2	1
1YD8	7	6	6	219	46	5	1	14	1
2AYO	532	3	3	74	54	70	1	11	1
2C7M	2	5	5	78	2	1	1	3	1
2G45	5	3	1	3654	40	117	2	54	1

[a]Results for bound and unbound ubiquitin are shown, when using a single orientation as well as 10 random orientations (see Methods). For bound ubiquitin, results are also shown after adding the native orientation to the pool of docking models obtained from 10 random orientations. Results for the RDC-based and MD ensembles are shown as "global" (best near-native rank after merging the docking models from all individual conformers and clustering) and as "best" (best near-native rank obtained by any individual docking for a given initial conformer).

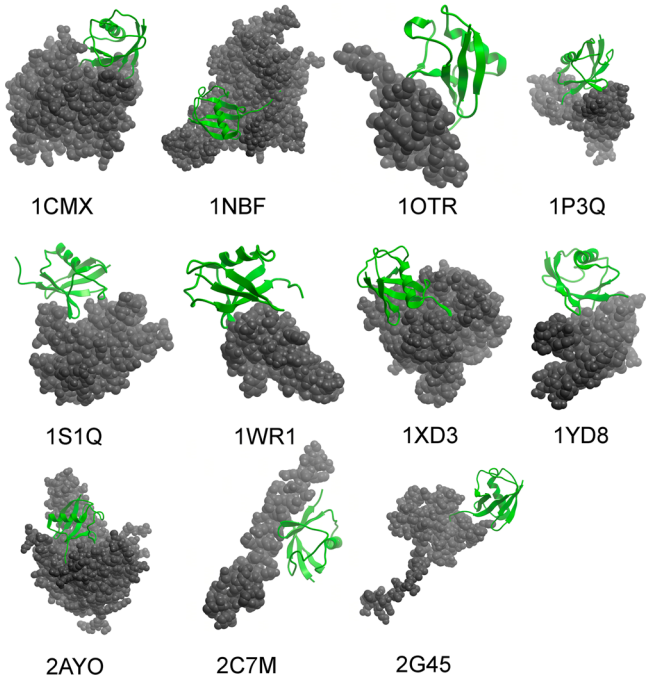


Figure 1. Ubiquitin complexes used in this work. Ubiquitin is shown in green ribbon, and the corresponding partner protein in gray CPK.

docked to the bound conformation ideally present in the partner ensemble and consequently the number of docking executions that need to be tested in the current work can be reduced without any loss of generality. Therefore, for the purposes of assessing the ubiquitin conformational ensemble as a useful tool for docking, we considered appropriate running docking with the bound partner protein, even though the success rates shown here arise from an optimal situation and would be difficult to achieve in a real case.

To define a control set of ideal docking results for this work, we first docked the bound structure of ubiquitin in each complex. Table 2 shows the single best rank obtained by a near-native solution as evaluated by the pyDock scoring function. We used 10 random initial rotations for each case (see Methods) to check if the results could improve by simple statistical reasons arising from small variations in the atomic positions due to the discrete representation in FTDock. This is indeed the case for 2AYO, in which the poor results obtained by a single bound structure dramatically improved after using 10 random initial rotations. But apart from this case, the results did not generally change after using different initial rotations. We obtained successful results, i.e., a near-native solution was found within the top 10 docking poses, for 7 of the 11 complexes (Table 2). Moreover, three complexes were correctly identified as the top ranking solution by pyDock (1CMX, 1NBF, and 1XD3). As recently discussed in the literature,[6,50,51] the poor docking performance for some complexes can be explained by different reasons related to scoring issues: (i) small interface areas, only three out of seven cases with an interface area smaller than 2000 Å^2 were successful, whereas all four cases with larger interfaces were successful; (ii) low affinity; (iii) the promiscuous nature of ubiquitin; and (iv) limited sampling of the grid-based FFT approach. We tested the latter by adding the native orientation to the pool of docking poses, thus ensuring that the exact bound orientation was sampled and scored. This improved the results in some of the cases (Table 2), like the low-affinity complex 1S1Q. However, the suboptimal docking results in several cases indicates that scoring of docking poses is not yet optimal for the reasons discussed above.

We then performed docking experiments using the unbound structure of ubiquitin to establish the expected performance of our protocol when using the available ubiquitin structure instead of a conformational ensemble (Table 2). We again used 10 random rotations from PDB code 1UBQ (see Methods) to account for possible statistical effects on the docking success rates, although the overall results did not vary (except for case 2C7M in which the results improved, and case 1NBF in which they became worse). As expected, the results were worse than those obtained when using the bound ubiquitin structure. In 3 out of 11 cases, a near-native solution was ranked by pyDock within the top 10 predictions of the 10 000 docking poses (Table 2). This success rate is within the range expected according to state-of-the-art docking benchmarks.[6] Previous studies showed that success rates of unbound docking largely depended on the extent of conformational change upon binding.[6] Consistently, here for the five complexes in which ubiquitin showed Cα-RMSD < 1.0 Å upon binding, docking was successful in three of them (moreover, in all cases a near-native solution ranked within the top 50 docking poses). In contrast, pyDock succeeded in none of the six remaining cases in which ubiquitin showed Cα-RMSD > 1.0 Å upon binding (and actually in only two cases a near-native solution ranked within the top 50 poses).

Docking with Ubiquitin Ensembles. The main goal of this work is to assess whether validated ensemble representations of proteins can improve the docking results obtained with the unbound ubiquitin structure. For this purpose, we individually docked the 320 ubiquitin conformations of a validated conformational ensemble and then combined and scored the resulting docking poses (see Methods). Table 2 shows that in 7 out of the 11 cases a near-native solution was found among the top 10 predictions (see Table 2), a significant improvement with respect to the results obtained with unbound ubiquitin. Remarkably, this global docking strategy showed essentially the same level of success as using the bound reference structure, although some of the successful cases were different in the two docking conditions.

Our ensemble docking strategy is similar to the approach used in the study of Grünberg et al.,[34] where ensemble docking for 17 complexes was used to characterize the surface complementarity between different members of their ensembles; however, that study only used 10 conformers per protein and did not reportedly attempt any scoring of the resulting docking poses. We can also compare our results with those of Smith et al.,[35] who used a similar strategy but using only 1 or 2 conformers per protein, which did not significantly improve the predictions obtained from the unbound structures. We speculate that the improvement that we observed is in part due to the improved sampling afforded by the larger size of our conformational ensemble. Indeed, Smith et al.[35] could not find a single conformer that fully sampled the bound state, although they observed that different conformers contained different parts of the bound state.

Finally, a different docking study found that an ensemble derived from MD could improve docking results with respect to those obtained when using the unbound protein, although sampling was limited to a small fraction of the interacting space.[36] In addition, that approach was not able to achieve the

same level of success as that achieved by docking with the bound reference structure.[36] Here, by contrast, we have been able to achieve this conceptually very relevant result. Moreover, for two cases that showed poor results when docking the bound subunits (1OTR and 1P3Q), our ensemble approach yielded much better performance (see Table 2). Indeed, we found that using conformational ensembles improved the docking results obtained when using a single unbound structure, and in some cases they could even outperform the results from the bound reference structure.

This of course does not necessarily mean that we could easily obtain these results in a realistic situation, in which only the structure of the unbound partner protein is available, but shows for the first time that validated ubiquitin ensembles can give the same docking results as when using the bound ubiquitin structure and suggests that future docking strategies based on precomputed ensembles for each of the partners can lead to accurate predictive results. The experimental ubiquitin ensemble used here was previously reported to contain conformers that were close to any of the known ubiquitin complex structures, but this is the first time that can be shown that these ensembles are indeed useful to predict the structure of ubiquitin complexes by docking.

For comparison, we have also performed ensemble docking with an ensemble generated by using unrestrained MD. For small globular proteins, current force fields give good agreement with experimental data and thus are expected to perform comparatively well for ubiquitin docking.[52,53] For the MD ensemble, the results are very similar (Table 2). Interestingly, the results achieved using the conformers of the ensemble derived from RDCs were in general better in the cases where the C-terminal disordered tail of ubiquitin was not involved in the interaction (1OTR, 1S1Q, 1WR1, 1YD8). In three of these cases (1WR1, 1S1Q, and 1YD8), the RDC-based ensemble yielded better results than bound ubiquitin (considering 10 random initial rotations), while the MD-based ensemble yielded worse results than bound ubiquitin. This can, to a significant extent, be attributed to the fact that the structural heterogeneity of the tail in the RDC-based ensemble is not defined by the experimental RDCs and is instead defined by the simulated annealing procedure. This is because only the RDCs of residues whose fluctuations do not affect the alignment tensor of the protein can be used as restraints in the approach that was used to determine the RDC-based ensemble as recently discussed.[45] As a consequence the structural heterogeneity of the tail in the experimental ensemble is not as reliable as that of the globular part of ubiquitin. These results suggest that experimental ensembles perform better in docking than MD-based ones when the region of structure involved in the protein−protein interaction is well-defined by the experimental data. In addition, it is relevant to mention that, although no experimental data was used to restrain the MD simulation, the resulting ensemble validates against experimental data better than the X-ray structure of ubiquitin (see Table S1), indicating that the MD-based ensemble is also a realistic representation of the structural heterogeneity of ubiquitin. These observations suggest that, while the best possible result is obtained when the structural heterogeneity can be defined by the experimental data, when such data are not available, MD can provide a comparable performance provided that the trajectories are realistic, as in the cases studied here (Table 2). This result contrasts with that obtained by Chaudhury et al.,[36] who observed worse results when using an NMR ensemble, however this result was probably due to the, at best, qualitative correlation between actual structural heterogeneity and apparent structural heterogeneity in the NMR structure ensemble. As an example, in the conformational ensembles derived from RDCs, a common pairwise RMSD is 0.8 Å between individual members, as compared to the pairwise RMSD of 0.1 Å for the NMR structural ensemble of ubiquitin (PDB code 1D3Z). This value of 0.8 Å is representative of the heterogeneity of biomolecular ensembles from MD simulations and X-ray crystal B-factors, while the structural heterogeneity within the NMR ensemble is not.[37,54]

Number of Near-Native Docking Solutions Generated by Individual Ensemble Members. In all of the complexes studied here we observed at least one near-native solution for each member of the ubiquitin conformational ensemble. The only exception to this was case 2G45, however, even in this difficult complex we still obtained near-native solutions for 92.2% of the ensemble members. The ability to obtain good solutions for nearly all conformations is consistent with Grünberg et al.,[34] where the authors found multiple optimal complementary conformers within the ensembles. The average number of near-native solutions generated for all the conformers in each of the 11 complexes was generally lower than the average number obtained when docking the 10 randomly oriented single bound or unbound ubiquitin structures (see Figure 2), indicating that not all conformers

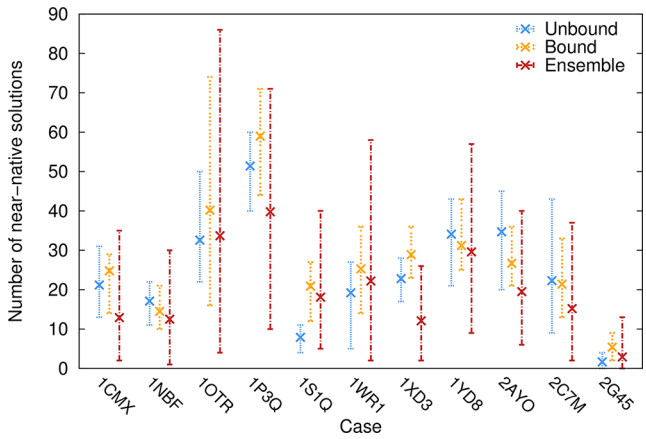


Figure 2. Near-native solutions generated by docking with a RDC-based ubiquitin ensemble. Minimum, maximum, and average number of near-native solutions generated for each complex with the conformers of the RDC-derived ensembles (in red). The results obtained with the bound (in orange) and unbound (in blue) ubiquitin (10 random rotations) are shown for comparison.

are binding competent. However, in all cases (except 1XD3), there were always conformers that generated more near-native docking solutions than bound ubiquitin (see Figure 2). For instance, in case 1OTR the average number of near-native solutions using the conformational ensemble was 33.7, below the average number obtained when using the bound (40.2) ubiquitin. Even so, 94 of the ensemble conformers (29.4% of the ensemble) generated more near-native solutions than when using the bound structure. Moreover, one of the members of the ensemble generated 86 near-native solutions, which remarkably doubles the average number of native-like solutions obtained with the bound structure.

Best Docking Results from Individual Conformers. We analyzed the pyDock scoring results individually for each single

conformer of the ensemble. In all cases there was always a conformer that yielded a near-native docking solution among the top 10 pyDock predictions, and in 8 of the 11 cases there was an individual docking run where a near-native solution was ranked 1 (see Table 2). For the complex that performed best (2C7M), as many as 273 (85.3%) of the members of the ensemble had near-native solutions among the top 10 predictions. For this example, 58.8% and 75.3% of the conformers outperformed the docking results with the unbound and bound ubiquitin structures (10 random rotations), respectively (see Figure 3). In the most challenging

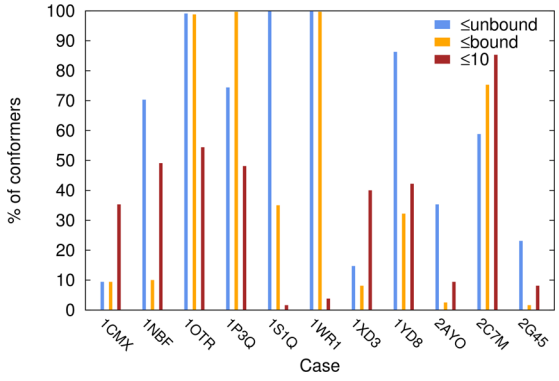


Figure 3. Docking results with ubiquitin ensemble. Percentage of ensemble members with the following docking results: (i) a near-native solution among the top 10 predictions (in red); (ii) a near-native solution ranked equally or better than that in the docking (10 random rotations) of unbound ubiquitin (in blue); (iii) a near-native solution ranked equally or better than that in the docking (10 random rotations) of bound ubiquitin (in orange).

complexes (1S1Q, 1WR1, 2G45, and 2AYO), less than 10% of the ensemble members produced a near-native solution within the top 10 predictions. However, even in these cases a large fraction of the ensemble members yielded better docking results than the unbound (10 random rotations) ubiquitin structure (100, 100, 23.1, and 35.3%, respectively). Interestingly, if a conformer was randomly chosen from the ensemble, the average number of successful cases would be 3.8 (as obtained from 10 000 random experiments), which is better than the results obtained when using the unbound form of ubiquitin (3 successful cases). This suggests that random selection of a conformer from the ensemble is sufficient to improve the unbound docking results.

Are the Best Conformers for Docking the Most Similar Ones to the Bound States? In the previous sections we have shown that some members of the conformational ensembles can produce more and better-ranked near-native docking orientations than others. In an attempt to better understand the results, we also analyzed the structural characteristics of the best-performing conformers. In all cases, there was an ensemble member that performed better than unbound or even bound ubiquitin. We observed that the best-performing conformers were different in the 11 cases. However, three specific ensemble members were successful (i.e., rank of a near-native solution $\leq$10) in as many as 8 out of the 11 cases, i.e., these three conformers showed in average better success than the unbound or bound ubiquitin (see example in Figure 4). Intriguingly, in some cases these three conformers showed very different Cα-RMSD with respect to the bound form of ubiquitin. For example in 1CMX, one of these conformers had

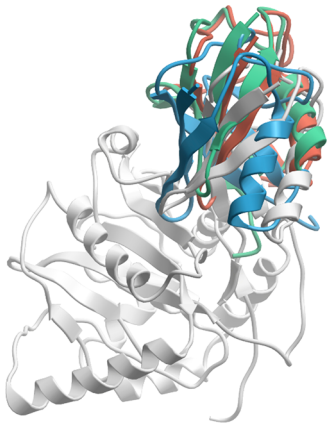


Figure 4. Docking structures obtained by best-performing conformers of the ubiquitin ensemble. Docking results for case 1XD3 with conformers no. 15 (near-native solution, ranked 1, is shown in red), no. 101 (near-native ranked 9 in green), and no. 130 (near-native ranked 3 in blue). Complex reference structure is shown in white for comparison.

a Cα-RMSD 0.6 Å with respect to the bound state in that complex, while another one showed a Cα-RMSD 1.7 Å. Despite the difference in the RMSDs with respect to the bound form, these two conformers were equally successful. Moreover, in most of the complexes, these three successful conformers were not actually the most similar to the bound form among all ensemble members. Intriguingly, therefore, the better performance of these successful conformers cannot be explained by the similarity to the bound structures.

In general, there was no obvious correlation between the docking success of a conformer and its similarity to the bound state of ubiquitin in a given complex (see Figure S1), in agreement with previous studies.[34] We only observed correlations above 0.10 between the best rank of a near-native solution obtained with a given conformer and its Cα-RMSD with respect to the bound ubiquitin for cases 2C7M, 2G45, and 1YD8 (see Figure 5). Correlations did not significantly improve when using only the Cα of residues at the interface or when taking into account all the atoms. Interestingly, we found that the total number of higher quality near-native solutions (those within 5 Å RMSD from the reference structure) generated by FTDock correlated with the similarity of each conformer to the bound state in many cases (Figure S2). The fact that the same correlation is not seen for the docking rank values (Figure S1) could be explained by technical reasons, such as the inherent noise from the grid-based representation of proteins (at 0.7 Å resolution) or from the soft potentials used in scoring. An alternative explanation would be that some of the successful conformers might favor the formation of initial contacts in some specific residues (e.g., anchor residues),[55] perhaps leading to encounter complexes, in which interface residue conformation might be slightly different from that in the final bound state. In any case, this shows the importance of using different conformers in docking to increase the probability of detecting specific favorable contacts. Indeed, when we analyzed the global docking results for each case, the quality of the ensemble is critical. In those cases in which the ensemble contains at least one conformer within 0.8 Å RMSD from the bound state (i.e., 1CMX, 1P3Q, 1XD3, 1YD8, 2C7M, 2G45), the global docking results were excellent, with a near-native ranked within top 10 in 83% of these cases. This shows that the global success rates

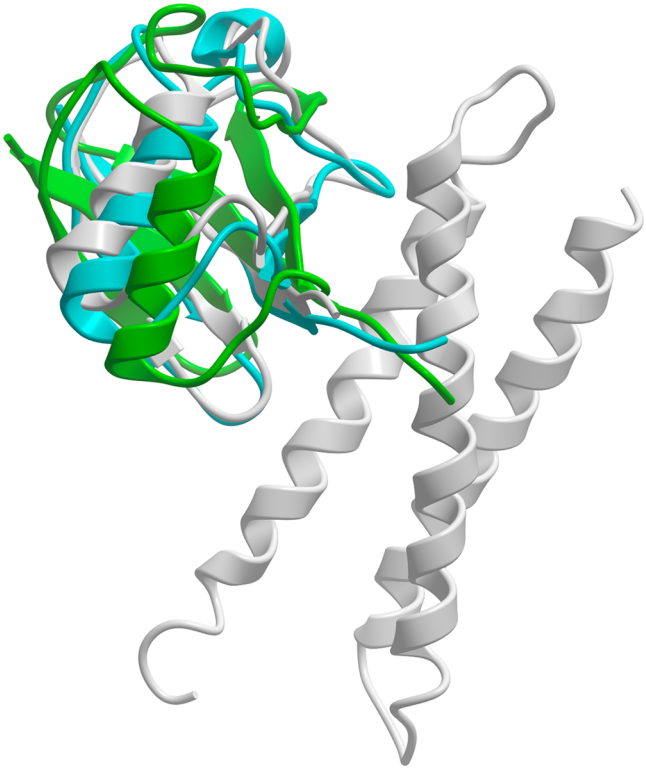


Figure 5. Predicted structures from docking using the ubiquitin ensemble. Docking results for case 1YD8. In green, best near-native solution from unbound docking (rank 46, RMSD 5.7 Å). In blue, best near-native solution from ensemble docking (rank 5, RMSD 2.1 Å). Complex reference structure is shown in white for comparison.

strongly improve when the bound state is sufficiently represented within the ensemble, and this is the case of the experimentally validated conformations here.

CONCLUSIONS

We have performed here a systematic docking study of 11 ubiquitin complexes, using a rigid-body docking protocol and a conformational ensemble of ubiquitin to mimic a possible conformational selection binding mechanism in ubiquitin. We used an ensemble of 320 conformers of ubiquitin previously derived from experimental data, which represent the largest ensemble used in a systematic docking study so far. The ensemble members were individually docked to the corresponding partner protein structure in each complex, in order to find whether the bound partner conformation was able to efficiently dock to suitable conformers and therefore improve the docking predictions. The global docking results, after merging and clustering the 3 200 000 models generated for each case in a cross-docking strategy, were clearly better than the ones obtained with the single unbound structure and of the same quality as those achieved by the bound conformation. This shows that current docking strategies can benefit from the structural heterogeneity provided by validated conformational ensembles. For all cases, we always found a conformer that generated better results than the ones obtained with the unbound or bound ubiquitin. Through a combination of docking and precomputed unbound ensembles to mimic the conformational selection mechanism in ubiquitin complexes, we have shown that validated ensembles can be automatically used to significantly improve the structural prediction of protein complexes.

ASSOCIATED CONTENT

Supporting Information

Two supporting figures, one supporting table, and methods describing the generation of the RDC restrained ensemble. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*E-mail: juanf@bsc.es; xavier.salvatella@irbbarcelona.org.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by the Spanish Ministry of Science (grant nos. BIO2010-22324 and CTQ2009-08850-BQU), ICREA, and Marató de TV3 (102030).

REFERENCES

(1) Rual, J.-F.; Venkatesan, K.; Hao, T.; Hirozane-Kishikawa, T.; Dricot, A.; Li, N.; Berriz, G. F.; Gibbons, F. D.; Dreze, M.; Ayivi-Guedehoussou, N.; Klitgord, N.; Simon, C.; Boxem, M.; Milstein, S.; Rosenberg, J.; Goldberg, D. S.; Zhang, L. V.; Wong, S. L.; Franklin, G.; Li, S.; Albala, J. S.; Lim, J.; Fraughton, C.; Llamosas, E.; Cevik, S.; Bex, C.; Lamesch, P.; Sikorski, R. S.; Vandenhaute, J.; Zoghbi, H. Y.; Smolyar, A.; Bosak, S.; Sequerra, R.; Doucette-Stamm, L.; Cusick, M. E.; Hill, D. E.; Roth, F. P.; Vidal, M. *Nature* **2005**, *437*, 1173−1178.
(2) Aloy, P.; Russell, R. B. *Nat. Rev. Mol. Cell Biol.* **2006**, *7*, 188−197.
(3) Robinson, C. V.; Sali, A.; Baumeister, W. *Nature* **2007**, *450*, 973−982.
(4) Alber, F.; Förster, F.; Korkin, D.; Topf, M.; Sali, A. *Annu. Rev. Biochem.* **2008**, *77*, 443−477.
(5) Ritchie, D. W. *Curr. Protein Pept. Sci.* **2008**, *9*, 1−15.
(6) Pons, C.; Grosdidier, S.; Solernou, A.; Pérez-Cano, L.; Fernández-Recio, J. *Proteins* **2010**, *78*, 95−108.
(7) Bonvin, A. M. J. J. *Curr. Opin. Struct. Biol.* **2006**, *16*, 194−200.
(8) Zacharias, M. *Curr. Opin. Struct. Biol.* **2010**, *20*, 180−186.
(9) Fischer, E. *Ber. Dtsch. Chem. Ges.* **1894**, *27*, 2985−2993.
(10) Koshland, D. E. *Proc. Natl. Acad. Sci. U.S.A.* **1958**, *44*, 98−104.
(11) Straub, F. B.; Szabolcsi, G. Remarks on the dynamic aspect of enzyme structure. In *Molecular biology: problems and perspectives*; Braunstein, A. E., Ed.; Nauka: Moscow, 1964; 182−187.
(12) Monod, J.; Wyman, J.; Changeux, J. P. *J. Mol. Biol.* **1965**, *12*, 88−118.
(13) Tsai, C. J.; Kumar, S.; Ma, B.; Nussinov, R. *Protein Sci.* **1999**, *8*, 1181−1190.
(14) Ma, B.; Kumar, S.; Tsai, C. J.; Nussinov, R. *Protein Eng.* **1999**, *12*, 713−720.
(15) Kumar, S.; Ma, B.; Tsai, C. J.; Sinha, N.; Nussinov, R. *Protein Sci.* **2000**, *9*, 10−19.
(16) Boehr, D. D.; Nussinov, R.; Wright, P. E. *Nat. Chem. Biol.* **2009**, *5*, 789−796.
(17) Fernández-Recio, J.; Totrov, M.; Abagyan, R. *Proteins* **2003**, *52*, 113−117.
(18) Dominguez, C.; Boelens, R.; Bonvin, A. M. J. J. *J. Am. Chem. Soc.* **2003**, *125*, 1731−1737.
(19) Wang, C.; Schueler-Furman, O.; Baker, D. *Protein Sci.* **2005**, *14*, 1328−1339.
(20) Mashiach, E.; Nussinov, R.; Wolfson, H. J. *Nucleic Acids Res.* **2010**, *38*, W457−461.
(21) Gabb, H. A.; Jackson, R. M.; Sternberg, M. J. *J. Mol. Biol.* **1997**, *272*, 106−120.

(22) Garzon, J. I.; Lopéz-Blanco, J. R.; Pons, C.; Kovacs, J.; Abagyan, R.; Fernandez-Recio, J.; Chacon, P. *Bioinformatics* **2009**, *25*, 2544−2551.
(23) Fernández-Recio, J.; Totrov, M.; Abagyan, R. *Protein Sci.* **2002**, *11*, 280−291.
(24) Cheng, T. M.-K.; Blundell, T. L.; Fernandez-Recio, J. *Proteins* **2007**, *68*, 503−515.
(25) Zacharias, M. *Protein Sci.* **2003**, *12*, 1271−1282.
(26) Ravikant, D. V. S.; Elber, R. *Proteins* **2010**, *78*, 400−419.
(27) Pons, C.; Talavera, D.; de la Cruz, X.; Orozco, M.; Fernandez-Recio, J. *J. Chem. Inf. Model.* **2011**, *51*, 370−377.
(28) Wlodarski, T.; Zagrovic, B. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 19346−19351.
(29) Fenwick, R. B.; Esteban-Martín, S.; Salvatella, X. *Eur. Biophys. J.* **2011**, *40*, 1339−1355.
(30) Andrusier, N.; Mashiach, E.; Nussinov, R.; Wolfson, H. J. *Proteins* **2008**, *73*, 271−289.
(31) Bastard, K.; Thureau, A.; Lavery, R.; Prévost, C. *J. Comput. Chem.* **2003**, *24*, 1910−1920.
(32) Bastard, K.; Prévost, C.; Zacharias, M. *Proteins* **2006**, *62*, 956−969.
(33) van Dijk, M.; van Dijk, A. D. J.; Hsu, V.; Boelens, R.; Bonvin, A. M. J. J. *Nucleic Acids Res.* **2006**, *34*, 3317−3325.
(34) Grünberg, R.; Leckner, J.; Nilges, M. *Structure* **2004**, *12*, 2125−2136.
(35) Smith, G. R.; Sternberg, M. J. E.; Bates, P. A. *J. Mol. Biol.* **2005**, *347*, 1077−1101.
(36) Chaudhury, S.; Gray, J. J. *J. Mol. Biol.* **2008**, *381*, 1068−1087.
(37) Esteban-Martín, S.; Bryn Fenwick, R.; Salvatella, X. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2012**, *2*, 466−478.
(38) Clore, G. M.; Schwieters, C. D. *J. Am. Chem. Soc.* **2004**, *126*, 2923−2938.
(39) Lange, O. F.; Lakomek, N.-A.; Farès, C.; Schröder, G. F.; Walter, K. F. A.; Becker, S.; Meiler, J.; Grubmüller, H.; Griesinger, C.; de Groot, B. L. *Science* **2008**, *320*, 1471−1475.
(40) Fenwick, R. B.; Esteban-Martín, S.; Richter, B.; Lee, D.; Walter, K. F. A.; Milovanovic, D.; Becker, S.; Lakomek, N. A.; Griesinger, C.; Salvatella, X. *J. Am. Chem. Soc.* **2011**, *133*, 10336−10339.
(41) Ban, D.; Funk, M.; Gulich, R.; Egger, D.; Sabo, T. M.; Walter, K. F. A.; Fenwick, R. B.; Giller, K.; Pichierri, F.; de Groot, B. L.; Lange, O. F.; Grubmüller, H.; Salvatella, X.; Wolf, M.; Loidl, A.; Kree, R.; Becker, S.; Lakomek, N.-A.; Lee, D.; Lunkenheimer, P.; Griesinger, C. *Angew. Chem., Int. Ed. Engl.* **2011**, *50*, 11437−11440.
(42) Peters, J. H.; de Groot, B. L. *PLoS Comput. Biol.* **2012**, *8*, e1002704.
(43) Richter, B.; Gsponer, J.; Várnai, P.; Salvatella, X.; Vendruscolo, M. *J. Biomol. NMR* **2007**, *37*, 117−135.
(44) De Simone, A.; Richter, B.; Salvatella, X.; Vendruscolo, M. *J. Am. Chem. Soc.* **2009**, *131*, 3810−3811.
(45) Salvatella, X.; Richter, B.; Vendruscolo, M. *J. Biomol. NMR* **2008**, *40*, 71−81.
(46) Cheng, T. M. K.; Blundell, T. L.; Fernandez-Recio, J. *BMC Bioinformatics* **2008**, *9*, 441.
(47) Fernández-Recio, J.; Totrov, M.; Abagyan, R. *J. Mol. Biol.* **2004**, *335*, 843−865.
(48) Grosdidier, S.; Pons, C.; Solernou, A.; Fernández-Recio, J. *Proteins* **2007**, *69*, 852−858.
(49) Pons, C.; Solernou, A.; Perez-Cano, L.; Grosdidier, S.; Fernandez-Recio, J. *Proteins* **2010**, *78*, 3182−3188.
(50) Vajda, S. *Proteins* **2005**, *60*, 176−180.
(51) Pons, C.; Glaser, F.; Fernandez-Recio, J. *BMC Bioinformatics* **2011**, *12*, 378.
(52) Showalter, S. A.; Brüschweiler, R. *J. Am. Chem. Soc.* **2007**, *129*, 4158−4159.
(53) Lindorff-Larsen, K.; Maragakis, P.; Piana, S.; Eastwood, M. P.; Dror, R. O.; Shaw, D. E. *PLoS ONE* **2012**, *7*, e32131.
(54) Kuzmanic, A.; Zagrovic, B. *Biophys. J.* **2010**, *98*, 861−871.
(55) Rajamani, D.; Thiel, S.; Vajda, S.; Camacho, C. J. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 11287−11292.