Modeling and minimizing CAPRI round 30 symmetrical protein complexes from CASP-11 structural models

Marwa El Houasli, Bernard Maigret, Marie-Dominique Devignes, Anisah W Ghoorah, Sergei Grudinin, David Ritchie

Modeling and Minimising CAPRI Round 30 Symmetrical Protein Complexes from CASP-11 Structural Models

Marwa El Houasli,

Bernard Maigret,

Marie-Dominique Devignes,

Anisah W. Ghoorah,

Sergei Grudinin,

David W. Ritchie*

September 29, 2016

Key Words

Protein docking; protein modeling; symmetrical oligomers; CAPRI; CASP.

Running header

Docking Symmetrical CASP/CAPRI Targets

Abstract

Many of the modeling targets in the blind CASP-11/CAPRI-30 experiment were protein homo-dimers and homo-tetramers. Here, we perform a retrospective docking-based analysis of the perfectly symmetrical CAPRI Round 30 targets whose crystal structures have been published. Starting from the CASP "stage-2" fold prediction models, we show that

---
*Corresponding author: e-mail: dave.ritchie@inria.fr, Tel. +33 3 83 59 30 45, Fax +33 3 83 41 30 79

using our recently developed "SAM" polar Fourier symmetry docking algorithm combined with NAMD energy minimisation often gives acceptable or better 3D models of the target complexes. We also use SAM to analyse the overall quality of all CASP structural models for the selected targets from a docking-based perspective. We demonstrate that docking only CASP "centre" structures for the selected targets provides a fruitful and economical docking strategy. Furthermore, our results show that many of the CASP models are dockable in the sense that they can lead to acceptable or better models of symmetrical complexes. Even though SAM is very fast, using docking and NAMD energy minimisation to pull out acceptable docking models from a large ensemble of docked CASP models is computationally expensive. Nonetheless, thanks to our SAM docking algorithm, we expect that applying our docking protocol on a modern computer cluster will give us the ability to routinely model 3D structures of symmetrical protein complexes from CASP-quality models.

Introduction

The 3-dimensional (3D) structures of many proteins in the Protein Data Bank (PDB) [1] are symmetric multimers [2]. While a large majority of the known protein multimers are homo-dimers and homo-tetramers, proteins have been observed to crystallise in a surprisingly large subset of the available cyclic, dihedral, tetrahedral, octahedral, and icosahedral point group symmetries. Thus, modeling the quaternary structures of symmetrical homo-oligomeric protein complexes provides an interesting and relevant challenge for protein structure modellers. In 2014, the CAPRI (Critical Assessment of PRedicted Interactions) blind docking experiment was directly linked to the CASP (Critical Assessment of Protein Structure Prediction) experiment [3]. While CASP aims to assess the state of the art in modeling the three-dimensional tertiary (3D) structures of proteins from their sequences, CAPRI serves a similar role for assessing algorithms to model the quaternary structures of protein-protein complexes from knowledge of their 3D structures. Both CASP and CAPRI represent genuinely blind 3D modeling challenges because the 3D crystallographic structures of the target structures are not made available to the participating groups until after each prediction round.

Many of the targets in Round 30 of CAPRI (CASP Round 11) were predicted by the CASP organisers to be symmetric homo-dimers, and were initially presented as CASP fold prediction targets. Thus, for the first time, the CAPRI docking community had the opportunity to predict the structures of the final complexes using the monomer structures predicted by the CASP community. During Round 30, we took this opportunity to test the ability of our "Kbdock" database of protein domain-domain interactions [4] to find homology templates, and the ability of our recently developed *ab initio* "SAM" (Symmetry AsseMbler) polar Fourier symmetry docking algorithm [5] to predict protein complexes by symmetry-constrained shape-based docking. We used Kbdock [4] to search for Pfam families which involved the given target domains and which had examples of structural homo-dimers in the PDB, and we used SAM to calculate symmetrical complexes according to the predicted stochiometry that was provided by the CASP organisers. However, due to the short time period available for each target, and because SAM was still under development, we attempted only 10 of the CAPRI targets in Round 30. According to the CAPRI assessment criteria, our approach found 2 acceptable solutions for 2 of the C_2 targets (T80 using Kbdock, and T86 using SAM). We also found low root means squared deviation (RMSD) solutions for targets T84 (Kbdock) and T87 (SAM) but these were rejected by the CAPRI assessors because their numbers of steric clashes exceeded the CAPRI

threshold. Nonetheless, these promising results showed that for several model-built CASP monomers, Kbdock and SAM were able to produce good C_2 docking models which could potentially be refined by molecular mechanics energy minimisation.

Here, we present a retrospective docking-based analysis of the Round 30 symmetrical targets for which the crystal structures have since been published in order to identify an *ab initio* strategy for modeling and refining symmetrical homo-multimers from CASP models. In particular, we investigate how many of the CASP monomer models are "dockable" by SAM, and how many SAM docking models should be energy-minimised in order to have a good chance of achieving an acceptable clash-free model of the biological complex.

Methods

The Symmetrical CAPRI Round 30 Targets

In CASP Round 11, some 30 predictor groups each submitted up to 5 models for each CASP target, giving up to 150 potential starting structures with which to model the corresponding complexes in CAPRI. Here, we consider only the CASP "stage-2" models, which represent a final set of high quality models, as assessed by the CASP organisers. These models were downloaded from the CASP web site [6].

Because SAM uses knowledge of point group symmetry to generate perfectly symmetrical homo-multimers, we analysed the published PDB structures for the Round 30 complexes [3], and we collected a set of 8 C_2 (two-fold cyclic symmetry) and 3 D_2 (two-fold dihedral symmetry) targets for the current study. These targets are summarised in Table 1. More detailed descriptions are provided by Lensink *et al.* [3]. Note that we consider a complex to be symmetric only if its monomers differ by no more than 2 amino acid residues, and if it can be superposed by a symmetry operation to within 1 Å RMSD or less. Several other CASP/CAPRI target complexes were called as homo-dimers or homo-tetramers by the CASP organisers but their PDB solution structures were found to be not strictly symmetric under the above criteria. While such complexes can be estimated by SAM, our NAMD scripts were not designed for this possibility, and so these cases are excluded from the current study.

During Round 30, typically only around 3 days were allotted to model each target, making it impossible for us to dock all of the stage-2 CASP models in the time available. Therefore, for each target, we used our "Kpax" protein structure alignment program [7, 8] to superpose all of the CASP models and to select a consensus, or "centre", structure that

gave the greatest similarity to all other models according to the Kpax "M-score". This sequence-independent cut-off free similarity score is calculated as a normalised sum of 3D Gaussian overlaps for the aligned residues in each pair of structure. We then used SAM to dock only the centre structure for each target on the assumption that this structure represents the community consensus prediction of the true 3D fold. As the crystallographic structures of the targets have since been made available, Table 1 also shows the M-score calculated for each selected centre model and the crystallographic structure to give a simple numerical indication of how well the selected centre model describes the target structure.

Table 1 also shows the CASP model and corresponding Kpax alignment statistics for the most similar model to the crystal structure ("Best PDB Alignment"). It is interesting to note that the calculated centre structure for T69 corresponds directly to the model most similar to the PDB solution structure (SAM_T08-server_TS3). In the remaining cases, the centre structures have M-scores greater than 0.6 in all except 3 cases (T77/C2, T94/C2 and T70/D2), indicating that the selected consensus structure represents a good way to pick a single representative model for the target monomer. However, the poor alignment statistics for T70, T77 and T94 show that these structures are difficult to model completely, with only 50–70% of these models being superposable onto the PDB crystal structure.

[Table 1 about here.]

Shape-Based SAM Symmetry Docking

While the majority of the CAPRI Round 30 targets involved C_2 symmetry, the SAM algorithm can construct protein complexes with arbitrary point group symmetry [5]. SAM uses a symmetry-constrained search using the "Hex" shape-based scoring function [9] to assemble cyclic C_n systems from a given protein monomer. Assuming the monomers are rigid, it can be shown that C_n systems have just 4 internal degrees of freedom, consisting of 3 rotations and one intermolecular translation, which need to be determined. Like Hex, SAM uses a spherical polar Fourier representation of protein shape using polynomial expansions to order $N=30$, and it uses an icosahedral tessellation of the sphere with 812 sample vertices to generate angular samples for the first 2 rotational degrees of freedom with an angular separation of approximately 7.5°. For each spherical sample, SAM performs an equi-spaced rotational search in the third rotational degree of freedom about the intermolecular axis between a pair of monomers. For C_n complexes, the axial rotational search is limited to by symmetry to the range [0:180°], and is accelerated by a one-dimensional (1D) fast Fourier

transform using 64 steps of approximately 2.8°. In order to cover the final translational degree of freedom, the 3D angular search is repeated for up to 64 translational steps of 0.8 Å, starting from an initial inter-monomer distance estimated from the monomer radius.

SAM builds D_2 complexes by performing a 2-dimensional Fourier domain rotational-translational search to dock the top 100 previously assembled pairs of C_2 sub-units. In this way, every SAM docking orientation is guaranteed to have precisely the desired symmetry. By exploiting the reduced search space, SAM is very fast, taking around 20 seconds for C_n complexes and 30 seconds for D_n complexes on a modern workstation. However, since the SAM scoring function is based only on surface skin volume overlaps, its docking solutions may contact physically forbidden steric clashes which should subsequently be energy-minimised using a more sensitive force-field model.

NAMD Energy Minimisation with Implicit Solvent

For each structure to be energy-minimised, 6,400 steps of conjugate gradient descent was applied using NAMD [10] with the CHARMM-22 force field and generalised Born implicit solvent model. We believe that at least this number of minimisation steps is necessary for stable convergence to a local energy minimum. Before running NAMD, any non-standard amino acids (e.g. MSE, MLY, LLP) were converted to standard types (i.e. MET and LYS), and all structures were prepared using VMD [11]. Any CASP model for which VMD detected improper or non-contiguous backbone geometry was eliminated from further consideration. Each NAMD minimisation run takes around 20 minutes on a 8-processor workstation.

Ligand RMSD-Based Quality Assessment

In pair-wise protein protein docking, it is conventional to call the larger protein the "receptor" and the smaller protein the "ligand". One way to measure the quality of a docking model is to fit the docked receptor C_α atoms to the native complex and then to calculate the deviation between the correspondingly transformed ligand C_α atoms and their coordinates in the native structure. This defines the ligand RMSD. The standard CAPRI assessment criteria for pair-wise complexes combine this ligand RMSD with a similarly calculated "interface RMSD" along with the fraction of native contacts ("fnat" score) in order to assign the quality of a docking model as either "high", "medium", "acceptable", or "incorrect". Docking models that contain more than a certain number of steric clashes are disqualified (full details are provided by Lensink *et al.* [3]). However, in symmetrical complexes, it is

not possible to uniquely assign distinct receptor and ligand monomers, and complexes with symmetries higher than C_2 can contain more than a single interface. Furthermore, because different CASP modeling approaches treat non-standard residues in different ways, different CASP models can contain different numbers of modeled interface residues. Therefore, because it is not obvious how one might sum or average the pair-wise CAPRI criteria over multiple interfaces in a fair way, and because in soft docking we do not wish to disqualify solutions with steric clashes, we use here only the ligand RMSD as a single measure of docking quality. For example, for C_2 dimers, we arbitrarily label the 2 monomers as "A" and "B", and we calculate the ligand RMSD by fitting on the A monomers and calculating the RMSD between the B monomers, as above.

Since a D_2 structure may be considered as a symmetrical dimer of 2 C_2 structures, it is easy to see that there are at least 2 distinct "ligand" environments with respect to the chosen A monomer. Therefore, for D_2 structures we label the next 2 monomers as "D" and "E", and we calculate the ligand RMSD as the lesser of the RMSD of the B+D and B+E monomer pairs after fitting the A monomers. For the sake of simplicity, we then consider a docking solution having such a ligand RMSD within 10 Å (i.e. the same upper limit as in CAPRI) of the native structure to be an "acceptable" (or better) docking solution. It is worth noting that several CASP models contain chain breaks which cause VMD to split them into multiple segments prior to NAMD minimisation. Therefore, we used in-house scripts to re-assemble the expected number of model chains from the known target stochiometry. We then used SAM in scoring mode to compute the NAMD ligand RMSD with respect to the native structure, as described above.

Results and Discussion

Table 2 shows the rank and RMSD of the first acceptable solution obtained when docking the crystallographic monomers from the PDB and our Kpax-calculated CASP centre models for the selected CAPRI targets. Clearly, docking PDB monomers represents the most favourable case for any docking algorithm since the monomers are already in the correct conformation of the target complex. As might be expected, Table 2 shows that SAM is able to generate acceptable or better complexes for many of the targets when starting from the PDB monomers, and that these solutions can be successfully refined using NAMD energy minimisation. SAM fails to find any solution in the top 50 for only 2 out of the 11 targets, namely T86/C2 and T71/D2.

Although it is to be expected that re-assembling crystallographic monomers should give

rather good docking solutions, it is interesting to see from Table 2 that the simple expedient of only docking CASP centre molecules leads to 4 acceptable models out of 11 in the top 32 SAM solutions. Furthermore, using NAMD to refine the top 50 SAM models for the CASP centre models gives a total of 6 acceptable or better solutions in the top 10 and 7 in the top 20. These results demonstrate that docking only the CASP centre model is a rather successful strategy, as well as being computationally economical.

[Table 2 about here.]

Since docking centre molecules can often lead to good models of the target complexes, it is interesting to ask how many of the CASP models might give rise to acceptable or better docking solutions. Therefore, we ran SAM on all of the CASP models and we extracted the 10 SAM solutions having the lowest energy for each model (giving a total of 1,618 SAM docking runs with up to 1,500 SAM docking solutions per target). Figure 1 shows scatter plots of SAM energy versus RMSD for these docking solutions. These plots show that an ensemble of candidate complexes gathered from the top 10 SAM docking solutions produced from each CASP model for each predictor group (normally 5 models per predictor group) always contains one or more acceptable docking solutions. Table 3 shows in more detail the numbers of acceptable SAM solutions when considering the top 10, 20, and 50 SAM solutions for each target. Despite the difficulty of modeling some of the CASP folds (in particular T70, T77, and T94: see Table 1), Table 3 confirms that for all targets except T71 (D_2), the ensemble of CASP stage-2 models always contain at least some, and often many, models that lead to acceptable complexes when docked by SAM.

[Figure 1 about here.]

[Table 3 about here.]

Table 3 also shows for each target the rank and RMSD of the CASP model which gives the best SAM docking solution. This shows that for 10 out of 11 targets, there exists a CASP model that leads to an acceptable complex within the top 5 solutions. The only target for which SAM failed to find any acceptable docking solutions is the tetrameric T71 complex. Table 1 shows that the monomer for this target could be sucessfully modeled by the CASP community, with one of the BAKER-ROSETTASERVER_TS3 models giving amongst the best model quality statistics of all 11 targets studied here (Table 1). However, the analysis of Lensink et al. [3] notes that modeling the T71 tetramer was particularly challenging during Round 30. This complex has 3 different domain-domain interfaces, all of

which are relatively small (from 470 to 720 Å^2), and while some predictor groups found one or other of the main interfaces, no predictor group produced a model containing all of the correct interfaces together [3]. Given that SAM uses a shape-based scoring function which favours large interfaces it is thus not particularly surprising that SAM also fails to dock this target, even when starting from the crystallographic monomer structure (Table 2).

Unfortunately, even for complexes with more favourable interfaces, it is not possible to know with certainty which particular CASP model structure might *a priori* lead to a good docking model. While Table 2 suggests that a simple strategy would be to use NAMD to minimise the top 5 or more SAM solutions for each CASP model, we estimate this would take at least 15 days per target on an 8-processor workstation, which would be rather impractical. One way to avoid this cost would be to use a statistical potential to further filter the docking solutions [12]. However, in our experience, statistical potentials are not sufficiently discriminating to distinguish good near-native conformations from large ensembles of conformations. Therefore, as a preliminary experiment, we energy-minimised in NAMD the top 50 SAM solutions gathered from the first 10 SAM solutions for each CASP model (Figure 1). This pulled out 2 good solutions for targets T80 and T87 (Rank=3, RMSD=3.88 Å from TASSER_VMT_TS4, and Rank=3, RMSD=2.33 Å from BAKER-ROSETTASERVER_TS2, respectively), but it failed to find any acceptable complexes in the top 10 for the remaining targets. This relatively low yield is perhaps not unsurprising, given that the CASP models for only these 2 targets yield favourable SAM energy-RMSD funnels (see Figure 1). Nonetheless, Figure 1 and Table 3 strongly suggest that rigorously energy-minimising more CASP-SAM models would be a successful docking strategy.

Overall, this study has shown that many of the CASP models can lead to acceptable or better models of symmetrical complexes when using NAMD to energy-minimise the top-10 solutions from each CASP-SAM docking model. However, applying our NAMD protocol to large numbers of candidate docking solutions involves a considerable computational cost. As mentioned above, the SAM solutions often contain steric clashes, and we believe that much of the time spent in NAMD is devoted to eliminating such clashes. It would therefore be interesting to test the utility of using shorter NAMD minimisation runs, or other more efficient refinement protocols and docking potentials, in order to detect and eliminate sterically infeasible solutions more rapidly. Long NAMD minimisations could then be applied only to the most promising solutions that remain. On the other hand, modern graphics processor units (GPUs) can considerably accelerate molecular dynamics simulations [13], and with the aid of only a modest number of GPUs we expect it would be feasible to apply the protocol described here in real time to future CASP-CAPRI targets.

Conclusion

We have investigated the ability of SAM shape-based symmetry docking and NAMD minimisation with implicit solvent to calculate the 3D structures of symmetrical protein complexes starting from 3D models from the CASP community. Our results show that *ab initio* docking of CASP models is often feasible. For the examples studied here, almost all of the CASP stage-2 models contain at least some 3D models that lead to acceptable shape-based SAM docking solutions. We have shown that docking only the Kpax-calculated CASP centre model for each target is a rather fruitful and economical docking strategy. However, it is still necessary to energy-minimise the SAM solutions using a more rigorous force-field model to pull out a near-native model of the complex.

We have also explored the possibility of using NAMD to pull out good solutions from large ensembles of CASP-SAM docking models. However, due to computational time constraints, we used NAMD to energy-minimise only the top 50 complex models gathered from the top 10 SAM solutions for each target. We expect that rigorously minimising a greater number of SAM models would lead to a greater number of acceptable docking solutions, but using NAMD to perform large-scale minimisations involves a considerable computational cost. Nonetheless, with the aid of our fast SAM docking algorithm, we expect that applying our protocol on a modest GPU cluster will give us the ability to routinely model and rigorously energy-minimise the 3D structures of symmetrical protein complexes from CASP models.

References

[1] HM Berman, T Battistuz, TN Bhat, WF Bluhm, PE Bourne, K Burkhardt, L Iype, S Jain, P Fagan, J Marvin, D Padilla, V Ravichandran, B Schneider, N Thanki, H Weissig, JD Westbrook, and C Zardecki. The protein data bank. *Acta Crystallographica*, D58:899–907, 2002.

[2] ED Levy, JB Pereira-Leal, C Chothia, and SA Teichmann. 3D Complex: a structural classification of protein complexes. *PLoS Computational Biology*, 2:e155, 2006.

[3] MF Lensink, S Velankar, A Kryshtafovych, S-Y Huang, D Schneidman-Duhovny, A Sali, J Segura, N Fernandez-Fuentes, S Viswanath, R Elber, S Grudinin, P Popov, E Neveu, H Lee, M Baek, S Park, L Heo, G Rie Lee, C Seok, S Qin, H-X Zhou, DW Ritchie, B Maigret, M-D Devignes, A Ghoorah, M Torchala, RAG Chaleil,

PA Bates, E Ben-Zeev, M Eisenstein, SS Negi, Z Weng, T Vreven, BG Pierce, TM Borrman, J Yu, F Ochsenbein, R Guerois, A Vangone, JPGLM Rodrigues, G van Zundert, M Nellen, L Xue, E Karaca, ASJ Melquiond, K Visscher, PL Kastritis, AMJJ. Bonvin, X Xu, L Qiu, C Yan, J Li, Z Ma, J Cheng, X Zou, Y Shen, LX Peterson, H-R Kim, A Roy, X Han, J Esquivel-Rodriguez, D Kihara, X Yu, NJ Bruce, JC Fuller, RC Wade, I Anishchenko, PJ Kundrotas, IA Vakser, K Imai, K Yamada, T Oda, T Nakamura, K Tomii, C Pallara, M Romero-Durana, B Jiménez-García, IH Moal, J Férnandez-Recio, JY Joung, JY Kim, K Joo, J Lee, D Kozakov, S Vajda, S Mottarella, DR Hall, D Beglov, A Mamonov, B Xia, T Bohnuud, CA Del Carpio, E Ichiishi, N Marze, D Kuroda, SS Roy Burman, JJ Gray, E Chermak, L Cavallo, R Oliva, A Tovchigrechko, and SJ Wodak. Prediction of homoprotein and heteroprotein complexes by protein docking and template-based modeling: A casp-capri experiment. *Proteins: Structure, Function, and Bioinformatics*, page In press, 2016.

[4] AW Ghoorah, M Smaïl-Tabbone, M-D Devignes, and DW Ritchie. KBDOCK 2013: a spatial classification of 3d protein domain family interactions. *Nucleic Acids Research*, 42:D389–D395, 2013.

[5] DW Ritchie and S Grudinin. Spherical polar fourier assembly of protein complexes with arbitrary point group symmetry. *Journal of Applied Crystallography*, 49:158–167, 2016.

[6] CASP-11. http://predictioncenter.org/download_area/CASP11/server_predictions/. Accessed: 2016-07-15.

[7] DW Ritchie, AW Ghoorah, L Mavridis, and V Venkatraman. Fast protein structure alignment using gaussian overlap scoring of backbone peptide fragment similarity. *Bioinformatics*, 28:3274–3281, 2012.

[8] DW Ritchie. Calculating and scoring high quality multiple flexible protein structure alignments. *Bioinformatics*, 32:in press, 2016.

[9] DW Ritchie and GJL Kemp. Protein docking using spherical polar fourier correlations. *Proteins: Structure, Function, Genetics*, 39(2):178–194, 2000.

[10] JC Phillips, R Braun, W Wang, J Gumbart, E Tajkhorshid, E Villa, C Chipot, RD Skeel, L Kale, and K Schulten. Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry*, 26:1781–1802, 2005.

[11] W Humphrey, A Dalke, and K Schulten. VMD – Visual Molecular Dynamics. *Journal of Molecular Graphics*, 14:33–38, 1996.

[12] P Popov and S Grudinin. Knowledge of native protein-protein interfaces is sufficient to construct predictive models for the selection of binding candidates. *Journal of Chemical Information and Modeling*, 55(10):2242–2255, 2015.

[13] JE Stone, DJ Hardy, IS Ufimtsev, and K Schulten. Gpu-accelerated molecular modeling coming of age. *Journal of Molecular Graphics and Modelling*, 29(2):116–125, 2010.

Figure 1 Caption

SAM energy-RMSD scatter plots for the top 10 SAM solutions from every CASP model for the 11 selected Round 30 CAPRI targets. X-axis: RMSD (Å); Y-axis: SAM docking energy (KJ/mol).

Target	CASP	PDB	#Res	#Model	Centre Model	M	#Align	RMSD	Best PDB Alignment	M	#Align	RMSD
T69/C2	T0764	4Q34	321	150	SAM-T08-server_TS3	0.825	258	1.67	SAM-T08-server_TS3	0.825	258	1.67
T72/C2	T0770	4Q69	453	150	3D-Jigsaw-V5_1_TS1	0.662	395	2.29	HHPredA_TS1	0.729	423	2.05
T75/C2	T0776	4Q9A	218	145	SAM-T08-server_TS4	0.815	200	1.83	nns_TS2	0.857	216	1.60
T77/C2	T0780	4QDY	198	150	FALCON_EnvFold_TS3	0.395	115	2.64	myprotein-me_TS5	0.426	116	2.37
T80/C2	T0801	4PIW	360	130	MULTICOM-CLUSTER_TS1	0.889	358	1.35	BAKER-ROSETTASERVER_TS1	0.912	360	1.27
T86/C2	T0815	4U13	109	150	FALCON_EnvFold_TS5	0.875	103	1.45	myprotein-me_TS3	0.905	105	1.31
T87/C2	T0819	4WBT	368	150	FALCON_TOPO_TS5	0.827	350	1.55	BAKER-ROSETTASERVER_TS2	0.871	361	1.44
T94/C2	T0852	4W9R	360	145	3D-Jigsaw-V5_1_TS3	0.510	260	2.54	Zhang-Server_TS5	0.541	277	2.50
T70/D2	T0765	4PWU	76	150	FALCON_TOPO_TS1	0.267	37	3.32	BAKER-ROSETTASERVER_TS3	0.826	75	1.76
T71/D2	T0768	4OJU	138	148	MULTICOM-CLUSTER_TS1	0.678	110	1.79	BAKER-ROSETTASERVER_TS3	0.883	135	1.38
T73/D2	T0772	4QHZ	236	150	3D-Jigsaw-V5_1_TS2	0.644	171	2.19	FFAS03_TS2	0.651	170	2.31

* In this table, CASP is the CASP identifier. PDB is the PDB structure of the crystallographic solution, #Res gives the number of residues in that structure, and #Model is the number of CASP models for that structure. Centre Model is the CASP model most similar to all other CASP models, as calculated by Kpax. M, #Align, and RMSD give the Kpax similarity score, number of superposed residues and their RMSD between the Centre model and the PDB structure. Note that only residue pairs within a 8 Å cut-off are included in the RMSD calculation. The "Best PDB Alignment" column shows the CASP model that was calculated by Kpax to be most similar to the PDB monomer. The final three columns give the corresponding Kpax alignment statistics.

Table 2: SAM and NAMD docking results for the crystallographic monomers and the Kpax-calculated CASP centre model.*

Target	PDB	SAM Rank	SAM RMSD	NAMD Rank	NAMD RMSD	Centre Model	SAM Rank	SAM RMSD	NAMD Rank	NAMD RMSD
T69/C2	4Q34_A	1	1.27	1	1.05	SAM-T08-server_TS3	8	3.18	1	1.01
T72/C2	4Q69_A	13	6.04	4	6.17	3D-Jigsaw-V5_1_TS1	13	5.55	1	6.18
T75/C2	4Q9A_A	3	2.47	1	3.68	SAM-T08-server_TS4	–	–	–	–
T77/C2	4QDY_A	2	4.44	4	4.30	FALCON_EnvFold_TS3	–	–	3	3.57
T80/C2	4PIW_A	1	1.23	2	1.24	MULTICOM-CLUSTER_TS1	32	5.75	–	–
T86/C2	4U13_A	–	–	–	–	FALCON_EnvFold_TS5	32	8.06	–	–
T87/C2	4WBT_A	1	0.98	1	0.58	FALCON_TOPO_TS5	–	–	1	0.58
T94/C2	4W9R_A	33	3.10	6	2.76	3D-Jigsaw-V5_1_TS3	–	–	9	2.75
T70/D2	4WPU_A	22	9.43	12	9.73	FALCON_TOPO_TS1	–	–	19	9.70
T71/D2	4OJU_A	–	–	–	–	MULTICOM-CLUSTER_TS1	–	–	–	–
T73/D2	4QHZ_A	12	8.86	7	3.61	3D-Jigsaw-V5_1_TS2	–	–	5	3.60

* This table shows the rank and RMSD of the first solution found within the top 50 solutions having a RMSD of 10 Å or less for the un-minimised SAM solutions, and when those solutions are energy minimised by NAMD using generalised Born implicit solvent. A hyphen denotes no solution in the top 50 is within 10 Å RMSD of the X-ray structure.

Table 3: Analysis of CASP protein models by SAM docking.*

Target	CASP-ID	#Top-10	#Top-20	#Top-50	Best SAM Energy	Rank	RMSD
T69/C2	T0764	57	82	137	FLOUDAS_SERVER_TS4	1	6.10
T72/C2	T0770	20	36	63	FALCON_TOPO_TS4	1	7.12
T75/C2	T0776	20	34	81	MULTICOM-NOVEL_TS4	1	6.99
T77/C2	T0780	2	4	18	FALCON_MANUAL_TS4	1	6.59
T80/C2	T0801	117	164	268	BAKER-ROSETTASERVER_TS3	1	2.85
T86/C2	T0815	18	44	99	MULTICOMCONSTRUCT_TS5	2	6.59
T87/C2	T0819	108	165	260	BAKER-ROSETTASERVER_TS4	1	3.57
T94/C2	T0852	4	7	10	ZHOU-SPARKS-X_TS2	2	6.27
T70/D2	T0765	5	9	17	MULTICOM-CONSTRUCT_TS4	5	8.18
T71/D2	T0768	0	0	0	–	–	–
T73/D2	T0772	2	2	4	QUARK_TS3	2	8.55

* This table summarises the results obtained when using SAM to dock all of the CASP models for each CAPRI target (a total of 1,618 SAM docking runs). For each SAM docking run, the first 50 solutions were compared to the X-ray structure in order to count the number of high-ranking solutions within 10 Å RMSD of the crystal structure. For example, from the 150 CASP models for T69, SAM produced at total of 57 solutions within 10 Å RMSD of the PDB structure which were ranked within the top-10 solutions for the corresponding model. The final three columns show the CASP model which gave the best SAM docking energy along with its rank with respect to the other SAM solutions for that model and its RMSD with respect to the crystal structure.

Figure 1: