



**HAL**  
open science

## Using Restraints in EROS-Dock Improves Model Quality in Pairwise and Multicomponent Protein Docking

Maria Elisa Ruiz Echartea, David Ritchie, Isaure Chauvot de Beauchêne

### ► To cite this version:

Maria Elisa Ruiz Echartea, David Ritchie, Isaure Chauvot de Beauchêne. Using Restraints in EROS-Dock Improves Model Quality in Pairwise and Multicomponent Protein Docking. *Proteins - Structure, Function and Bioinformatics*, 2020, 88 (8), pp.1121-1128. 10.1002/prot.25959 . hal-02930827

**HAL Id: hal-02930827**

**<https://hal.science/hal-02930827v1>**

Submitted on 4 Sep 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Using Restraints in EROS-Dock Improves Model Quality in Pairwise and Multicomponent Protein Docking

## Running Title

EROS-DOCK multi-body and with restraints

## Authors

Maria Elisa Ruiz Echartea,  
David W. Ritchie,  
Isaure Chauvot de Beauchêne

## Corresponding Author

isaure.chauvot-de-beauchene@loria.fr

## Acknowledgments

M. E. Echartea Ruiz is funded by a CORDI-S (Inria) doctoral contract.

## Abstract

Protein docking algorithms aim to predict the 3D structure of a protein complex from the structures of its separated components. In the past, most docking algorithms focused on docking pairs of proteins to form dimeric complexes. However, attention is now turning towards the more difficult problem of using docking methods to predict the structures of multi-component complexes. In both cases, however, the constituent proteins often change conformation upon complex formation, and this can cause many algorithms to fail to detect near-native binding orientations due to the high number of atomic steric clashes in the list of candidate solutions. An increasingly popular way to retain more near-native orientations is to define one or more restraints that serve to modulate or override the effect of steric clashes. Here, we present an updated version of our “EROS-DOCK” docking algorithm which has been extended to dock arbitrary dimeric and trimeric complexes, and to allow the user to define residue-residue or atom-atom interaction restraints. Our results show that using even just one residue-residue restraint in each interaction interface is sufficient

to increase the number of cases with acceptable solutions within the top 10 from 51 to 121 out of 173 pairwise docking cases, and to successfully dock 8 out of 11 trimeric complexes.

## Keywords

Protein docking; CAPRI; distance restraints; multi-body docking.

## Introduction

Protein docking aims to predict the three-dimensional (3D) structure of a protein complex starting from knowledge of the structures of the component proteins. In the past, most docking algorithms were designed to dock rigidly a single pair of proteins. However, because it is becoming increasingly evident that many protein complexes in Nature consist of more than two components, there is a growing interest in developing algorithms that can assemble multiple protein components in order to build the desired target complex. On the other hand, a confounding problem for many rigid body docking algorithms is that many proteins change their 3D conformations on binding. Thus, from a computational point of view, the docking problem may be considered as a 3D jigsaw puzzle in which none of the pieces fit together properly. A practical consequence of this is that, during a docking search, near-native candidate solutions may be rejected due to steric clashes between their surfaces.

An increasingly popular way to retain near-native orientations in a list of docking solutions is to define one or more restraints that serve to modulate or override the effect of steric clashes. Such restraints may often be derived from existing knowledge of catalytic residues, or from experimental data such as site-directed mutagenesis, or NMR chemical shift data. In the context of protein docking, several pairwise docking algorithms have been adapted to allow the user to define interaction restraints [8, 4, 2, 10, 15]. For instance, SwarmDock restricts the search to the surrounding area of the restraint residues, avoiding the starting points on the other side of the molecule [9]. ClusPro allows the user to define a range distance restraints for atom groups. During the docking process, for each rotation only translations that fulfill the restraints are kept to be evaluated. Finally, a total of 1,000 solutions that satisfy the restraints are clustered and minimized [8]. The HADDOCK approach uses “ambiguous interaction restraints (AIRs)” to describe the restraints. AIRs are defined by two kind of residues, active and passive. The active residues are those with high probability of belonging to the binding site, and the passives are the residue neighbors of

the binding site. As soon as two restraint atoms of the interacting proteins are in contact, the AIRs will contribute favorably to the scoring function [4]. In pyDock, an additional term is added to the scoring function which represents the percentage of satisfaction of distance restraints defined by the user [2].

On the other hand, only relatively few algorithms can deal with multiple components. We believe this is mainly due to the high combinatorial complexity that arises in the sampling stage and the additional programmatic complexity of keeping track of potentially vast numbers of candidate partial solutions. It is therefore common in multi-body docking approaches to dock the proteins by pairs as a first stage, and then to use the pairwise solutions to assemble higher order complexes. For example, DockTrina forms trimers using the combinations of transformations obtained from pairwise docking. It then takes one of the proteins as a reference and moves it by applying a combination of three transformations from the pairwise solutions. The RMSD between the transformed protein and its initial position is then used as a measure of the quality of the docked trimer, because in a near native trimer solution the moving protein should be transformed back onto itself [11].

As another example, CombDock creates spanning trees in a hierarchical way, the size of the trees being increasing at each stage. Then at each stage, parts of the new tree that were generated in previous stages are connected to generate a larger tree which is validated by checking the level of inter-penetration between the subunits [6].

Some multi-body docking algorithms can also use experimental information and/or bioinformatics data to deal with such problems have been used by some approaches with success. For example, HADDOCK using experimental data as explained above allows to model symmetric multicomponent assemblies composed of up to six sub-units [7]. Another example is 3D-MOSAIC, that relies on information obtained from the previous pairwise docking of the proteins involved in the target complex. This is useful to define the approximate location of the interaction interfaces, and to find suitable poses of the monomers during the formation of each possible target complex solution. In such a way that a monomer may occupy an interface if it does not cause clashes. Then similar poses of the new monomer regarding the already retained units are searched in the corresponding pairwise solutions. The possible solutions are ranked according the sum of their pairwise scores [3].

Recently we described EROS-DOCK [11], a docking algorithm that uses a new rotational sampling strategy based on sub-dividing a quaternion “ $\pi$ -ball”. EROS-DOCK rigidly docks two proteins using a series of exhaustive 3D rotational searches in which non-clashing orientations are scored using the ATTRACT coarse-grained force field model. The rota-

tional space is represented as a quaternion “ $\pi$ -ball”, which is systematically sub-divided in a “branch-and-bound” manner, allowing efficient pruning of rotations that will give steric clashes. Here we present an updated version of EROS-DOCK that allows to (i) use residue-residue or atom-atom interaction restraints as an additional pruning criteria and (ii) dock trimeric complexes by a combination of pairwise docking solutions. We report here the results of the new EROS-DOCK for pairwise docking on the protein benchmark v4, and for 3-body docking on a home-made benchmark of 11 cases.

## Methods

### Pairwise Docking with EROS-DOCK

Briefly, EROS-DOCK uses the ATTRACT coarse grained force field to reduce the high computational cost of calculating interaction energies at atomic resolution. We showed previously that most of the complexes in the Protein Docking Benchmark (v5) [16] have at least one pair of beads at their minimal energy distance in that force field. We therefore let each attractive pair of beads of the receptor and ligand surfaces corresponds to an initial docking pose. Conceptually, the ligand is then rotated about the current surface bead position, and a list of pairs of beads that clash and the corresponding list of 3D rotations that lead to steric clash is computed for each initial pose. However, rather than calculating clashes explicitly in Cartesian space, EROS-Dock uses the notion of “cone angle thresholds”, which are defined by applying a cosine rule to the triangle formed from the closest possible contact distance between two beads and the distances of those beads from the global rotation centre. If an arbitrarily rotated ligand bead makes a cone angle with a receptor bead that is less than the cone angle threshold for that pair, then it will certainly produce a steric clash. Thus, that particular rotation, *and any other similar rotations*, should be excluded from the search space. Hence, more generally, cone angles may be used to exclude regions of the 3D rotational space for a given pair of initial receptor and ligand bead contacts. This is achieved by using the subdivided quaternion  $\pi$ -ball representation of the search space, as mentioned above. In typical cases, when using this approach, around 95% of the search space may be eliminated from consideration. In other words, ATTRACT CG bead interactions need only be calculated for about 5% of the rotational search space [12].

## Defining Restraints in EROS-DOCK

Restraints are provided to EROS-DOCK by using a restraint file in which each line in the file defines one restraint. Restraints may concern residues or atoms pairs and their maximum separation distance. The minimum number of restraints to be satisfied may be specified by a command line parameter, the default value being one. Residue restraints are considered as satisfied as soon as one pair of atoms of the residues are separated by a distance less than or equal to the restraint distance.

## Defining Initial Docking Poses Using Restraint Specifications

Using the data from the restraints definition file, EROS-DOCK constructs a constraints  $\pi$ -ball data structure in a similar manner to the  $\pi$ -ball clash tree data structure. Furthermore, initial poses that will never satisfy the minimum number of restraints are discarded. Such useless initial docking poses are identified by computing the vector length differences between the pair of beads  $a$  and  $b$  of each restraint. Since such differences represent the minimal separation distance of the position vectors  $R_a$  and  $L_b$  if they are moved by any rotation as illustrated in Figure 1. As soon as a pair of beads of a restraint is at the restraint distance  $D$  or closer, such a restraint is marked as “possibly satisfied” at that initial pose. On the other hand, if all of the vector length differences are larger than  $D$ , logically such a restraint will never be satisfied. Hence, initial poses are discarded when the number of “possibly satisfied” restraints is less than the minimum required. This is illustrated in Figure 2.

## Multibody Docking

EROS-DOCK was adapted to assemble trimers by docking in a first stage all possible combinations of pairs of proteins involved in the multibody complex. Possible trimer solutions are assembled by fixing one protein, the “root-protein” (protein A, say) at the origin and by placing the other two around it using the transformations  $T_{AB}$  and  $T_{AC}$ , from the corresponding pairwise solution lists, as illustrated in Figure (a). Hence, one of the transformations  $T_{BC}$  yet to be determined. However, if the three transformations together form a near-native trimer, then it is natural to suppose that  $T_{BC}$  should be found in the list of B-C pairwise solutions.

More specifically, since the pairwise solutions were calculated independently, we may expect to find a matrix in the B-C list that *is similar to*  $T_{BC} = T_{AB} \cdot T_{AC}^{-1}$ , (see Figure (b),

(c). To search for such a matrix,  $T_{BC}$  is decomposed in a rotational part,  $RS_{BC}$ , and a translational part,  $TS_{BC}$ , to search for B-C nodes that contain similar rotations and rotations to  $T_{BC}$ . If  $RS_{BC}$  is inside the radius of some tree node  $N$ , the RMSD is computed between  $T_{BC}$  and the transformations composed by the rotation  $R$  and the translations stored in the node  $N$ . We use a threshold of 4 Å to recognize that such transformations are similar, and will therefore produce similar solutions. If no matching transformation is found at node  $N$ , the search will continue with its descendants.

Because the B-C list might contain several translations at each rotational node, this search is carried out using a separate 3D rotational tree for each set of pairwise docking solutions in order to reduce the computational cost. The search is performed three times, in such a way that every protein in the triplet is used as the root protein. At the end, the energy of the unknown interaction is computed to obtain the total energy of the triplet by adding the energies of the two other interactions two from the pairwise solution list.

In cases where combining transformations from pairwise docking A-B and B-C leads only to transformations A-C that are not found in the list of solutions for the corresponding pairwise docking, the combinations of A-B and B-C that provide the best global docking score (sum of A-B and B-C scores) are retained as best 3-body solutions. This way, correct solutions could in principle be found even for trimers where only two pairs of proteins are in contact.

## Results and Discussion

### Results of Pairwise Docking Applying Restraints

We docked 173 unbound complexes from the Protein Docking Benchmark (v4) with EROS-DOCK using one contact restraint. For each complex, a restraint was generated by randomly selecting one pair of residues that have at least one pair of atoms separated by a distance shorter than 5 Å after fitting the unbound structure on the bound complex. In EROS-DOCK, a rotational sub-space is discarded as soon as it is found that it will lead to a clash. However, the user can add a command line parameter to specify how many steric clashes can be produced by a sub-space before it is discarded. Thus, the benchmark was docked allowing the use of rotational sub-spaces that contain up to two, three and four clashes. As illustrated in the Figure 4, the number of near-native models grows when the number of clashes allowed increases. This is particularly noticeable for medium quality models.

As expected, Eros-Dock could find correct solutions for much more targets with restraints than without restraints, for almost all combinations of solution quality and number of top-ranked poses. The ranking of the first correct solution is especially greatly improved: Eros-Dock with restraints and allowing four clashes finds  $\approx 12\%/41\%$  more  $\star/\star\star$ -solutions in the 1000 top-ranked poses,  $\approx 50\%/69\%$  more  $\star/\star\star$ -solutions in the 100 top-ranked poses, and  $\approx 137\%/96\%$  more in the 10 top-ranked poses compared to docking without restraints. Regarding  $\star\star\star$ -solutions, the number of successes in the 10 top-ranked poses is increased from 0 to 3 by using restraints, but it is reduced from 7 to 6 in the 1000 top-ranked poses. A summary of program execution times is shown in Table 1.

## Results of Multibody Docking Applying Restraints

We present the results of multibody docking of 11 asymmetric trimers taken from the Protein Data Bank. We modeled the 3D unbound structure of the trimers by searching sequence homologous for each chain involved in the trimers using the HHpred tool from the Bioinformatics Toolkit of Max Planck Institute [14, 1], and by doing homology modeling with MODELLER [13]. If no unbound template could be found, we used a template from another structure of an homologous complex to create pseudo-unbound models. To reduce the resulting bias, if two pseudo-unbound models had to be created for the same complex, their templates were taken from different structures. Details about the trimers and the templates used to model the 3D structures are shown in the Table 2.

We defined one residue-residue restrain per interface in the trimer. We considered as hits those trimer solutions whose global RMSD is less or equal than 10 Å.

For 7 from the 11 complexes, the first hit was obtained within the top 100-ranked solutions, and for 5 within the top 50. Four of the complexes in the Table 3 are linear, and the 3 failed targets correspond to linear cases of the benchmark, in which two proteins are not in interaction. While EROS-DOCK is in principle able to retrieve 3-body docking solutions with only two interfaces, such configuration makes the docking obviously much harder if it is not known and not taken into account in the docking. Some experimental knowledge of the absence of interface between 2 of the 3 proteins could in principle be included in the docking process, which we will test in further studies.

Regarding the difficulty of the targets in terms of bound/unbound RMSDs, we could not find any correlation between the quality or rank of the hits obtained and the target difficulty. No clear correlation was found either between the quality of the results and the number of unbound / pseudo-unbound models in each trimer.



## Conclusion

We have presented an updated version of our EROS-DOCK protein docking algorithm, which can model protein complexes having up to three protein sub-components, and which allows the user to define simple restraints between pairs of residues at known or hypothesized protein-protein interfaces. We tested our approach on 173 pairwise complexes from the Protein Docking Benchmark (v4) [5] and on 11 trimeric complexes that we extracted from the PDB. Our results show how the used restraints increase the rank and quality of solutions. Moreover, for 7 of the trimeric complexes, a hit was obtained within the 100 top-ranked solutions.

## References

- [1] V. Alva, S.-Z. Nam, J. Söding, and A. N. Lupas. The mpi bioinformatics toolkit as an integrative platform for advanced protein sequence and structure analysis. *Nucleic acids research*, 44(W1):W410–W415, 2016.
- [2] V. Chelliah, T. L. Blundell, and J. Fernandez-Recio. Efficient restraints for protein-protein docking by comparison of observed amino acid substitution patterns with those predicted from local environment. *Journal of Molecular Biology*, 357(5):1669 – 1682, 2006.
- [3] M. Dietzen, O. V. Kalinina, K. Taškova, B. Kneissl, A.-K. Hildebrandt, E. Jaenicke, H. Decker, T. Lengauer, and A. Hildebrandt. Large oligomeric complex structures can be computationally assembled by efficiently combining docked interfaces. *Proteins: Structure, Function, and Bioinformatics*, 83(10):1887–1899, 2015.
- [4] C. Dominguez, R. Boelens, and A. M. J. J. Bonvin. HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. *Journal of the American Chemical Society*, 125:1731–1737, 2003.
- [5] H. Hwang, T. Vreven, J. Janin, and Z. Weng. Protein-protein docking benchmark version 4.0. *Proteins: Structure, Function, Bioinformatics*, 78(15):3111–3114, 2010.
- [6] Y. Inbar, H. Benyamini, R. Nussinov, and H. J. Wolfson. Prediction of multimolecular assemblies by multiple docking. *Journal of molecular biology*, 349(2):435–447, 2005.

- [7] E. Karaca, A. S. Melquiond, S. J. de Vries, P. L. Kastritis, and A. M. Bonvin. Building macromolecular assemblies by information-driven docking: introducing the haddock multibody docking server. *Molecular & Cellular Proteomics*, 9(8):1784–1794, 2010.
- [8] D. Kozakov, D. R. Hall, B. Xia, K. A. Porter, D. Padhorny, C. Yueh, D. Beglov, and S. Vajda. The ClusPro web server for protein–protein docking. *Nature Protocols*, 12(2):255, 2017.
- [9] I. H. Moal, R. A. Chaleil, and P. A. Bates. Flexible protein-protein docking with SwarmDock. In *Protein Complex Assembly*, pages 413–428. Springer, 2018.
- [10] B. G. Pierce, K. Wiehe, H. Hwang, B.-H. Kim, T. Vreven, and Z. Weng. ZDOCK server: interactive docking prediction of protein–protein complexes and symmetric multimers. *Bioinformatics*, 30(12):1771–1773, 2014.
- [11] P. Popov, D. W. Ritchie, and S. Grudinin. Docktrina: Docking triangular protein trimers. *Proteins: Structure, Function, and Bioinformatics*, 82(1):34–44, 2014.
- [12] M. E. Ruiz Echartea, I. Chauvot de Beauchêne, and D. W. Ritchie. Eros-dock: protein–protein docking using exhaustive branch-and-bound rotational search. *Bioinformatics*, 2019.
- [13] A. Šali, L. Potterton, F. Yuan, H. van Vlijmen, and M. Karplus. Evaluation of comparative protein modeling by modeller. *Proteins: Structure, Function, and Bioinformatics*, 23(3):318–326, 1995.
- [14] J. Söding, A. Biegert, and A. N. Lupas. The hhpred interactive server for protein homology detection and structure prediction. *Nucleic acids research*, 33(suppl\_2):W244–W248, 2005.
- [15] M. Torchala, I. H. Moal, R. A. G. Chaleil, J. Fernandez-Recio, and P. A. Bates. SwarmDock: a server for flexible proteinprotein docking. *Bioinformatics*, 29(6):807–809, 01 2013.
- [16] T. Vreven, I. H. Moal, A. Vangone, B. G. Pierce, P. L. Kastritis, M. Torchala, R. Chaleil, B. Jiménez-García, P. A. Bates, J. Fernandez-Recio, et al. Updates to the integrated protein–protein interaction benchmarks: docking benchmark version 5 and affinity benchmark version 2. *Journal of molecular biology*, 427(19):3031–3041, 2015.

## Tables

Table 1: Summary of EROS-DOCK execution times using restraints, grouped by number of clashes allowed.

Num. Clashes Allowed	Execution Time/min		
	2 Clashes	3 Clashes	4 clashes
Shortest Time	1.56	1.33	1.36
Longest Time	234.23	182	185.63
Average Time	20.16	22	18.4

Table 2: Trimeric targets and the templates used to model the (pseudo-)unbound forms.

Target	RMSD Bound - Model / Å	Template	unbound/pseudo-unbound	% Identities	Topology	Missing
6o07	Structure and mechanism of acetylation by the N-terminal dual enzyme NatA-Naa50 complex; Resolution: 2.702 Å					
6o07_A	5.9	6C9M_A	Pseudo-unbound	32	Triangular	-
6o07_B	1.9	2OB0_B	Unbound	25		
6o07_C	2.3	5ICV_A	Unbound	22		
6eqi	Structure of PINK1 bound to ubiquitin; Resolution: 3.1 Å					
6eqi_A	2.5	5L9U_S	Unbound	81	Triangular	-
6eqi_B	1.7	6OQ8_C	Unbound	63		
6eqi_C	3.2	5YJ9_D	Unbound	59		
6cp2	SidC in complex with UbcH7 Ub; Resolution: 2.9 Å					
6cp2_A	1.3	4TRH_B	Unbound	100	Triangular	-
6cp2_B	1.2	1WZV_A	Unbound	54		
6cp2_C	1.4	5L9U_S	Unbound	78		
6ath	Cdk2/cyclin A/p27-KID-deltaC; Resolution: 1.82 Å					
6ath_A	2.6	6GU2_A	Pseudo-unbound	64		

Table 2: Trimeric targets and the templates used to model the (pseudo-)unbound forms.

Target	RMSD Bound - Model / Å	Template	unbound/pseudo-unbound	% Identities	Topology	Missing
6ath_B	2.3	1W98_B	Pseudo-unbound	28	Triangular	-
6ath_C	7.3	1JSU_C	Pseudo-unbound	100		
5y6q	Structure of an aldehyde oxidase from <i>Methylobacillus</i> sp. KY4400; Resolution: 2.5 Å					
5y6q_A	1.4	1RM6_F	Pseudo-unbound	43	Triangular	-
5y6q_B	2.8	5G5G_B	Pseudo-unbound	40		
5y6q_C	2.4	2W55_F	Pseudo-unbound	24		
5wgb	Structure of the Human mitochondrial Cysteine Desulfurase in complex with ISD11 and <i>E. coli</i> ACP1 protein; Resolution: 2.75Å Å					
5wgb_A	2.1	3LVM_B	Unbound	57	Linear	A-C
5wgb_B	2.0	6GCS_P	Pseudo-unbound	22		
5wgb_C	2.1	6G2J_U	Pseudo-unbound	44		
5xfs	Structure of PE8-PPE15 in complex with EspG5 from <i>M. tuberculosis</i> ; Resolution:2.9 Å					
5xfs_A	1.5	4W4K_A	Pseudo-unbound	33	Linear	A-C
5xfs_B	3.5	2G38_B	Pseudo-unbound	33		
5xfs_C	3.2	5VBA_A	Unbound	24		
5xs5	Structure of Coxsackievirus A6 (CVA6) virus procapsid particle; Resolution: 3.3 Å					
5xs5_A	1.4	3VBR_A	Pseudo-unbound	60	Triangular	-
5xs5_B	2.1	1XYR_2	Unbound	55		
5xs5_C	4.0	5K0U_A	Pseudo-unbound	40		
6gwj	protein complex; Resolution: 1.95 Å					
6gwj_A	1.8	4WXA_C	Pseudo-unbound	24	Linear	B-C
6gwj_B	1.5	6GMH_C	Unbound	20		
6gwj_C	3.8	4K25_A	Pseudo-unbound	28		
6mac	Ternary structure of GDF11 bound to ActRIIB-ECD and Alk5-ECD; Resolution: 2.34 Å					
6mac_A	1.2	5NTU_A	Unbound	90	Linear	B-C
6mac_B	0.9	4FAO_F	Pseudo-unbound	99		

Table 2: Trimeric targets and the templates used to model the (pseudo-)unbound forms.

Target	RMSD Bound - Model / Å	Template	unbound/ pseudo-unbound	% Identities	Topology	Missing
6mac_C	3.4	1ES7_B	Pseudo-unbound	32		
6q84	Crystal structure of RanGTP-Pdr6-eIF5A export complex; Resolution: 3.7 Å					
6q84_A	9.1	3ZKV_A	Unbound	16	Triangular	-
6q84_B	2.5	1Z2A_A	Unbound	30		
6q84_C	1.4	5HY6_A	Unbound	65		

Table 3: Results of the docking of trimers using restraints, considering solutions with RMSD lower or equal than 10 Å as hits.

Target	Rank First Hit	Global LRMSD First Hit	Rank Best Hit	Global RMSD Best Hit	Num. Hits Top 100	Num. Hits Top 1000
6o07	23	8.94	40	7.61	4	14
6eqi	54	9.91	9643	6.72	2	30
6cp2	508	9.50	4005	8.03	0	1
6ath	10	7.93	336	5.49	6	99
5y6q	51	9.57	1938	6.73	2	5
5wgb	-	-	-	-	0	0
5xfs	-	-	-	-	0	0
5xs5	36	8.49	98	5.56	6	113
6gwj	73	8.98	1948	7.08	1	16
6mac	-	-	-	-	0	0
6q84	54	9.72	6618	6.10	3	40

## Figure Legends

Figure 1: Illustration of the application of restraints in EROS-DOCK. Shown are the position vectors  $R_{a1}$  and  $L_{b1}$  for the beads  $a1$  and  $b1$ , and the difference distance  $d_1$  between their vector lengths. Since  $d_1$  is shorter than the restraint distance  $D$ , such a restraint is marked as “possibly satisfied”. On the other hand, the pair of beads  $a2$  and  $b2$  do not satisfy the restraint since the distance  $d_2$  between their position vectors is greater than  $D$ .

Figure 2: Illustration of how distance restraints are represented in a quaternion  $\pi$ -ball search tree. The red nodes represent rotational sub-spaces that will not be used in the following stages of the docking since they do not satisfy the restraints. The right side of the figure shows how a rotation  $R$  of the node  $N$  is used to move the ligand beads of the restraints to test if  $R$  will satisfy the restraints. In general, one restraint definition can lead to many red (forbidden) nodes in a search tree.

Figure 3: General illustration of the construction of trimers: (i) Possible solutions are assembled using two transformations,  $T_{AB}$  and  $T_{AC}$ , from the corresponding pairwise solutions list, while A acts as the “root-protein”. (ii) The ensemble is transformed to place the centre of mass of B at the origin. (iii) Transformations  $T_{P1}$  to  $T_{Pn}$  are formed by applying the rotation  $R$  and translations  $T$  of tree node  $N$ . (iv) In this example, these transformations are compared to  $T_{BC}$  to see if some of them are similar.

Figure 4: Results from docking of the benchmark (v4) using one residue restraint. The graph shows the number of complexes for which at least one hit was obtained according to the CAPRI quality criteria, for EROS-DOCK without restraints and using restraints and allowing two, three and four clashes (EROS NR, EROS 2C, EROS 3C and EROS 4C, respectively).

# Figures

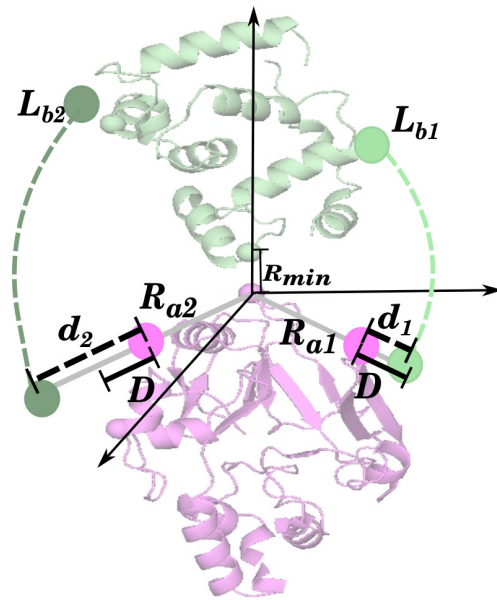


Figure 1:

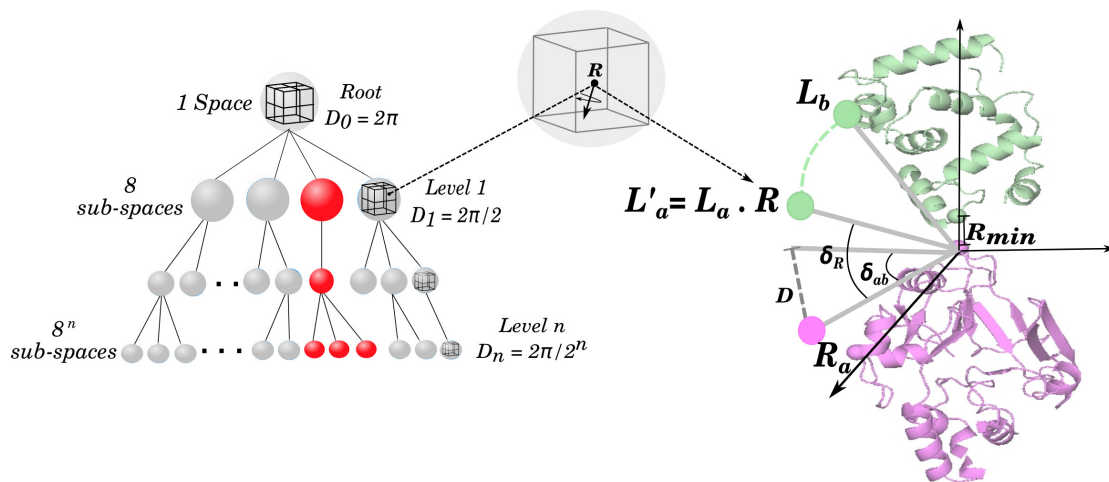


Figure 2:



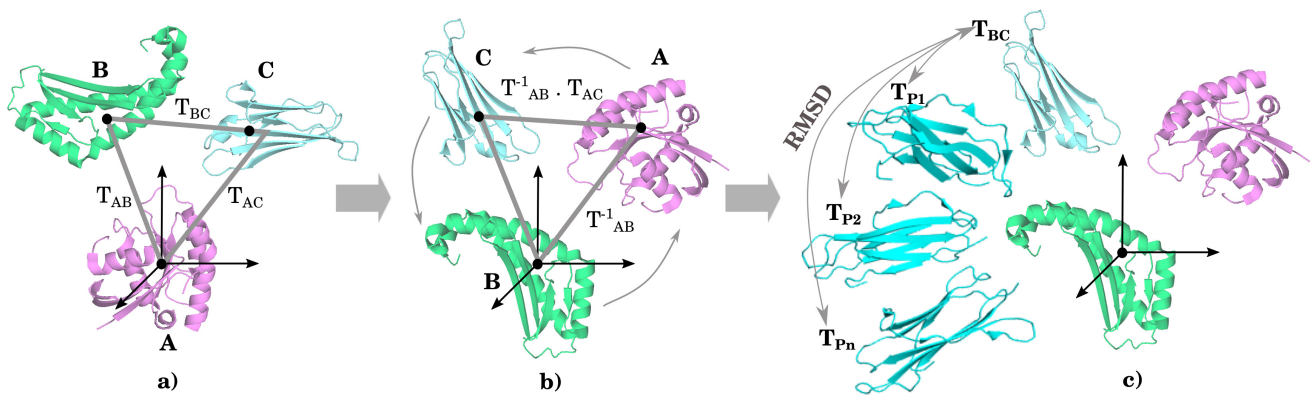


Figure 3:

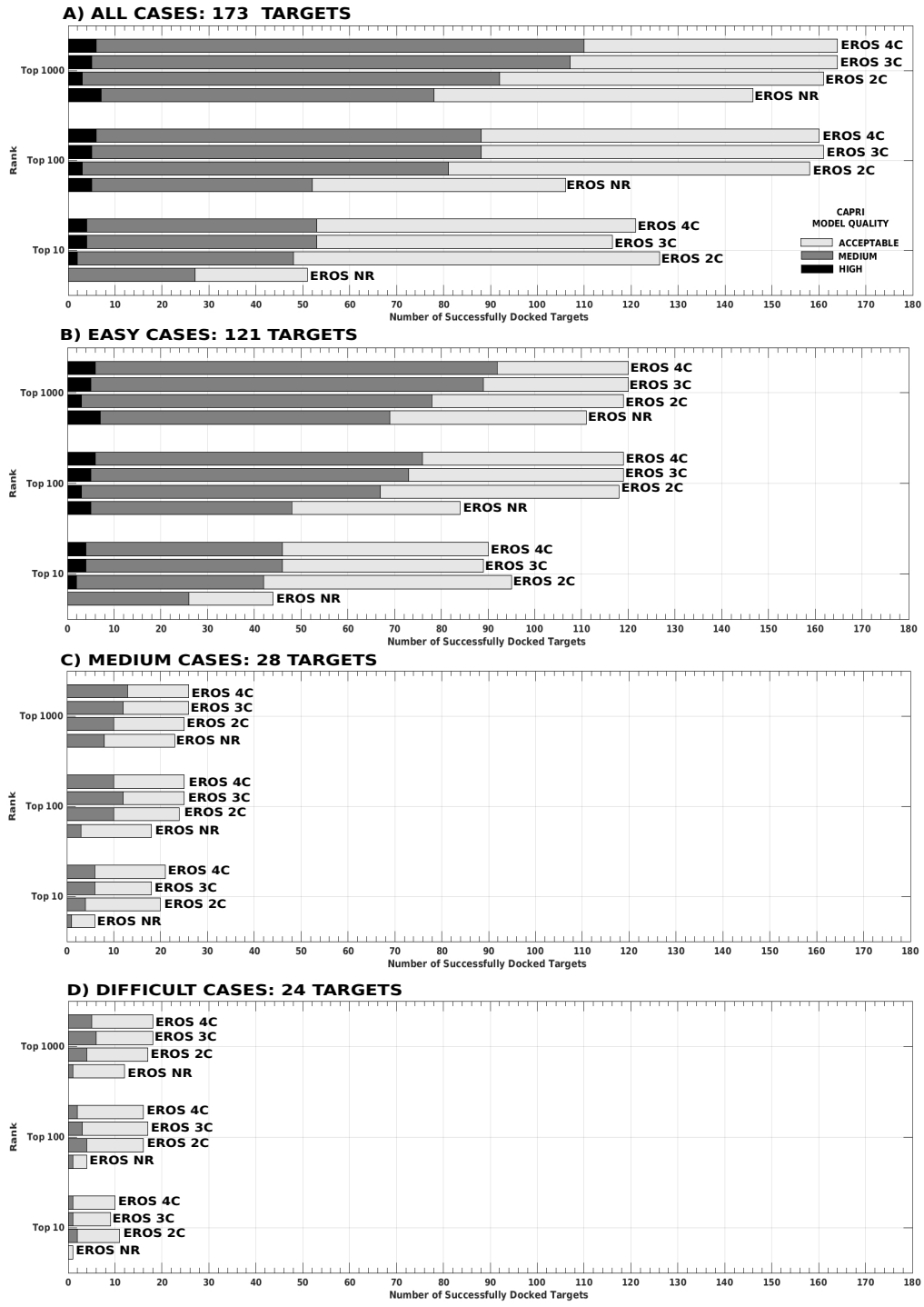


Figure 4: