EROS-DOCK: Protein-Protein Docking Using Exhaustive Branch-and-Bound Rotational Search
Maria Elisa Ruiz Echartea, Isaure Chauvot de Beauchêne, David Ritchie

To cite this version:
Maria Elisa Ruiz Echartea, Isaure Chauvot de Beauchêne, David Ritchie. EROS-DOCK: Protein-Protein Docking Using Exhaustive Branch-and-Bound Rotational Search. Bioinformatics, 2019, 35 (23), pp.5003-5010. 10.1093/bioinformatics/btz434. hal-02269812

HAL Id: hal-02269812
https://hal.science/hal-02269812
Submitted on 23 Aug 2019

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.
EROS-DOCK: Protein-Protein Docking Using Exhaustive Branch-and-Bound Rotational Search
Maria Elisa Ruiz Echartea, Isaure Chauvot de Beauchêne, David Ritchie

To cite this version:
Maria Elisa Ruiz Echartea, Isaure Chauvot de Beauchêne, David Ritchie. EROS-DOCK: Protein-Protein Docking Using Exhaustive Branch-and-Bound Rotational Search. Bioinformatics, Oxford University Press (OUP), 2019, 10.1093/bioinformatics/xxxxxx . hal-02269812

HAL Id: hal-02269812
https://hal.archives-ouvertes.fr/hal-02269812
Submitted on 23 Aug 2019

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.
Protein Docking by Branch-and-Bound Search

EROS-DOCK: Protein-Protein Docking Using Exhaustive Branch-and-Bound Rotational Search

Maria Elisa Echartea Ruiz,1 Isaure Chauvot de Beauchêne,1 and David W. Ritchie1,*

1University of Lorraine, CNRS, Inria, LORIA, 54000 Nancy, France.
* To whom correspondence should be addressed.
Associate Editor: XXXXXXX
Received on XXXXX; revised on XXXXX; accepted on XXXXX

Abstract

Motivation: Protein-protein docking algorithms aim to predict the 3D structure of a binary complex using the structures of the individual proteins. This typically involves searching and scoring in a six-dimensional space. Many docking algorithms use FFT techniques to exhaustively cover the search space and to accelerate the scoring calculation. However, FFT docking results often depend on the initial protein orientations with respect to the Fourier sampling grid. Furthermore, Fourier-transforming a physics-base force field can involve a serious loss of precision.

Results: Here, we present EROS-DOCK, an algorithm to rigidly dock 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 rotational space is represented as a quaternion “π-ball”, which is systematically sub-divided in a “branch-and-bound” manner, allowing efficient pruning of rotations that will give steric clashes. The algorithm was tested on 173 Docking Benchmark complexes, and results were compared with those of ATTRACT and ZDOCK. According to the CAPRI quality criteria, EROS-DOCK typically gives more acceptable or medium quality solutions than ATTRACT and ZDOCK.

Availability: The EROS-DOCK program is available for download at http://erosdock.loria.fr.

Contact: dave.ritchie@inria.fr.

Supplementary information: Supplementary data are available at Bioinformatics online.

1 Introduction

Protein-protein docking algorithms aim to predict how two proteins interact to form a complex. Docking algorithms usually involve two main tasks: (1) sampling the possible relative orientations of the two proteins, and (2) calculating an interaction energy or docking score at each position.

Although the protein docking problem has been studied for over 25 years, developing accurate and efficient protein docking algorithms remains a challenging problem due to the size of the search space, the approximate nature of the scoring functions used, and often the inherent flexibility of the protein structures to be docked (for reviews, see e.g. Halperin et al. (2002); Bonvin (2006); Ritchie and Kemp, 2000) and FRODOCK (Garzon et al., 2009) algorithms apply similar principles using spherical polar representations in order to accelerate the docking search in rotational coordinates. Other algorithms such as ZDOCK (Chen et al., 2003) and PIPER (Kozakov et al., 2006) use 3D FFT translational searches over scoring functions derived from knowledge of known protein-protein interfaces. On the other hand, docking algorithms such as HADDOCK (Dominguez et al., 2003) and ATTRACT (Zacharias, 2003) explicitly move and score atomistic or “coarse-grained” (CG) representations of the ligand and receptor using Monte-Carlo or gradient-based techniques to guide the search towards local energy minima. Thus, such approaches require multiple initial starting orientations in order to cover the 6D search space.
space. At each trial orientation, the receptor-ligand interaction energy is calculated using pair-wise distances between the relevant atoms or CG "beads", respectively. Thus, the scoring functions in such approaches may be considered as physics-based, rather than knowledge-based. However, since pair-wise atom or CG bead distances must be calculated explicitly in physics-based scoring functions, FFT-based acceleration cannot be used and the computational cost scales as $O(N^4 + M^4)$ per trial orientation for $N$ ligand and $M$ receptor atoms or beads, respectively. However, in ATTRACT, this cost can be greatly reduced by pre-calculating the receptor potential energies on a 3D grid (de Vries and Zacharias, 2017).

Here, we present a novel docking algorithm which retains the exhaustive nature of FFT-based search algorithms while still using a sensitive physics-based CG scoring function. However, rather than calculating an $O(N^4 + M^4)$ interaction energy explicitly at every grid point, we use a quaternion $\pi$-ball to represent the space of all possible 3D Euler angle rotations, and we recursively sub-divide the $\pi$-ball in order to cover the rotational space in a systematic way. It has been shown previously that there is a mapping between points in the $\pi$-ball space and Euler angle rotations, and that distances calculated between pairs of points in the $\pi$-ball are always greater or equal to the angular distances between the corresponding pairs of Euclidean space rotation matrices (Hartley and Kahl, 2009). In other words, coordinate distances in the quaternion $\pi$-ball representation provide upper bounds for the corresponding rotational distances in Euler angle rotation space. This important property has been exploited previously to develop efficient branch-and-bound based search algorithms for the problem of finding the optimal registration of two 3D point clouds (Chin et al., 2014; Bustos et al., 2014) which is a common problem in computer vision. In this paper, we apply for the first time a similar branch-and-bound based rotational search to the 6D rigid-body protein docking problem. However, instead of aiming to optimize the 3D registration of two objects represented by point clouds, our aim here is to find the global maximum of all possible pair-wise CG bead docking energies while simultaneously avoiding regions of the search space that lead to forbidden steric clashes. Since rigid body docking is essentially a 6D search problem, we divide the search space into multiple 3D rotational sub-problems, each of which can be treated in parallel using a separate $\pi$-ball search tree. The $\pi$-ball allows potentially very large regions of a 3D rotational search space to be pruned as soon as it can be established that any rotation within a well-defined sub-region of the search space will cause more than a given number of steric clashes.

In the current implementation of our approach, which we call "EROS-DOCK" (for Exhaustive Rotational Search based Docking), we use CG beads from the standard ATTRACT (attractive plus repulsive) CG force field model (Fiorucci and Zacharias, 2010), where the atoms of each amino acid are represented using from 2 to 4 beads. This reduces considerably the $O(N^4 + M^4)$ cost of each energy calculation compared to an all-atom representation. We use all attractive pairs of receptor and ligand surface beads to define the initial starting orientations for a full 6D docking search. Additionally, in order to detect steric clashes more efficiently, we pre-calculate "super-beads" from clusters of buried (i.e. non-surface) receptor and ligand beads.

EROS-DOCK has several advantages over existing FFT-based and real-space docking algorithms. In particular, (i) the chosen force field model is calculated exactly, since no potential grids (real or complex) are used; (ii) there are no limits on the sizes of the proteins to be docked; (iii) the protein starting orientations are determined automatically, and (iv) all docking runs are completely deterministic (i.e. two separate runs give identical results), since no random sampling is involved.

We tested EROS-DOCK on the structures of 173 protein-protein complexes taken from the Protein Docking Benchmark (v4) (Hwang et al., 2010) in order to compare our search algorithm with that of ATTRACT, since both EROS-DOCK and ATTRACT use exactly the same physics-based CG scoring function. Additionally, we also compared EROS-DOCK with ZDOCK, as an example of a widely used FFT-based docking algorithm. These tests were performed using the 3D structures of the unbound components of each target complex, and were assessed using the standard CAPRI quality classes (i.e. acceptable, medium, and high quality) (Méndez et al., 2003) with respect to the known crystal structures of the bound complexes.

2 Methods

2.1 Defining Initial Docking Contact Poses

It is reasonable to suppose that the interface in many protein complexes will have several pairs of ligand and receptor beads whose distances are close to the optimal distance for the corresponding bead types. Therefore, we first studied the distribution of ATTRACT CG bead distances in existing protein complexes in the Protein Docking Benchmark (v5) (Vreven et al., 2015). To do this, we used FATCAT (Godzik and Ye, 2004) to superposed each unbound structure onto its complex, and we calculated its intermolecular interaction energy curve. Because a deviation of only 0.2 Å between a trial orientation and the optimal bead distance may be considered to be negligible, and because it is almost certain that every protein complex will have at least one pair of such beads, it follows that all possible pairs of receptor and ligand attractive surface beads may be used to define a set of initial docking contact poses.

More specifically, for each such pair of receptor and ligand surface beads, $(i, j)$, the receptor bead $i$ is placed at the coordinate origin and the receptor’s centre of mass is placed on the negative $z$ axis. Similarly, ligand bead $j$ is placed on the positive $z$ axis at a distance $R_{min}$ from the origin, and the ligand’s centre of mass is placed on the positive $z$ axis. This is illustrated in Figure 2. Since the action of making the receptor and ligand centres of mass co-linear with the $z$-axis is purely for convenience, it can be seen that each placement of one pair of beads absorbs three degrees of freedom, thus leaving a purely 3D rotational search problem. Clearly, when starting a docking search from such an initial configuration, any rotation of the ligand about the coordinate origin will keep ligand bead $j$ in perfect contact with the receptor bead $i$.
2.2 3D Rotation Searches using a Quaternion π-Ball

Figure 5 illustrates the notion of a quaternion π-ball. In order to sub-divide this 3D rotational space, it is convenient to consider the π-ball as being inscribed in a cube of side 2π, in which any point within the π-ball may be mapped to an Euler rotation defined by the three Euler rotation angles, (α, β, γ). Points within the π-ball may be represented as a unit quaternion, \( Q = (\cos(\theta/2), \sin(\theta/2)u) \), where \( u \) is a unit vector from the coordinate origin, and \( \theta \) represents the radial distance from the origin. A mapping from the π-ball coordinate system to Euler rotation angles (α, β, γ) in conventional 3D space (using the “z-y-z” convention for Euler angle rotations) may be achieved by setting

\[
\alpha = \theta \quad \text{and} \quad u = (\sin \beta \cos \gamma, \sin \beta \sin \gamma, \cos \beta).
\]

Conceptually, a series of sample rotations is generated by dividing the initial π-ball into 8 cubes, and by then recursively sub-dividing each such cube into smaller cubes until a given angular threshold is reached. From 3D geometry, the distance \( \Delta_s \), from the centre of cube \( s \) to any one of its vertices is given by

\[
\Delta_s = \frac{\sqrt{3}}{2} \Delta_c,
\]

where \( \Delta_c \) is the length of the side of cube \( s \) (initially \( \Delta_0 = 2\pi \)). Thus, \( \Delta_s \) may be considered as the bounding radius of cube \( s \). At each iteration of an angular search, the centre of the \( s \)-ball may be calculated by successively doubling the leaf node cube size until \( \Delta_0 \geq 2\pi \). Any cube centre \( Q(L, u) \) having \( |u| > \pi \) represents an invalid rotation and is ignored. However, as described below, it is often not necessary to evaluate a docking energy for every node in the search tree.

2.3 Distance Measures in 3D Rotation Spaces

It is intuitively obvious that two similar quaternions or rotation matrices will rotate a given object into similar orientations. However, unlike ordinary Cartesian space, many different angular distance metrics may be defined for rotational spaces. Here, we take as our starting point the angular distance relations (see Lemmas 1 and 2 of Hartley and Kahl (2009))

\[
\theta(Q, Q') = \min(\theta(Q, Q_1), \theta(Q, Q_2)),
\]

where \( \theta(Q, Q_1), \theta(Q, Q_2) \) represents the Euclidean distance between a pair of quaternion points, \( Q, Q_1 \), in the π-ball space, \( R \) and \( R' \) represent the 3D rotation matrices that correspond to \( Q \) and \( Q' \) respectively, and \( \theta(R, R') \) represents the angle between two Cartesian space vectors. The functions \( \theta(Q, Q_1) \) and \( \theta(Q, Q_2) \) may be calculated as the inverse cosine of the dot product of the corresponding vector components, whereas the angular distance, \( d_{\theta}(Q, Q') \), may be calculated by extracting the angular part of the matrix \( M = R^{-1}R' \) in the axis-angle representation of \( M \) (Hartley and Kahl, 2009).

Since we are mainly concerned with how a sample 3D rotation matrix might move a ligand bead compared to some reference rotation, Equation 2 says that an upper bound for the angular difference in the bead’s positions after having applied the two rotations may be obtained from their quaternion representations in the π-ball. Conversely, if the angular distance \( d_{\theta}(\cdot, \cdot) \) between two rotation matrices is greater than the radius \( \Delta_s \) of a π-ball sub-cube, then it follows that the corresponding quaternion rotation coordinates must fall within different π-ball sub-cubes. This property is used to define a branch-and-bound search in rotational space.

2.4 Branch and Bounds Search using Bead Cone Angles

In order to prune the rotational search efficiently, we begin each 3D rotational docking search by building a list of all possible receptor and ligand attractive surface bead pairs, (a, b), and for each pair we use the corresponding ATTRACT potential energy curve to define a minimum allowed contact distance \( \sigma_{ab} \), such that a pair-wise bead distance less than \( \sigma_{ab} \) is considered a steric clash (see Figure 1). Letting \( R_a \) and \( L_a \) represent the position vectors of beads a and b, and letting \( R_a = L_a \) and \( L_a = L_b \) denote the corresponding vector lengths, then clearly beads a and b will never give a steric clash under any ligand rotation if \( |R_a - L_b| \geq \sigma_{ab} \). Otherwise, it will be necessary to calculate explicitly whether a particular rotation might cause a steric clash.

While steric clashes are commonly calculated according to a Euclidean distance threshold, here it is more convenient to work with angular distances. More specifically, we first use \( R_a \) and \( L_b \) to calculate the rotation \( R_{ab} \) that will place the ligand bead centre \( L_a \) as closely as possible to the centre of the receptor bead, \( R_a \). We call \( R_{ab} \) a “clash rotation”, because it will cause a steric clash if \( |R_a - R_{ab}| < \sigma_{ab} \) (see Figure 4(A)). A list of clash rotations may be calculated just once for each starting pose. Now, if \( R_{ab} \) causes a steric clash between beads a and b then there must exist an infinite number of sample rotations, \( R_{ab}^{b} \), which are “near” to \( R_{ab} \) and which will cause the ligand bead to sweep
out a cone in 3D space while remaining in contact with the receptor bead (Figure 4(B)). Hence we use the cosine rule to define a “cone angle”, \( \beta \), as

\[
\cos \beta_{ab} = (R_a^2 + R_b^2 - \alpha^2)/(2R_a R_b).
\]

Then, letting \( \omega \) represent the angular difference in the ligand position when rotated by a sample rotation \( R_c \) and its position when rotated by the clash rotation \( R_{\text{clash}} \), we have

\[
\omega = \theta(\hat{R}_a \cdot \hat{R}_b; \hat{R}_{\text{clash}} \cdot \hat{R}_b).
\]

In this way, we may compare the angles \( \omega \) and \( \beta_{ab} \) to determine whether the rotation \( R_c \) causes beads \( a \) and \( b \) to clash.

More importantly, since \( \Delta \alpha \) represents an upper bound on the angular difference between \( \hat{R}_a \) and any other point in sampling cube \( s \), then if \( \omega > \Delta \alpha \), we infer that the rotation \( R_{\text{clash}} \) must belong outside cube \( s \). In a similar manner, if \( \omega > \beta + \Delta \alpha \), we can infer that no rotation within cube \( s \) can cause a steric clash between beads \( a \) and \( b \) (see Figure 5(A)). Conversely, if \( \omega < \beta - \Delta \alpha \), we can infer that any rotation from within cube \( s \) will cause a steric clash between beads \( a \) and \( b \) (Figure 5(B)). Finally, as noted above, if \( \omega < \beta \), we infer that the rotation \( R_{\text{clash}} \) causes a steric clash between \( a \) and \( b \). However, in the context of a systematic search, subdividing cube \( s \) could yield further rotational samples that might not cause clashes.

In a similar manner, we note here that some sampling cubes may intersect the boundary of the \( \pi \)-ball. In such cases, if the centre of a cube lies outside the \( \pi \)-ball, then its rotational sample \( R_{\text{clash}} \) is not meaningful and is discarded. However, the cube remains a candidate for sub-division because the centres of some of its children may still correspond to meaningful rotations.

2.5 Coloring the 3D Rotation Search Tree

As indicated above, each node in the rotation search tree is visited recursively for each bead pair in order to color it according to whether it gives a steric clash or not. In order to eliminate sample rotations that lead to steric clashes as early as possible, we first use a simple clustering algorithm to assign any overlapping non-surface beads to a small number of “super-beads” (details not shown). These super-beads are then added to the list of potential clash pairs, and the list is sorted in order of decreasing cone angle because bead pairs having large clash cone angles are more likely to allow a node that always clashes to be detected and colored early in the search. Then, in a first pass, each pair of beads from the clash list is used to color the nodes in the tree according to whether a node always gives a steric clash or whether only its central sample rotation gives a clash. As soon as a node has been colored as “Always Clashing”, it and all of its children may be ignored by subsequent bead pairs, and a counter in the parent node is incremented. Thus, whenever all of the children of a given node are colored as Always Clashing, then the parent node is assigned Always Clashing as well.

2.6 Calculating Non-Clashing Docking Energies

After the clash status of each \( \pi \)-ball node has been determined, the tree is traversed once more to calculate exact ATTRACT energies for only the non-clashing nodes. The list of non-clashing orientations is then sorted by ATTRACT energy, and the top 100 solutions per \( \pi \)-ball are saved into a global list. Once all of the top 100 solutions per bead pair have been gathered in the global list, the global list is sorted and the top 50,000 orientations are saved as the best solutions found for that target complex.

2.7 Adjustable Parameters

EROS-DOCK has a small number of adjustable parameters. Here, we describe the three most important parameters. Firstly, as mentioned above, the angular search resolution, \( \alpha \), is the most important parameter, since the computational cost of the algorithm scales as \( \text{O}(1/\alpha^3) \). A large value of \( \alpha \) gives faster execution, but if \( \alpha \) is too large, then many good solutions will be missed. We currently use a default of \( \alpha = 7.5^\circ \), which corresponds to a resolution of \( 7.5^\circ \) in each of the three Euler rotation angles. Secondly, in order to calculate which beads are surface beads, the accessibility of each bead is initially tested by rolling a probe bead over the surface of each protein. From some early experiments, we determined heuristically that a good value for the rolling bead probe radius is 2.5 Å. Finally, again based on heuristic tests, we determined that up to two bead clashes may be tolerated per pair-wise orientation before calling a steric clash for the corresponding node of the \( \pi \)-ball. Note that clashes are calculated using the cone angle representation (Section 2.4) and that ATTRACT energies are calculated only for rotational orientations that pass the clash count threshold.

3 Results

3.1 Defining the Initial Docking Poses

Our study of protein-protein complexes from the Protein Docking Benchmark (v5) revealed that a large number of protein interfaces contain at least one pair of beads at almost the optimal distance, according to their distance-dependent interaction energy in the ATTRACT CG force field. For example, 90% of the benchmark complexes have at least one pair of receptor-ligand interface beads within 0.2 Å of their optimal separation.
residues and 3,414 starting orientations gave the shortest execution time of 4.33 min. On the other hand, the target 1MR8 has 863 residues and 93,442 surface bead pairs, and gave the longest execution time of 184.14 min. Table 2 shows the overall shortest, longest, and average execution times for each target category.

### 3.4 Comparing EROS-DOCK with ATTRACT and ZDOCK

Because EROS-DOCK uses the ATTRACT coarse-grained force field model, we first compare the results of EROS-DOCK with those of ATTRACT in order to study the effect of our new sampling strategy. However, because ATTRACT performs energy minimizations whereas EROS-DOCK does not, for a fairer comparison we apply energy minimizations using the ATTRACT toolkit to the top 50,000 solutions of each target docked by EROS-DOCK. These results are subsequently called EROS-MIN.

We also compare results with ZDOCK version 3.0.2 (Pierce et al., 2011) in order to examine the difference between the use of exhaustive CG sampling and regular FFT sampling using a pairwise statistical interaction potential. The results presented for ZDOCK were obtained using default parameters and random starting orientations for both receptor and ligand. Since EROS-DOCK performs dense rotational sampling, we also ran ZDOCK using its dense (6°) sampling option. However, the results were less favorable than using ZDOCK’s default 15° sampling mode. Therefore, we show here only ZDOCK results using 15° sampling.

For the ATTRACT runs, the ligand starting positions were generated by the standard ATTRACT search procedure which gave a set of points evenly distributed over the receptor surface (the actual number depends on the size of the receptor), and at a distance from the receptor surface that depends on the ligand’s radius of gyration. The ligand was placed on each starting point, and 228 ligand rotations were applied to generate approximately equally distributed ligand orientations. For each receptor starting position and ligand orientation, 1,000 minimization steps were applied using the ATTRACT force-field with grid acceleration, a final sum of pairwise atom-atom energies was calculated, the structures were ranked by ATTRACT energy, and redundant structures (RMSD < 0.2 Å) were discarded.

Figure 6 summarises the number of successfully docked targets obtained by ZDOCK, ATTRACT, EROS-DOCK, and EROS-MIN for the 173 benchmark complexes, as a function of the CAPRI docking quality criteria. For example, Figure 6(A) shows the distribution of targets having at least one acceptable, medium, or high quality docking solution within the ranks 1, 10, 100, and 1,000 for the 173 benchmark complexes. At each rank threshold the number of successful docking cases is represented by a bar. In the same way, Figures 6(B), (C), and (D) show the results according to the “easy”, “medium”, and “difficult” classifications, as determined by the Benchmark authors. More detailed results are provided as Supplementary Material.

It should be noted that in Figure 6, the total number of acceptable solutions includes the number of medium and high quality solutions. Hence, for example, Figure 6(A) shows that EROS-MIN found acceptable solutions ranked within the top 100 solutions for 156 out of 173 target complexes, of which 88 are also classed as medium quality solutions and 21 as high quality solutions. Because different proteins will often have different numbers of surface beads, EROS-DOCK generally calculates a different number of initial docking poses for each target complex. However, we did not find any relationship between the quality of the docking solutions and the number of starting poses (details not shown).

In general, Figure 6 (A) shows that EROS-MIN produces more acceptable solutions than the other algorithms, except at the top 10 where the results are comparable with those of EROS-DOCK and ZDOCK. This indicates that several of the basic EROS-DOCK solutions are close to
enough to a near-native local energy minimum to benefit from a subsequent minimization step. Regarding medium quality solutions, Figure 6 (A) shows that the performance of EROS-DOCK, EROS-MIN, and ZDOCK is generally comparable at each level, except that ZDOCK finds noticeably more acceptable solutions in the top 10 while ATTRACT generally finds fewer acceptable or better solutions. For high quality solutions, EROS-MIN performs better than the other methods.

Since EROS-MIN and ATTRACT use the same force field and scoring function, any difference in their performance must be due to their different sampling strategies. ATTRACT uses a heuristic sampling scheme, while EROS uses an exhaustive search. Therefore, we believe that ATTRACT is prone to miss some energy minima when the energy landscape fluctuates rapidly, but it will find the local minimum in each energy basin it explores. On the other hand, EROS-DOCK is less likely to miss basins, but will not find the minimum in each basin.

To investigate this further, we compared the energy of the top-ranked solutions found by ATTRACT and by EROS-DOCK before minimization. We found that EROS-DOCK finds solutions with lower energy than the lowest-energy solution of ATTRACT in 163 out of 173 cases (in 142/173 cases when considering only differences above 1 Kcal/Mol). These lower-energy solutions found by EROS-DOCK correspond to basins not explored by ATTRACT. This confirms that the better performance of EROS-MIN over ATTRACT is due to a more exhaustive initial sampling by EROS-DOCK, allowing to find more local minima after minimization of the low-energy basins found by EROS-DOCK.

When considering the results by target difficulty, Figure 6 (B) shows that the best solutions produced by each algorithm for the easy targets are mainly of acceptable and medium quality, and the number of successfully docked targets is comparable, especially among EROS-MIN, EROS-DOCK, and ZDOCK. On the other hand, EROS-DOCK (i.e. without minimisation) finds fewer high quality solutions than the other algorithms. For medium difficulty targets, Figure 6 (C) shows that the best solutions obtained by each algorithm are mainly of acceptable quality, and the number of successfully docked targets is again comparable. A similar profile of results is seen for the difficult targets (Figure 6 (D)). However, the total number of targets and number of high quality solutions obtained by any method for the medium and difficult target groups are generally quite small, making it difficult to make meaningful comparisons between the different algorithms. Nonetheless, it interesting to note that ATTRACT is the only algorithm to obtain high quality solutions in the top 1,000 for some difficult targets (Figure 6 (D)).

As mentioned above, energy minimizing the basic EROS-DOCK solutions (EROS-MIN) increased the number of targets with high quality models for the easy targets, and it increases the number of targets with acceptable or medium quality solutions at all the difficulty classifications. This demonstrates the utility of using EROS-DOCK as an exhaustive initial docking search engine to propose high quality trial orientations which could be refined using flexible docking procedures or short molecular dynamics simulations.

### 4 Conclusion

We have developed an exhaustive real space CG docking algorithm called EROS-DOCK. A novel feature of our approach is the use of a quaternion n-ball representation of 3D rotational space which allows the notion of branch-and-bound search to be applied for the first time to the protein docking problem. We have demonstrated that our branch-and-bound search using the ATTRACT CG force field model typically gives more acceptable or better solutions, especially when a final energy minimization step is applied, when compared to the well-known and highly optimised ATTRACT and ZDOCK docking programs.

### Table 2. Summary of EROS-DOCK execution times, grouped by benchmark category.

<table>
<thead>
<tr>
<th>Target</th>
<th>No. Residues</th>
<th>No. Starting Pairs</th>
<th>Execution Time / min</th>
</tr>
</thead>
<tbody>
<tr>
<td>Easy Targets</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shortest Time</td>
<td>ZO0B</td>
<td>111</td>
<td>3,414</td>
</tr>
<tr>
<td>Longest Time</td>
<td>19R</td>
<td>863</td>
<td>93,442</td>
</tr>
<tr>
<td>Average Time</td>
<td></td>
<td>453</td>
<td>29,323</td>
</tr>
<tr>
<td>Medium Targets</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shortest Time</td>
<td>1SYX</td>
<td>191</td>
<td>6,971</td>
</tr>
<tr>
<td>Longest Time</td>
<td>1BGX</td>
<td>1,230</td>
<td>195,965</td>
</tr>
<tr>
<td>Average Time</td>
<td></td>
<td>512</td>
<td>40,951</td>
</tr>
<tr>
<td>Difficult Targets</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shortest Time</td>
<td>1PXV</td>
<td>282</td>
<td>10,070</td>
</tr>
<tr>
<td>Longest Time</td>
<td>1DE4</td>
<td>1,641</td>
<td>245,456</td>
</tr>
<tr>
<td>Average Time</td>
<td></td>
<td>573</td>
<td>55,073</td>
</tr>
</tbody>
</table>

While the current implementation of EROS-DOCK is slower than ATTRACT, we believe there is scope to optimize the EROS-DOCK code and search parameters. Furthermore, we expect that our approach will be particularly suitable for docking very large protein structures that will have many more local energy minima than the examples studied here. We also expect that our angular search algorithm will be particularly useful when some knowledge of the interface residues is available (as in “data-driven” docking) and when multiple interfaces must be considered in multi-component docking problems. We are currently investigating both such possibilities.

### Acknowledgements

### Funding

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

### Conflict of Interest

None declared.
Fig. 6. Results obtained by EROS-DOCK, ATTRACT and ZDOCK for 173 unbound target complexes from the Protein Docking Benchmark (v4). The plots show the number of complexes docked with acceptable, medium, and high quality according to the CAPRI quality criteria.
References