

Article

SwarmDock and the Use of Normal Modes in Protein-Protein Docking

Iain H. Moal and Paul A. Bates *

Biomolecular Modelling Laboratory, Cancer Research UK London Research Institute, Lincoln's Inn Fields Laboratories, 44 Lincoln's Inn Fields, London, WC2A 3LY, UK

* Author to whom correspondence should be addressed; E-Mail: Paul.Bates@cancer.org.uk;
Tel.: +44-20-7269-3223; Fax: +44-20-7269-3258.

Received: 2 July 2010; in revised form: 29 July 2010 / Accepted: 16 September 2010 /

Published: 28 September 2010

Abstract: Here is presented an investigation of the use of normal modes in protein-protein docking, both in theory and in practice. Upper limits of the ability of normal modes to capture the unbound to bound conformational change are calculated on a large test set, with particular focus on the binding interface, the subset of residues from which the binding energy is calculated. Further, the SwarmDock algorithm is presented, to demonstrate that the modelling of conformational change as a linear combination of normal modes is an effective method of modelling flexibility in protein-protein docking.

Keywords: elastic network model; normal mode analysis; particle swarm optimisation; PSO; protein flexibility; RTB; CAPRI

1. Introduction

Protein-Protein interactions are fundamental to almost all biological processes. Disruption of molecular recognition is integral to many diseases including cancer. Docking attempts to predict the structure of complexes from their monomeric constituents. The computational approach has the potential to confirm or dismiss putative interactions as well as provide structural knowledge which can be exploited for the design of therapeutic interventions in a range of diseases, should a sufficiently accurate model be produced.

The docking problem presents two main challenges; the generation of structures and the discrimination between structures with a scoring function. In this study, the focus is on the first of these challenges, for which a number of approaches have been used. The simplest method of docking two structures is to treat them as rigid bodies, usually using the Fast Fourier Transform (FFT) technique [1–8] which may include a refinement stage [9], although other rigid-body methods have been used [10,11]. For many proteins, however, there are flexible deformations upon binding which can alter their geometric and electrostatic properties. Attempts to cater for flexibility include rigid-body cross-docking of an ensemble of structures generated with molecule dynamics (MD) [12–14] and flexible refinement of rigidly docked poses [6,15–22]. Soft potentials have also been used to allow minor clashes [10,17,22], whilst others use a mean-field technique to select a conformations from a multiple copy representation [6,23] or assemble independently docked subunits connected by hinge regions [24]. A common flexible docking approach consists of an energy minimisation protocol (such as Monte Carlo, simulated annealing, simplex, MD, adopted basis Newton-Raphson or steepest decent) and a set of parameters to minimise (such as multiple copy weightings, atomic cartesian coordinates, internal coordinates or rotamers). In this paper, the degree to which low frequency normal modes can capture the unbound to bound transition is quantified, both on their own and in linear combination. A novel docking algorithm, SwarmDock, is explained and tested on a number of cases of varying difficulty.

Linear, harmonic vibrational motions around a single minima can be calculated using normal mode analysis. These motions may then be ordered by vibrational frequency. The lowest frequency eigenvectors correspond to the motions that can be excited with the least amount of energy and hence are the most accessible with thermal energy. The elastic network model simplifies the analysis, by only taking account of the shape of the molecule under analysis, an approximation which has shown to be capable of reproducing thermal B-factors at an atomistic [25] and coarse-grained [26] level of resolution, as well as NMR [27] and MD atomic fluctuations [28].

Despite the approximation only being valid for small linear motions and ignoring multiple optima and solvent dampening, many protein motions resemble a single low energy normal mode [29–32]. Dobbins *et al.* investigated normal mode analysis in the context of protein-protein docking. They showed the use of normal modes for predicting protein flexibility and demonstrated that global motions of high-flexibility proteins can, in many cases, be approximated with a single low energy normal mode [33]. While these studies have investigated the overlap between single normal modes and molecular motions, of interest in the context of constructing a docking algorithm is how closely a linear combination of normal modes can capture the conformational change upon binding, especially that of interface residues. Cui *et al.* used normal modes as a basis for understanding motion in the F₁ ATPase molecular motor, using RMSD as a guide [34]. Petrone and Pande used a similar approach when analysing the motions of four proteins, including one protein-protein binding event [35], whilst Tama *et al.* have used elastic network normal modes to fit atomic structures to low resolution cryo-electron microscopy maps [36]. Mustard and Ritchie used a method related to normal mode analysis to study motions that occur upon the binding of 10 test complexes [37]. To date, there has been no investigation of the ability of normal modes to recreate binding associated conformational change at the binding interface, the subset of residues from which binding free energy is calculated.

Due to the reduced number of degrees of freedom, a linear combination of low frequency eigenvectors has been discussed as a method of modelling flexibility in docking algorithms [38–40]. Normal modes have been used to generate ensembles for cross-docking [41], relax docked structures [42,43] and dock small molecules [44–48]. Whilst these methods use normal modes in docking, the only algorithms to date that adjust the coefficients of a linear combination of normal modes during the descent of the protein-protein binding funnel, are those that use the coarse grain model, ATTRACT [49] and FiberDock [50]. In the ATTRACT algorithm, up to 5 coefficients represent the amplitude along low frequency normal modes and the eigenvectors represent conformational degrees of freedom. A similar approach is used in FiberDock, in which 10 modes are selected based upon the repulsive forces acting between the binding partners.

Here, we investigate the ability of fine-grained, all-atom elastic network normal modes to analytically map various components of the unbound structures onto their equivalents in the bound structures using least squares linear regression and a large data set of 236 conformational changes upon binding [51]. As many protein interface refinement protocols do not optimise the backbone and sidechains concertedly, the ability of normal modes to capture conformational change of the backbone alone is also investigated. A normal mode representation of the backbone could be used in combination with a rotamer based side-chain optimisation algorithm, such as SCWRL [52]. Similarly, algorithms such as PULCHRA [53] can be used to reconstruct detail from a coarse-grained C_α structural representation. Hence, we also perform the mapping at this level of resolution. Furthermore, we investigate the rotation-translation-of-blocks (RTB) method of calculating elastic network normal modes, and its ability to capture the unbound to bound transition.

During docking, concerted change in conformation, position and orientation are desirable. SwarmDock, a memetic docking algorithm in which the translational, orientational and conformational degrees of freedom are simultaneously optimised using the Particle Swarm Optimisation (PSO) metaheuristic [54], combined with a Solis and Wets local search algorithm [55], was developed to perform flexible docking using normal modes. The PSO algorithm is effective as a robust optimiser of non-linear, multi-modal functions. It has been successfully applied to many optimisation problems including neural network training, design optimisation, data mining [56], gene clustering [57] and sequence alignment [58]. More recently, PSO has been used for small molecule docking and has been shown to outperform methods based on simulated annealing and genetic algorithms [59–61].

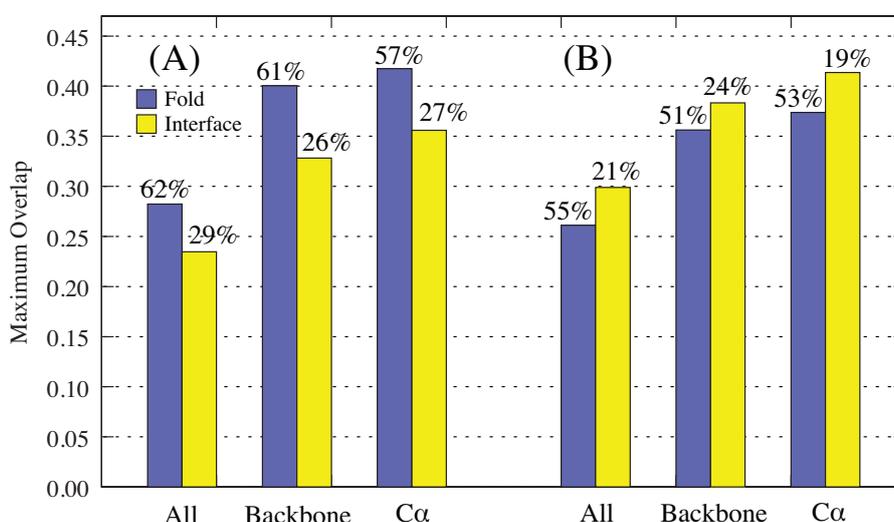
In the SwarmDock protocol, the PSO algorithm is used to optimise coefficients of a linear combination of Hessian eigenvectors, position and orientation. Local and global, flexible, unbound-unbound docking is performed on a number of complexes. Whilst SwarmDock has only been implemented with a simplistic energy function, it can successfully dock flexible structures which undergo significant conformational changes upon binding. The success of the algorithm as a function of the number of soft modes included for flexible deformation of the binding partners, is reported. A brief overview of SwarmDock has been given elsewhere [62], however the details of the algorithm are elaborated here for the first time.

2. Results and Discussion

2.1. Overlap

The intermolecular potential is determined by interacting residues, defined here as residues which have a non-hydrogen atom within 6 Å of a non-hydrogen atom on the binding partner. It is the ability of normal coordinates to capture the transition of these atoms that is most important for the purpose of docking. Figure 1A shows the mean maximum overlap (maximum overlap averaged over the test set) over the whole fold and at the interface, at three levels of resolution. As the size of the subsystem for which the overlap is calculated decreases, the mean maximum overlap increases. As expected, the difference between the backbone and the C_α mean maximum overlaps is small, as backbone atoms move collectively with their adjacent atoms. The drop in mean maximum overlap when we compare backbone to all atoms demonstrates that whilst the side chains do move in concert with their adjacent backbone atoms, they can also move independently and change their rotameric state.

Figure 1. Mean Maximum overlap across the whole fold and the interface, for all atoms, backbone atoms and C_α atoms, calculated with (A) standard diagonalisation of the Hessian and (B) the RTB approach. The percentage of complexes for which the mode of maximum overlap is one of the first 5 non-trivial modes is shown above the bars.



However, when we compare the mean maximum overlap of the whole protein to that of the interface, we observe a decrease at all levels of resolution. Furthermore, while the global motion is best represented by one of the 5 lowest frequency modes (57–62% of the modes of maximum overlap), conformational change at the binding interface is not (26–29% of the modes of maximum overlap). A Kolmogorov-Smirnov test shows that there is no evidence against uniformity in the distribution of modes of maximum overlap for the interface C_α atoms ($D = 0.070$, $D_{0.01} = 0.107$). The same test for the global C_α conformation shows that the mode of maximum overlap is not drawn from a uniform distribution above the 99% significance level ($D = 0.368$, $D_{0.01} = 0.107$).

This analysis was extended to the the modes of maximum overlap for the lowest 500 normal modes, and a difference was seen between the motion of interface C_α atoms and all interface atoms

(see Figure 2). However, when compared to the distribution of modes of maximum overlap for global motion, it is clear that much higher frequency modes are involved in the subtle changes that occur at the binding interface. This demonstrates that whilst the global change in protein conformation usually resembles one of the five lowest frequency normal modes, the conformational change of the binding interface does not. This is in agreement with Petrone and Pande [35], who suggest that higher energy modes can be activated by the presence of the binding partner. This discrepancy between the interface and global motion raises the question of the origin of the global change observed when comparing crystal structures. A dramatic example is given by the complex between influenza hemagglutinin and its paratope (complex 2VIS), as shown in Figure 3. The lowest frequency normal mode in the antigen binding protein corresponds to hinge bending motion which overlaps greatly with the unbound to bound transition. Whilst these types of conformational change can be involved in function, in this case the ligand binds away from the interdomain region and this motion does not correspond to the conformational change at the binding interface.

Figure 2. Maximum overlap and respective mode for (A) interface residues and (B) the whole fold.

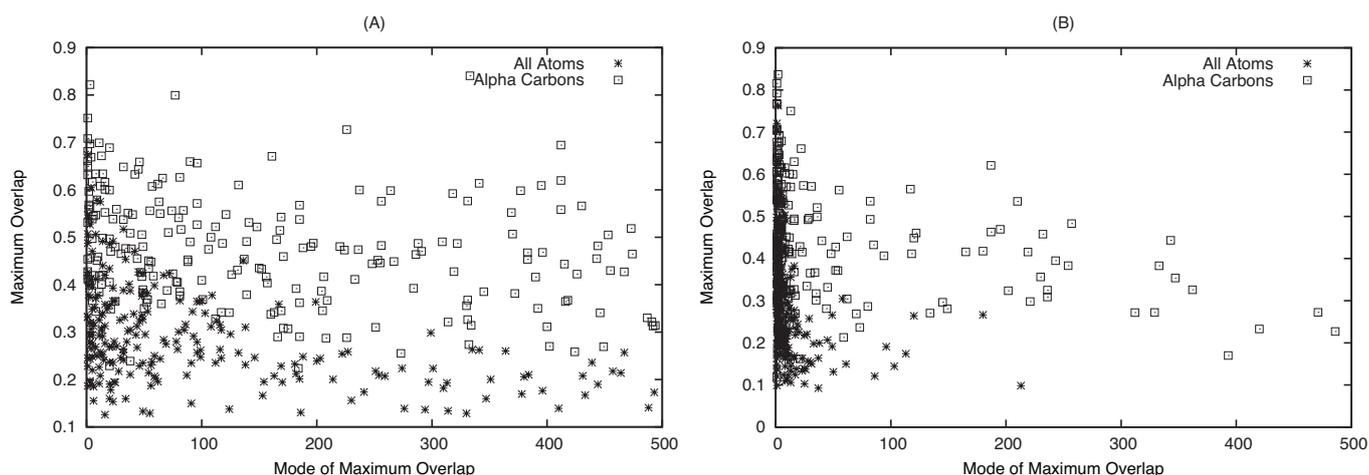
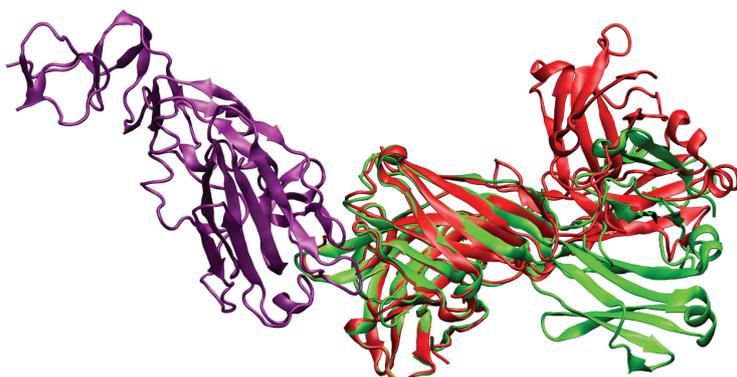


Figure 3. Complex 2VIS, of influenza hemagglutinin (purple) and murine IgG1, λ HC19 antibody in the bound (green) and unbound (red) crystal structures, superimposed on the binding domain.



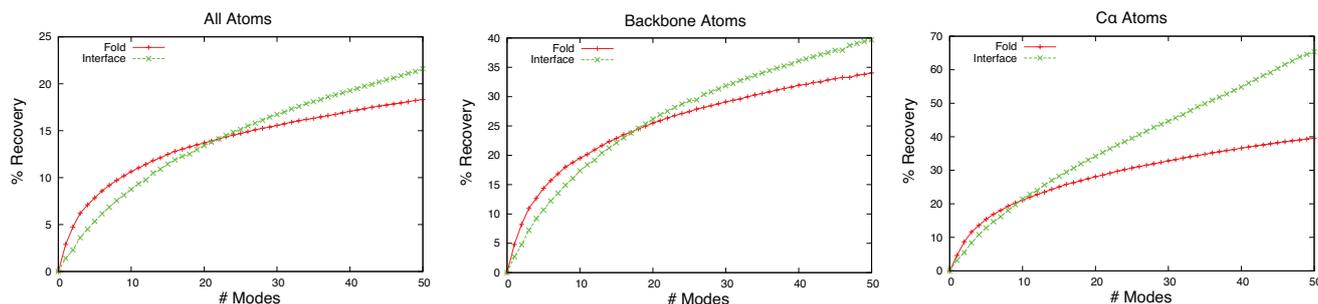
It has been shown that different crystal structures of the same protein can vary in their hinge bending angle, such as in the cubic and rhombic space group structures of the ligand free ribose binding protein [63]. Additionally, MD simulations of crystal structures of the T4 lysozyme, varying in their degree of hinge bending, have shown convergence of incipient structures after 500 ps [64]. Evidently, care must be taken when comparing crystal structures so as not to confuse binding associated change with crystal-packing induced conformational rearrangement facilitated by the intrinsic flexibility of the protein in question. Due to localisation of the binding interface, conformational change upon docking is expected to be less susceptible to this effect when considering the interface only, as this change is either not associated with global change, or the global movements are functional aspects of the binding, as in the case of a ligand binding to a cleft between two domains.

2.1.1. Rotations-Translation-in-Blocks Method

The above analysis was repeated for modes calculated using RTB diagonalisation, and a comparison with modes derived from exact diagonalisation of the all-atom Hessian is shown in Figure 1B. The trend as we go to coarser levels of resolution in the fitting (all-atom to backbone to C_α) is the same. However, there is a noticeable difference in the methods ability to model the conformational change at the interface compared to across the whole fold. The RTB method is better at capturing the local changes at the interface, with mean overlap values of 0.41 and 0.30 for C_α and all-atom respectively. This exceeds the ability of the more computationally expensive standard Hessian diagonalisation method at capturing the interface unbound-bound change (0.36 and 0.23) and is on a par with its well known ability to capture global motions (0.42 and 0.28). The reasons for this are not clear, but may be related to the perturbation of low frequency modes with higher modes calculated at the residue level. Nevertheless, this does suggest that the RTB method is not only faster, but also a better choice for pre-calculating normal modes for use in docking.

2.2. Linear Combinations of Normal Modes

While overlap gives a good measure of how close a single normal mode is to the unbound-bound transition, it is not possible to know which mode has the greatest overlap without already knowing the bound structure. Hence, the inclusion of a number of modes is preferable for use in a docking algorithm. Furthermore, using these modes in linear combination may be able to significantly enhance the ability to recapture the important conformational changes that occur upon binding beyond that achievable by considering single modes, while still significantly reducing the search space compared to Cartesian or internal co-ordinates. In order to investigate the contribution of multiple modes, the unbound-bound transitions were decomposed into linear combinations of normal modes. It should be noted that, in terms of representing dynamics with a linear combination of normal modes, phase angles and amplitudes of motion must be considered. However, the purpose of flexible docking is not to model protein dynamics, but to model the conformational change that occurs upon binding. Hence, least square fitting is an appropriate method of combining modes in this context, and it is also probable that the coefficients generated by least-square fitting correspond to a structure accessible by a combination of harmonic motions.

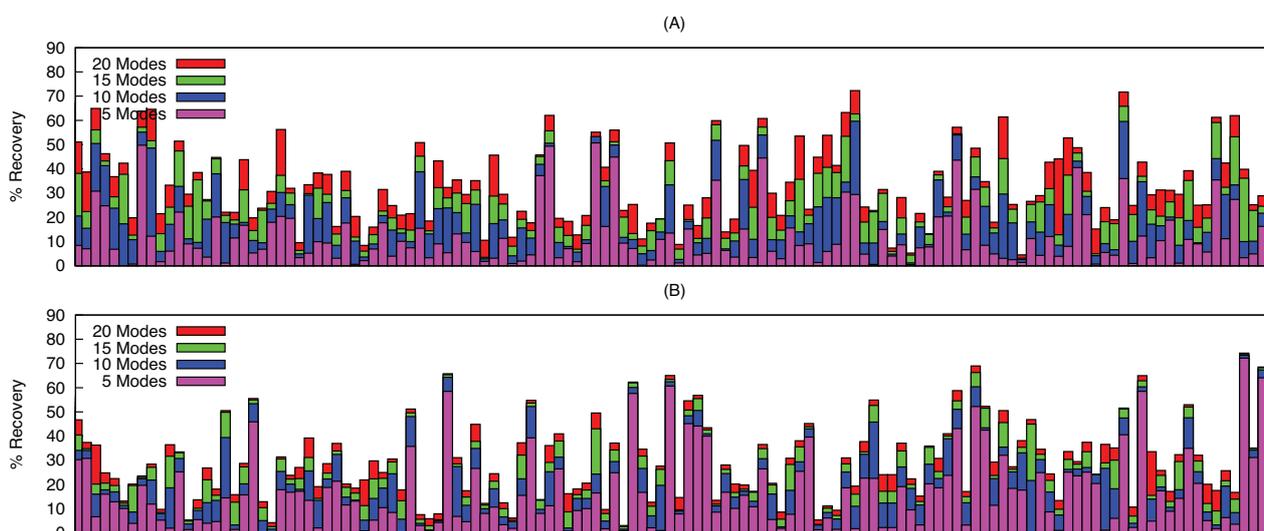
Figure 4. Mean percentage reduction in RMSD as a function of the number of modes used.

Linear least squares regression of unbound structures onto bound structures using a set of low frequency normal modes as a basis were performed. This yields the coefficients for the structure, of the infinite structures which can be generated with those normal modes as a basis, that minimises the RMSD against the bound form. Just as a certain subset of atoms, such as C_{α} atom or interface atoms, can be used to determine the local overlap between a normal mode and the docking transition, subsets of atoms can be used in the fitting. Figure 4 shows the mean percentage reduction in RMSD, from the initial RMSD between bound and unbound, plotted against the number of low frequency modes used. Using 5 normal modes, only 44% of complexes have greater recovery in the interface than across the whole fold. This figure increases to 55%, 60% and 64% when using 10, 15 and 20 modes respectively. A similar trend can be seen when analysing the results for all atoms (41%, 45%, 50% and 52%). It is evident that as the number of modes used increases, the ability to model the conformational change at the interface improves at a greater rate than it does across the whole fold. This shows that although low-frequency modes have greater overlap with global motions than with local rearrangements at the interface, when taken in combination, conformational rearrangement at the interface can be better modelled than those across the fold, as more modes are included. This demonstrates that whilst a single mode of large overlap may well be of much higher frequency than those usually used to describe protein motion, greater recovery of RMSD can be garnered without the need to use hundreds of modes.

Generally, significant improvements in RMSD are observed in a subset of the modes considered, with other modes offering only a smaller improvement. For each protein, the mode which gives the greatest decrease in RMSD is the mode of greatest influence. Using twenty modes, the mean mode of greatest improvement for the whole fold is 6.8 and 5.0 when fitting the C_{α} and all atom models respectively. These increase to 9.3 and 8.1 when only the interface is considered, in line with the overlap data shown in Figure 2 and further demonstrating the involvement of higher frequency motions occurring at the interface during binding.

Figure 5 shows the improvement in RMSD for C_{α} interface and fold for a number of higher-flexibility proteins. A number of conclusions can be drawn from this graph. When using 20 modes, the improvement in RMSD is generally greater for the interface than across the whole fold. Furthermore, the conformational rearrangements across the fold are usually modelled by the lowest 5 modes (purple), compared to the higher frequency modes that are more often involved in conformational rearrangements at the interface (red and green).

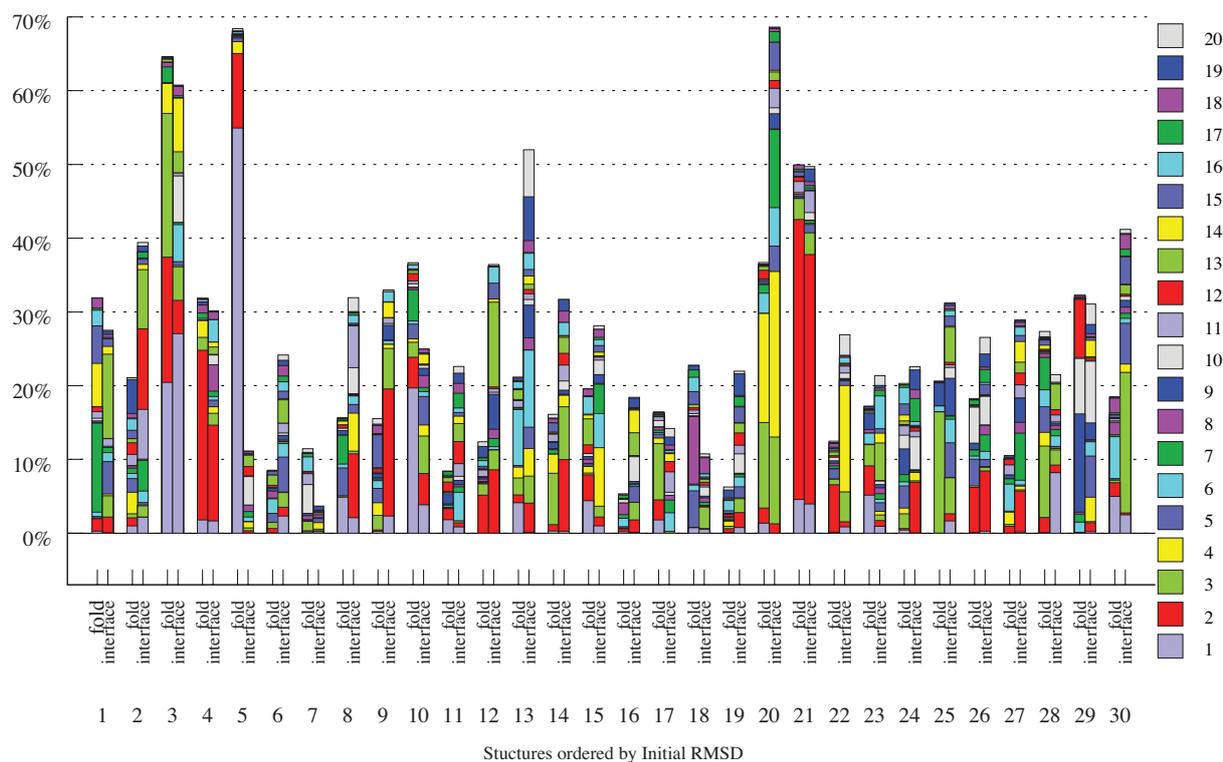
Figure 5. Percent reduction in RMSD upon analytical fitting of C_{α} atoms to the unbound-bound transition, using 5, 10, 15 and 20 normal modes respectively as a basis. Values for the 130 proteins of highest initial C_{α} RMSD, for both the interface (A) and the whole fold (B), are shown. The structures are ordered left to right in ascending order of initial RMSD.



A more detailed breakdown of mode contributions for 30 complexes of high flexibility, as calculated using 20 modes, is shown in Figure 6. One of the most salient feature of this graph is structure 5, the influenza hemagglutinin and its paratope (structure 2VIS), discussed earlier. The global hinge-bending motion associated with the first mode is clearly unrelated to the conformational change occurring at the interface and most likely the influence of crystal packing forces. The distribution of modes for the interface is distinctly different from that of the fold for most cases, with some notable exceptions: 1I1BR (Ran/Importin β complex), 1E4K (immunoglobulin G Fc fragment/Fc γ RIII complex) and 1Y64 (actin/BNI1 complex), which all correspond to larger columns in Figure 6. For 1I1BR, the receptor wraps around the ligand, which spans the entirety of the receptor. For 1Y64, the ligand binds to a cleft between two domains, which opens up with hinge bending motion. In 1E4K, the ligand binds to two distinct domains which again opens up with hinge bending motion. In these three cases, the global motion is functionally involved in the binding process.

It is interesting to note, however, that despite the differences between the global and local change, the predicted flexibility measure found by Dobbins *et al.* [33], which they showed to be effective at predicting global flexibility, can also be used to predict local flexibility. A Wilcoxon rank-sum test shows that, when ordered by C_{α} , backbone and all-atom initial RMSD, the predicted flexibilities are not drawn from a different distribution, with significance of 0.0197, 0.0197 and 0.0006, respectively.

Figure 6. Percentage recovery of initial RMSD for 30 structures of greatest initial fold RMSD, ordered from left to right in descending order of initial RMSD, using all atoms in the fine-grained elastic network model. Effect of inclusion of 1 to 20 modes is shown for both the interface and the whole fold. Structures included are (1) 1IRA_r; (2) 1H1V_l; (3) 1Y64_r; (4) 1FAK_r; (5) 2VIS_r; (6) 1R8S_r; (7) 1IBR_r; (8) 1EER_r; (9) 1FC2_r; (10) 2FD6_l; (11) 2COL_l; (12) 1FQ1_l; (13) 1JMO_r; (14) 1BKD_l; (15) 1GPW_r; (16) 1I2M_r; (17) 1YVB_l; (18) 2AJF_l; (19) 1NW9_r; (20) 1E4K_r; (21) 1IBR_l; (22) 2CFH_l; (23) 1HIA_l; (24) 2OT3_l; (25) 1KKL_r; (26) 1PXV_r; (27) 1KTZ_r; (28) 1BKD_r; (29) 1EER_l and (30) 1IB1_r.

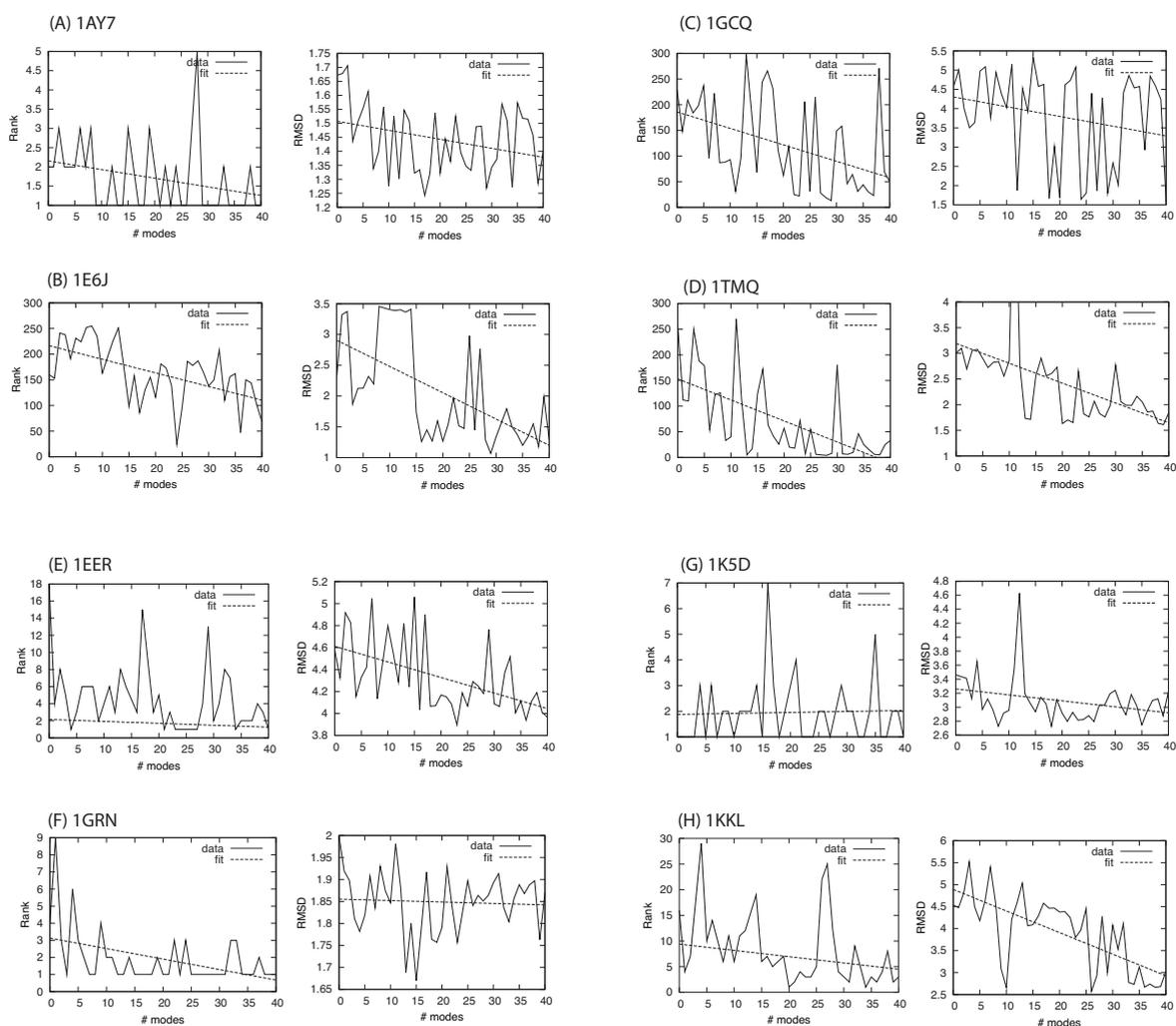


2.3. Docking as a Function of Normal Modes

In solution, the unbound proteins and the complex adopt an ensemble of conformational states. Much debate has focussed on the degree of overlap between the conformational ensembles of the binding partners unbound and in complex. At one end of the spectrum lies the ‘conformational sorting and population shift’ mechanism, in which members of the bound state ensemble are accessible by the unbound protein. At the other end lies the ‘induced fit’ hypothesis, in which bound conformations are essentially inaccessible by the unbound protein, but stabilised by the presence of the binding partner [65]. Significant evidence supporting the conformational sorting model come from a number of sources (see Boehr *et al.* and references therein). This gives credence to the idea of modelling flexible binding using information based upon the conformational diversity of the unbound binding partners, such as NMR ensembles, clustered MD snapshots, normal mode analysis or principle component analysis of MD trajectories. Such an approach is taken here. While induced structural changes undoubtedly play a role in molecular recognition, for many cases it is expected that the conformational diversity available

using a set of pre-calculated low-frequency normal modes should be able to generate structures within or close to the bound ensemble, and with sufficient interaction energy to distinguish it from generated structures which do not correspond to the genuine biological interaction. Further, as higher frequency modes, which are visited less frequently in solution, are included, conformational change along these modes can be induced by the presence of the binding partner since, during the docking, both partners adjust their conformation in the presence of the field generated by the other.

Figure 7. Docking results and linear regression for global docking (A–D) and local docking (E–H). I_RMSD is the lowest value found in the docking run.



Global docking of four small complexes was done using up to 40 normal modes in both the receptor and the ligand, as calculated using the RTB method. For each complex, the SwarmDock algorithm was run 480 times at points spaced evenly around the receptor (4 times at each of 120 points surrounding the receptor). The rank and I_RMSD are reported in Figure 7 (A–D). For the docking simulations involving a larger number of normal modes, structural distortions became apparent and so each structure was minimised in CHARMM prior to clustering to remove clashes and unphysical geometry in the internal structure. The general trend is shown by fitting a line to the data, all of which have a negative gradient for both the rank and the I_RMSD, indicating that as more modes are included in the simulation, there

is an improvement in the ability to discriminate correctly bound poses from false positives and capture conformational change at the interface. Interestingly, these complexes only undergo small backbone changes when comparing crystal structures. It may be the case that the inclusion of conformational dimensions preferentially broadens the correct binding funnel, enhancing the focus of the swarm, or deepening the energy minimum, as will be discussed shortly.

Local docking was performed for four high-flexibility models which undergo significant conformational change upon binding. In these docking runs, the search is focussed on the binding site region by performing more runs from the starting positions near the binding site and removing the starting positions away from the binding site. Of 120 positions evenly spaced around the receptor, the algorithm was run 8 times from the 10 positions nearest the binding site. For these complexes, 1EER, 1GRN, 1K5D and 1KKL, interface C_{α} RMSD after superimposition of the interface, are 2.44 Å, 1.22 Å, 1.19 Å and 2.20 Å respectively. Rank and LRMSD are shown in Figure 7 (E–H). Low ranking conformations could be found for all these complexes, and the RMSD of the two most flexible complexes improved as more modes were included.

For many, but not all, of the false positives, the cluster size is small and they are not found consistently. However, the correct binding site is usually found multiple time during the docking runs. Consideration of the mechanism of the algorithm and the nature of biological interfaces offers a potential explanation. Structures near to the native complex also have low energies and, as a rule of thumb, poses further away from the binding site tend to have higher energies, resulting in an energy funnel surrounding the native complex. This is essential for the evolution of the interface, as it overcomes the binding equivalent of Levinthal's paradox. SwarmDock is very different from other docking methods which rely upon finding low-energy structures by combining the results of many single independent trajectories through search space, or by filtering a list of putative structures using FFT correlations, geometric hashing or other scoring functions. Instead, communication between members of the swarm can result in, as an emergent property of the system, switching between exploration of diffuse regions of search space, or exploitation of narrower regions containing numerous positions which correspond to lower energy structures, depending on the nature of the energy landscape. SwarmDock takes advantage of a correlated energy landscape surrounding the true binding site, as low energy positions found by members of the swarm act as an attractor for a subset of the swarm. Further, the equation governing the velocities of the members of the swarm has a distance dependent repulsion term which acts against the contraction of a diffuse swarm, but has less effect on the contraction of a swarm which is focussing in on a particular region containing numerous low energy structures, such as the true biological interface. Indeed, previous work has shown that the mean Euclidean distance between the translational components of the particles (corresponding to the centres of mass of the ligands), decreases significantly earlier for SwarmDock runs which find the binding site compared to those that do not (data not published), indicating that fewer iterations are required to form a consensus in the population, and the presence of a wider binding funnel than for false positive sites - this information may be used in the future as part of a post-docking re-scoring step. It can now be seen that the inconsistent discovery of low energy false poses may not be because the number of starting positions is insufficient to find all low-energy poses, but because these poses are not surrounded by a wide binding funnel and hence not scrutinised by the swarm. Further information can be gleaned from the cluster size of the true positive cluster. Table 1 shows the average

number of times the correct binding site was found during docking. For all but one of the complexes, the correct binding site was found most frequently in the runs with 21–30 modes or 31–40 modes included, despite the increasing dimensionality of search space. This is indicative that the width of the true binding funnel is increasing as higher modes are added, and is hence being found by SwarmDock more frequently during the search. This is consistent with the notion that the interface has evolved for efficient recognition of proteins, all of which possess greater or lesser degrees of intrinsic flexibility.

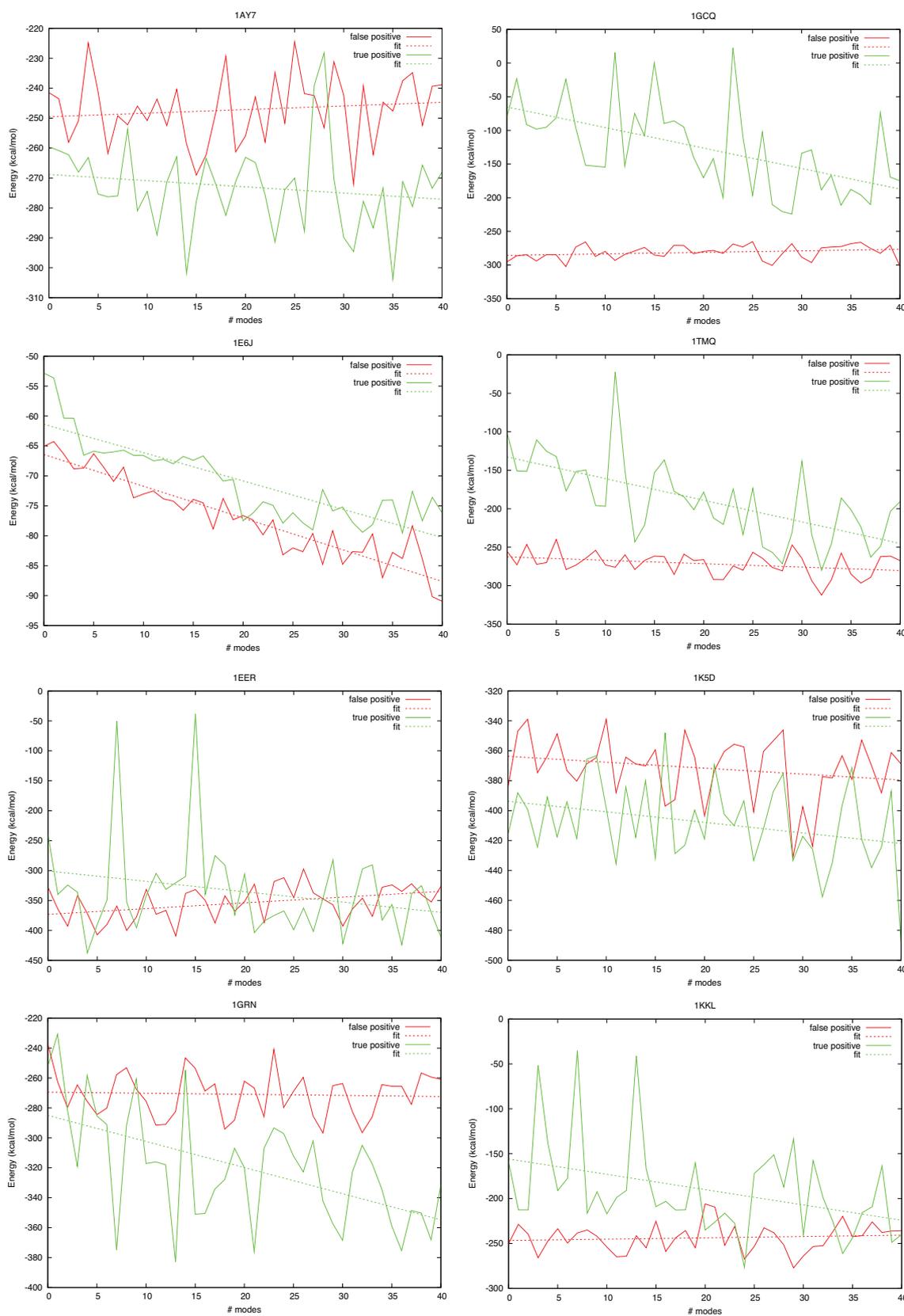
Figure 8 shows how the energy of true and false positive clusters vary as more modes are included. It is interesting to note that, with the exception of 1E6J, the energy of the false positives did not decrease significantly as greater flexibility was added. However, the gradient of the fitted line for the true positive solutions was negative in all cases, and resulted in a significant lowering of interaction energy for most cases. This indicates that the true binding funnels are preferentially deepened compared to false positive binding funnels as higher modes are included.

Table 1. The mean cluster size for the correct binding site ($I_{\text{RMSD}} < 5 \text{ \AA}$), averaged over runs including different numbers of modes. The greatest values are marked in bold.

Complex	Modes:	1–10	11–20	21–30	31–40
1AY7		15.1	11.2	12.3	14.3
1GCQ		1.2	1.9	2.9	2.6
1E6J		10.0	9.9	11.0	11.5
1TMQ		2.1	2.9	4.7	2.9
1EER		3.6	3.1	5.3	5.9
1K5D		5.9	6.9	7.4	6.6
1GRN		26.6	28.7	29.3	27.3
1KKL		2.8	2.9	3.9	5.6

As well as the examples given here, the SwarmDock algorithm has also been used in rounds 16–19 of the CAPRI blind docking trials. In the CAPRI experiment, the protein docking community test their algorithms blind. Target predictions are collected, the experimental bound structure revealed, and the predictions are ranked based on their quality. SwarmDock was able to successfully produce the bound structure of three target complexes: targets 37, 40 and 41. For these targets, semi-local docking was performed by eliminating certain regions around the receptor protein as potential binding sites, as described previously [62]. In target 37, acceptable predictions were made for Arf6 in complex with the LZ2 leucine zipper of JIP4 (pdb 2W83 [66]) by docking the unbound Arf6 (pdb 2A5D [67]) to a homology model of the leucine zipper built with the POPULUS server [68] using the GCN4 zipper as a template (pdb 2ZTA [69]). For target 40, both trypsin binding sites of the double-headed arrowhead inhibitor complex (pdb 3E8L) were predicted, one with medium accuracy and the other with high accuracy, by docking unbound bovine trypsin (pdb 1BTY [70]) to the bound form of the inhibitor. For target 41, a medium accuracy prediction was made for the Colicin E9/Im2 complex (pdb 1BXI), by docking the unbound Colicin E9 DNase domain (pdb 1FSJ) to the unbound Im2 immunity protein (pdb 2NO8).

Figure 8. Energy of lowest energy true positive structure and mean energy of the lowest energy member of the 5 lowest energy false positive clusters.



3. Conclusions

Conformational change upon complexation has been studied with the construction of docking algorithms in mind. The capability of fine-grained elastic network models to capture unbound to bound transitions was quantified by assessing the ability of normal modes to move all atoms, backbone (C, O, N, C_α) and C_α atoms from the unbound to bound conformation on a large test set. This analysis confirms the observation that the differences in the crystal structures of bound and unbound proteins can usually be approximated with a single low-frequency normal mode. However, both *in silico* and *in vivo*, protein-protein binding is driven by the lowering of the Gibbs free energy, which is dominated by the exclusion of water at the interface and the enthalpy of binding. Both these factors are largely determined by change that occurs at the protein-protein interface. Hence, the computational reconstruction of this change is of great importance when designing a docking algorithm. As previous studies have only looked at global motion, we paid close attention to the residues involved in binding. Global protein motion was shown to be distinct from localised conformational rearrangement at the interface, except in a minority of cases where global motion is functionally involved in docking. Despite this however, flexibility prediction designed to predict the extent of global change, can still be used to predict the degree of conformational rearrangement at the interface. Conformational changes at the interface correspond to motions associated with higher frequency modes than those associated with global structural rearrangements. While hundreds of modes need to be considered to find a single mode which sufficiently captures the unbound-bound transition at the interface, a linear combination of low-frequency normal modes is shown to do so with much fewer. Furthermore, despite involving higher frequency modes, greater improvement in RMSD can be made at the interface than across the whole fold, as increasing numbers of normal modes are used.

The effect of the RTB method was also investigated. While this technique is shown to be less accurate for predicting conformational change across the whole fold, it has a greater ability to model conformational change at the interface. For this reason, and due to its computational efficiency, this method makes a better choice for generating pre-calculated modes for docking. Fine-grained RTB normal modes were calculated and used in a novel docking algorithm, SwarmDock. SwarmDock is the first protein-protein docking algorithm based on particle swarm optimization, and was developed for flexible docking using normal modes as a method of modeling conformational change. The algorithm works well on the test cases presented here and in the blind CAPRI experiment. Further, RMSD and rank of docked structures can be demonstrably improved by the inclusion of normal modes, even for difficult cases and there is evidence to suggest that the true binding funnel is both widened and deepened, relative to false positive binding funnels, as higher modes are included. This demonstrates that the algorithm can effectively exploit the conformational freedom given by a set of low frequency modes. Further work, focussing on an improved scoring function and post-docking refinement and filtering, should further the ability for the SwarmDock protocol to generate potential bound structures and discriminate biological interfaces from artifacts.

4. Experimental Section

4.1. Normal Mode Analysis

In normal mode analysis, the potential energy basin is approximated as a harmonic well, allowing analytical solution to the equations of motion. Using a typical molecular mechanics energy function requires lengthy minimisation to ensure that the minima is found. To avoid this difficulty, we use the elastic network model (ENM) [25,26,32], with a Hookian potential used between close nodes, and the initial structure accepted as the minima.

$$U_{AB} = \frac{1}{2}k(R_{AB} - R_{e,AB})^2 \quad U_{tot} = \sum_{R_{AB} < C} U_{AB}$$

Where the force constant, k , is equal for all atom pairs, R_{AB} is the Euclidean distance between nodes A and B and $R_{e,AB}$ is their equilibrium separation. A cutoff, C , of 10 Å was used in all ENM calculations, where all atoms are nodes. The all-atom elastic network model is robust to cutoff [71], and the value used is large enough to dampen excessive motions of peripheral atoms, the tip effect. The Hessian matrix contains the second derivatives of the potential with respect to mass-weighted coordinates. Construction and diagonalisation of the Hessian matrix yields solutions in which all nodes have the same angular frequency, given by the square root of the eigenvalues. The respective motions are given by the normal coordinates, the eigenvectors. Any structural change can be constructed as a linear combination of normal coordinates, the complete set of which form an orthogonal basis. More detailed accounts are given elsewhere [72,73]. We use the package Elnemo to construct the Hessian matrix [74] and diagonalise it using the Basic Linear Algebra Subprograms (BLAS) library [75]. All calculations were done on unbound structures, including all small molecule ligands, excluding crystallographic waters. Small molecules were removed from the results for the remainder of the analysis and for docking.

4.1.1. The Rotation-Translation-of-Blocks Method

In the building block (RTB) technique for diagonalising the Hessian matrix, atoms are grouped into blocks, usually of one residue or more, which can translate and rotate as a rigid unit. The all-atom Hessian is projected into a block translation and rotation subspace by the use of a projection matrix. The projected Hessian is then diagonalised, yielding the vibrational frequencies and eigenvectors [76,77]. The eigenvectors are then translated back into all-atom space by applying the transpose of the projection matrix, to produce approximate low-frequency modes. Subsequently, these approximate all-atom modes are perturbed toward the exact solution iteratively by accounting for higher frequency modes calculated for each block [78,79]. This method can reproduce frequencies and atomic fluctuations with accuracy comparable to that achieved when diagonalising the all-atom Hessian. Single residue blocks are used in this paper, as calculated using the `diagrtb` program distributed with Elnemo [74].

4.1.2. Overlap

The overlap value, O_j describes how well the j^{th} normal mode captures the displacement of atoms involved in a protein motion [80].

$$O_j = \frac{|\sum_{i=1}^{3n} a_{ij}(r_i^b - r_i^u)|}{\sqrt{\sum_{i=1}^{3n} a_{ij}^2 \sum_{i=1}^{3n} (r_i^b - r_i^u)^2}}$$

Where r_i^b and r_i^u are the i^{th} coordinates of the bound and unbound structures respectively and a_{ij} is the i^{th} coordinate of the j^{th} normal mode. Here, the maximum overlap, O_{max} , is defined as the greatest overlap between an unbound to bound transition of the first 20 modes, unless stated otherwise.

4.1.3. Data Set

The 124 complexes in the Protein-Protein docking benchmark v3.0 were used [51]. Both binding partners were analysed which, after removing the 12 complexes for which the unbound structure is only known for one of the binding partners, gives 236 conformation changes upon binding. All structures have a resolution below 3.25 Å with redundancy alleviated as described by Chen *et al.* [81].

4.2. Analytical Mapping of Unbound to Bound

This section describes the method used to determine the extent of deformation along each normal mode in order to generate the closest structure to the bound state from the unbound state. When fitting unbound structures to the bound structures using normal coordinates as a basis, a one-to-one correspondence of atoms is necessary. For some complexes, the bound and unbound structures have this correspondence. For the remainder, sequences of the two structures were aligned and non-matching residues were ignored in the mapping, but not in the construction of the Hessian matrix. When considering subsets of the structure, such as interface or backbone only, all atoms not in that subset are also ignored, leaving the n atoms of interest. When investigating the ability of an orthogonal set of vectors to capture protein motion, the contribution of each mode can be calculated independently as a projection [37,43]. However, when investigating atoms which do not correspond to the complete set from which the modes were calculated, such as the interface, orthogonality can no longer be assumed and an alternative approach is necessary. Hence, the unbound structures were mapped onto the bound structures using linear least squares regression, *ie* the unbound-bound transition is decomposed through basis expansion of a limited number of modes. The transition, \mathbf{T} , from unbound structure, \mathbf{D} , to bound structure \mathbf{E} , is fitted to a linear combination of normal modes, $f(i, \beta) = \sum_{j=1}^m \beta_j \mathbf{M}_{ij}$, by finding the minimum sum of square residuals, S .

$$\mathbf{T} = \mathbf{E} - \mathbf{D} = \mathbf{r} + \sum_{j=1}^m \beta_j \mathbf{M}_j \quad S = \sum_{i=1}^{3n} r_i^2$$

As the sum squared residuals are quadratic with respect to the coefficients, its derivative with respect to the coefficients is linear, with one unique solution which, as the square residuals are convex, correspond to a minimum.

Setting the gradient to zero,

$$\frac{\partial S}{\partial \beta_j} = 2 \sum_{i=1}^{3n} r_i \frac{\partial r_i}{\partial \beta_j} = 0 \quad \text{for} \quad j = 1 \dots m$$

which can be arranged to a series of j normal equations,

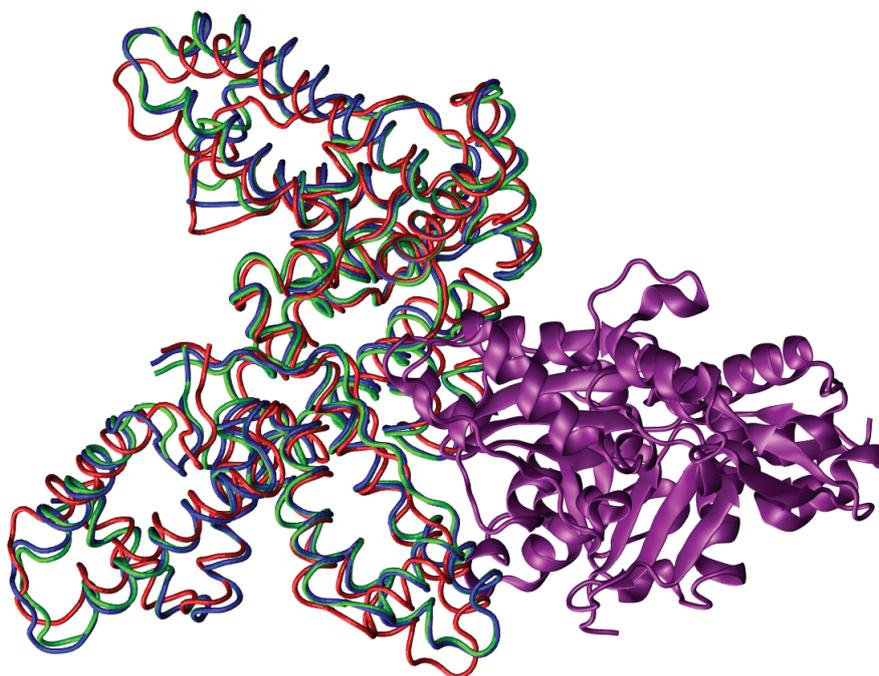
$$\sum_{i=1}^{3n} \sum_{k=1}^m M_{ij} M_{ik} \beta_k = \sum_{i=1}^{3n} M_{ij} T_i \quad \text{for } j = 1 \dots m$$

Written in matrix form and inversion shows how the analytically fit coefficients, β , are obtained. These coefficients correspond to the magnitudes of deformation along their respective normal coordinates.

$$(\mathbf{M}^T \mathbf{M}) \beta = \mathbf{M}^T \mathbf{T} \quad \beta = (\mathbf{M}^T \mathbf{M})^{-1} \mathbf{M}^T \mathbf{T}$$

Fitting was done for all 136 unbound-bound transitions in the data set, an example from which is shown in Figure 9.

Figure 9. The complex, 1KXP, between actin (purple) and vitamin D binding protein (VDBP) in the bound (green) and unbound (red) conformations. A fitted structure (blue), obtained by linear regression to the VDBP unbound to bound transition, using 20 normal modes as a basis, has a C_α RMSD of 0.87 Å compared to an initial RMSD of 2.12 Å. This corresponds to the closest structure to the bound that is attainable using the 20 lowest non-trivial normal modes.

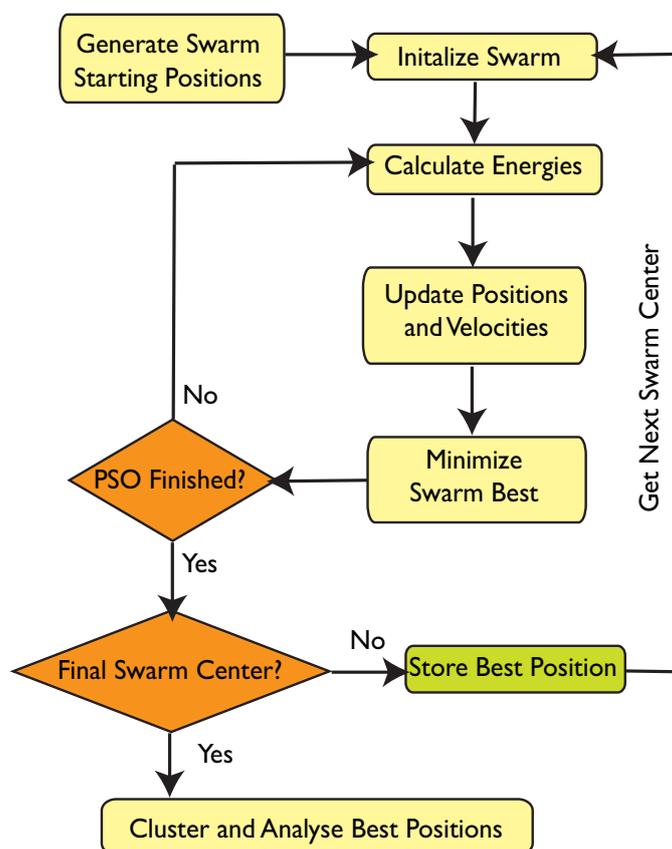


4.3. SwarmDock

The SwarmDock algorithm is a novel iterative population based mimetic algorithm. Each member of the population, i , is a particle in search space with both position in search space, χ_i , and velocity, \mathbf{v}_i . Upon each iteration, the energy of each particle is evaluated and the positions and velocities are updated.

Additionally, the lowest energy particle undergoes a local search based on that of Solis and Wets [55]. A typical SwarmDock run takes around 10 minutes on a 2.66 GHz CPU. A schematic of the algorithm is given in Figure 10.

Figure 10. An overview of the SwarmDock algorithm.



4.3.1. Search Space

The search space is composed of the ligand centre of mass (3 translational dimensions) and a quaternion representation of ligand orientation (4 orientational dimensions). The receptor is kept fixed at the origin. N_r and N_l normal mode coefficients ($N_r + N_l$ conformational dimensions) are also included for the receptor and ligand respectively. Hence, χ_i is a vector pointing to a position in $7 + N_r + N_l$ dimensional search space.

4.3.2. Initialisation

SwarmDock is run at approximately evenly spaced positions surrounding the receptor, generated using a point distribution algorithm explained elsewhere [62]. The translational component of each particles position is moved randomly away from this position (drawn from a Gaussian $\sigma = 10 \text{ \AA}$) and each particle is assigned a random orientation. Normal mode coefficients for the initial swarm are drawn from a Gaussian. To ensure reasonable internal bonded geometry of the initial swarm, normal

mode coefficients are drawn from a low distribution ($\sigma = 3.0$) and the algorithm searches from here. Throughout this paper, the same number of modes are used in the receptor and the ligand.

4.3.3. Propagation

The energy function is evaluated for all members of the initial population, then the velocities and positions are updated using the following transition functions, a variation on the well studied PSO equations [54]. These are used iteratively to propagate the algorithm.

$$\chi_i(t+1) = \chi_i(t) + \mathbf{v}_i(t+1)$$

$$\mathbf{v}_i(t+1) = w\mathbf{v}_i(t) + c_1r_{1,i}(\mathbf{p}_i(t) - \chi_i(t)) + c_2r_{2,i}(\mathbf{p}_{n,i}(t) - \chi_i(t)) + r_{3,i}(\mathbf{p}_{\text{rand}}(t) - \chi_i(t))$$

where w , the inertial weight, is the degree to which previous velocity contributes to the new velocity. $r_{1,i}$, $r_{2,i}$ and $r_{3,i}$ are random numbers taken from a uniform distribution between 0 and 1. Parameters c_1 and c_2 are respectively known as the cognitive and social aspect, both set to 2.05. The cognitive aspect is the propensity for the particle to move toward the lowest energy position it has previously experienced. The social aspect is the degree to which the particle moves toward the lowest energy position found by any particle in the neighbourhood of the particle being updated. \mathbf{p}_i and $\mathbf{p}_{n,i}$ are position vectors that correspond to the lowest energy position found by particle i and the best position found by any particle within the neighbourhood of particle i , respectively. \mathbf{p}_{rand} is the position of a randomly selected particle.

Note that the use of the word neighbourhood has a specific, non-colloquial meaning, as it appears in the PSO literature. A particle is deemed to be in the neighbourhood of particle i if it is one of the k particles adjacent to particle i in the vector array of the computer memory, where k is the neighbourhood size, which is set to 114 for all docking runs, out of the total population of 350. Neighbourhoods wrap around in the standard way, so that particles connected to their neighbours form a ring network topology.

4.3.4. Velocity Clamping

To avoid explosions and control the exploration/exploitation trade-off, velocity clamping is used to impose a limit on the distance the particles can move in any one iteration:

$$v_{ij}(t+1) = \begin{cases} v'_{ij} & \text{if } -V_{j,max} < v'_{ij} < V_{j,max} \\ V_{j,max} & \text{if } v'_{ij} > V_{j,max} \\ -V_{j,max} & \text{if } v'_{ij} < -V_{j,max} \end{cases}$$

where v'_{ij} is the calculated velocity of iteration $t+1$. V_{max} is set at 5 Å for translational parameters, 0.2 rad for the angular term in the quaternion and 0.5 Å for spatial terms in the quaternion. As the conformational parameters do not reach excessive velocities, no clamping is set on the normal mode coefficients.

4.3.5. Local Minimisation

At each iteration, the lowest energy particle in the swarm undergoes a local search step which is based on that of Solis and Wets [55]. This is an adaptive minimisation technique which does not require

gradient information. Its adaptivity arises from its ability to adjust its step size according to the number of previous successes or failures. Additionally, there is a bias vector in search space, \mathbf{b} which gives preferential searching in the direction that has previously been most successful.

The algorithm works as follows. For each dimension in search space, j , a deviate is taken from a gaussian distribution centered around the bias b_j , with a standard deviation ρ_j . These deviates form a vector \mathbf{d} . The energy of the particle is evaluated at $\chi + \mathbf{d}$. If this energy is lower than the energy at position χ then the bias vector is adjusted according to the equation below, the success counter is incremented and the fail counter is set to zero. If not, the energy is evaluated at $\chi - \mathbf{d}$ and if lower in energy than at χ then again, the bias is updated, the success counter is incremented and fail counter is set to zero. If a lower energy position is not found then the success counter is set to zero and the fail counter is incremented.

$$\mathbf{b}(t+1) = \begin{cases} 0.6\mathbf{b}(t) + 0.4\mathbf{d} & \text{if } E(\chi + \mathbf{d}) < E(\chi) \\ 0.6\mathbf{b}(t) - 0.4\mathbf{d} & \text{if } E(\chi + \mathbf{d}) > E(\chi) \text{ and } E(\chi - \mathbf{d}) < E(\chi) \\ \mathbf{b}(t)/2 & \text{if } E(\chi + \mathbf{d}) > E(\chi) \text{ and } E(\chi - \mathbf{d}) > E(\chi) \end{cases}$$

If the success or fail counters reach five, then the the step size is expanded or contracted, by doubling or halving the standard deviations, ρ . Initial values for ρ_j are 0.5 Å for the translational parameters, 5° for the angular term in the quaternion, 0.25 Å for the spatial terms in the quaternion and 0.15 for the normal mode coefficients. The algorithm is repeated and terminates after 5 consecutive contractions.

4.3.6. Clustering

After a set number of iterations, the lowest energy complex found is returned. The results for all runs are clustered. The lowest energy structure forms the first cluster. In ascending order of energy, all subsequent structures are added to a cluster if they are within 2.5 Å RMSD of the first structure in any existing cluster, otherwise they form a new cluster. A ranked list of structures is returned, with each structure corresponding to the lowest energy member of the cluster.

4.3.7. Analysis of Docked poses

The analysis of the quality of docked poses uses interface RMSD, a metric that is used in the community-wide CAPRI experiment [82]. Two residues are deemed to be in contact if they have a pair of atoms within 5 Å of each other. Interface residues are those in contact with the binding partner. The interface RMSD, I_RMSD, is defined as the minimum RMSD between the backbone atoms of the predicted structure and the crystal structure, just considering the interface residues.

4.3.8. Energy Function

A simple energy function is employed in the docking algorithm, composed of a van der Waals term and a Coulombic term between i and j interacting atoms, on the receptor and ligand respectively. A switching function is used between 7 and 9 Å ($r_{on} = 7$ and $r_{off} = 9$), in order to eliminate discontinuity

in the energy function and avoid calculating long-distance interactions with negligible contribution to the interaction energy.

$$E_{int} = \sum_i^{atoms} \sum_j^{atoms} E_{i,j}$$

$$E_{i,j} = \begin{cases} \frac{q_i q_j}{\epsilon r_{i,j}} + \sqrt{\epsilon_i \epsilon_j} \left[\left(\frac{R_{min_{i,j}}}{r_{i,j}} \right)^{12} - \left(\frac{R_{min_{i,j}}}{r_{i,j}} \right)^6 \right] & \text{if } r_{i,j} < 7 \\ \left(\frac{(r_{off} - r_{i,j})^2 (r_{off} + 2r_{i,j} - 3r_{on})}{(r_{off} - r_{on})^3} \right) \left[\frac{q_i q_j}{\epsilon r_{i,j}} + \sqrt{\epsilon_i \epsilon_j} \left[\left(\frac{R_{min_{i,j}}}{r_{i,j}} \right)^{12} - \left(\frac{R_{min_{i,j}}}{r_{i,j}} \right)^6 \right] \right] & \text{if } 7 < r_{i,j} < 9 \\ 0 & \text{if } 9 < r_{i,j} \end{cases}$$

All parameters are taken from the Charmm19 force field [83]. No potential was used to model the internal energy. While structures very rarely wandered into regions of search space which correspond to significantly distorted structures, this did result in some structural deformations when higher modes were included in the calculation. However, the structures were extensively minimised (500 steps conjugate gradient, 4000 steps adopted basis Newton-Raphson and 600 steps steepest decent) in CHARMM prior to final interaction energy calculation, clustering and analysis, ameliorating perturbations in structural integrity. In the future, changes in internal geometry may be penalised during the simulation in accordance to the extent of deformation along each mode and the eigenvalue association with that mode, so that a more diverse initial swarm can be generated.

Acknowledgements

The authors would like to thank Marcin Krol, Ozge Kurkcuoglu and the members of the biomolecular modelling laboratory for their helpful comments. This work was funded by Cancer Research UK.

References

1. Katchalski-Katzir, E.; Shariv, I.; Eisenstein, M.; Friesem, A.A.; Aflalo, C.; Vakser, I.A. Molecular surface recognition: Determination of geometric fit between proteins and their ligands by correlation techniques. *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 2195–2199.
2. Chen, R.; Weng, Z. Docking unbound proteins using shape complementarity, desolvation, and electrostatics. *Proteins* **2002**, *47*, 281–294.
3. Schneidman-Duhovny, D.; Inbar, Y.; Nussinov, R.; Wolfson, H.J. PatchDock and SymmDock: servers for rigid and symmetric docking. *Nucleic Acids Res.* **2005**, *33*, W363–367.
4. Shentu, Z.; Al Hasan, M.; Bystroff, C.; Zaki, M.J. Context shapes: Efficient complementary shape matching for protein-protein docking. *Proteins* **2008**, *70*, 1056–1073.
5. Li, N.; Sun, Z.; Jiang, F. SOFTDOCK application to protein-protein interaction benchmark and CAPRI. *Proteins* **2007**, *69*, 801–808.
6. Jackson, R.M.; Gabb, H.A.; Sternberg, M.J. Rapid refinement of protein interfaces incorporating solvation: application to the docking problem. *J. Mol. Biol.* **1998**, *276*, 265–285.
7. Mandell, J.G.; Roberts, V.A.; Pique, M.E.; Kotlovyy, V.; Mitchell, J.C.; Nelson, E.; Tsigelny, I.; Ten Eyck, L.F. Protein docking using continuum electrostatics and geometric fit. *Protein Eng.* **2001**, *14*, 105–113.

8. Ritchie, D.W.; Kemp, G.J. Protein docking using spherical polar Fourier correlations. *Proteins* **2000**, *39*, 178–194.
9. Tovchigrechko, A.; Vakser, I.A. Development and testing of an automated approach to protein docking. *Proteins* **2005**, *60*, 296–301.
10. Zacharias, M. Protein-protein docking with a reduced protein model accounting for side-chain flexibility. *Protein Sci.* **2003**, *12*, 1271–1282.
11. Gardiner, E.J.; Willett, P.; Artymiuk, P.J. Protein docking using a genetic algorithm. *Proteins* **2001**, *44*, 44–56.
12. Smith, G.R.; Sternberg, M.J.; Bates, P.A. The relationship between the flexibility of proteins and their conformational states on forming protein-protein complexes with an application to protein-protein docking. *J. Mol. Biol.* **2005**, *347*, 1077–1101.
13. Grunberg, R.; Leckner, J.; Nilges, M. Complementarity of structure ensembles in protein-protein binding. *Structure* **2004**, *12*, 2125–2136.
14. Krol, M.; Chaleil, R.A.; Tournier, A.L.; Bates, P.A. Implicit flexibility in protein docking: Cross-docking and local refinement. *Proteins* **2007**, *69*, 750–757.
15. Krol, M.; Tournier, A.L.; Bates, P.A. Flexible relaxation of rigid-body docking solutions. *Proteins* **2007**, *68*, 159–169.
16. Dominguez, C.; Boelens, R.; Bonvin, A.M. HADDOCK: A protein-protein docking approach based on biochemical or biophysical information. *J. Am. Chem. Soc.* **2003**, *125*, 1731–1737.
17. Gray, J.J.; Moughon, S.; Wang, C.; Schueler-Furman, O.; Kuhlman, B.; Rohl, C.A.; Baker, D. Protein-protein docking with simultaneous optimization of rigid-body displacement and side-chain conformations. *J. Mol. Biol.* **2003**, *331*, 281–299.
18. Comeau, S.R.; Gatchell, D.W.; Vajda, S.; Camacho, C.J. ClusPro: An automated docking and discrimination method for the prediction of protein complexes. *Bioinformatics* **2004**, *20*, 45–50.
19. Camacho, C.J.; Gatchell, D.W. Successful discrimination of protein interactions. *Proteins* **2003**, *52*, 92–97.
20. Li, L.; Chen, R.; Weng, Z. RDOCK: Refinement of rigid-body protein docking predictions. *Proteins* **2003**, *53*, 693–707.
21. Andrusier, N.; Nussinov, R.; Wolfson, H.J. FireDock: Fast interaction refinement in molecular docking. *Proteins* **2007**, *69*, 139–159.
22. Fernandez-Recio, J.; Totrov, M.; Abagyan, R. ICM-DISCO docking by global energy optimization with fully flexible side-chains. *Proteins* **2003**, *52*, 113–117.
23. Bastard, K.; Prvost, C.; Zacharias, M. Accounting for loop flexibility during protein-protein docking. *Proteins* **2006**, *62*, 956–969.
24. Schneidman-Duhovny, D.; Inbar, Y.; Nussinov, R.; Wolfson, H.J. Geometry-based flexible and symmetric protein docking. *Proteins* **2005**, *60*, 224–231.
25. Tirion, M.M. Large Amplitude Elastic Motions in Proteins from a Single-Parameter, Atomic Analysis. *Phys. Rev. Lett.* **1996**, *77*, 1905–1908.
26. Bahar, I.; Atilgan, A.R.; Erman, B. Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. *Fold Des.* **1997**, *2*, 173–181.

27. Yang, L.W.; Eyal, E.; Chennubhotla, C.; Jee, J.; Gronenborn, A.M.; Bahar, I. Insights into equilibrium dynamics of proteins from comparison of NMR and X-ray data with computational predictions. *Structure* **2007**, *15*, 741–749.
28. Rueda, M.; Chacon, P.; Orozco, M. Thorough validation of protein normal mode analysis: A comparative study with essential dynamics. *Structure* **2007**, *15*, 565–575.
29. Yang, L.; Song, G.; Jernigan, R.L. How well can we understand large-scale protein motions using normal modes of elastic network models? *Biophys. J.* **2007**, *93*, 920–929.
30. Krebs, W.G.; Alexandrov, V.; Wilson, C.A.; Echols, N.; Yu, H.; Gerstein, M. Normal mode analysis of macromolecular motions in a database framework: Developing mode concentration as a useful classifying statistic. *Proteins* **2002**, *48*, 682–695.
31. Tama, F.; Sanejouand, Y.H. Conformational change of proteins arising from normal mode calculations. *Protein Eng.* **2001**, *14*, 1–6.
32. Atilgan, A.R.; Durell, S.R.; Jernigan, R.L.; Demirel, M.C.; Keskin, O.; Bahar, I. Anisotropy of fluctuation dynamics of proteins with an elastic network model. *Biophys. J.* **2001**, *80*, 505–515.
33. Dobbins, S.E.; Lesk, V.I.; Sternberg, M.J. Insights into protein flexibility: The relationship between normal modes and conformational change upon protein-protein docking. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 10390–10395.
34. Cui, Q.; Li, G.; Ma, J.; Karplus, M. A normal mode analysis of structural plasticity in the biomolecular motor F(1)-ATPase. *J. Mol. Biol.* **2004**, *340*, 345–372.
35. Petrone, P.; Pande, V.S. Can conformational change be described by only a few normal modes? *Biophys. J.* **2006**, *90*, 1583–1593.
36. Tama, F.; Miyashita, O.; Brooks, C.L. Normal mode based flexible fitting of high-resolution structure into low-resolution experimental data from cryo-EM. *J. Struct. Biol.* **2004**, *147*, 315–326.
37. Mustard, D.; Ritchie, D.W. Docking essential dynamics eigenstructures. *Proteins* **2005**, *60*, 269–274.
38. Andrusier, N.; Mashiach, E.; Nussinov, R.; Wolfson, H.J. Principles of flexible protein-protein docking. *Proteins* **2008**, *73*, 271–289.
39. Bonvin, A.M. Flexible protein-protein docking. *Curr. Opin. Struct. Biol.* **2006**, *16*, 194–200.
40. May, A.; Zacharias, M. Accounting for global protein deformability during protein-protein and protein-ligand docking. *Biochim. Biophys. Acta* **2005**, *1754*, 225–231.
41. Rueda, M.; Bottegoni, G.; Abagyan, R. Consistent improvement of cross-docking results using binding site ensembles generated with elastic network normal modes. *J. Chem. Inf. Model.* **2009**, *49*, 716–725.
42. Zacharias, M.; Sklenar, H. Harmonic modes as variables to approximately account for receptor flexibility in ligand-receptor docking simulations: Application to DNA minor groove ligand complex. *J. Comp. Chem.* **1999**, *20*, 287–300.
43. Lindahl, E.; Delarue, M. Refinement of docked protein-ligand and protein-DNA structures using low frequency normal mode amplitude optimization. *Nucleic Acids Res.* **2005**, *33*, 4496–4506.
44. May, A.; Zacharias, M. Protein-ligand docking accounting for receptor side chain and global flexibility in normal modes: Evaluation on kinase inhibitor cross docking. *J. Med. Chem.* **2008**, *51*, 3499–3506.

45. Floquet, N.; Marechal, J.D.; Badet-Denisot, M.A.; Robert, C.H.; Dauchez, M.; Perahia, D. Normal mode analysis as a prerequisite for drug design: application to matrix metalloproteinases inhibitors. *FEBS Lett.* **2006**, *580*, 5130–5136.
46. Sander, T.; Liljefors, T.; Balle, T. Prediction of the receptor conformation for iGluR2 agonist binding: QM/MM docking to an extensive conformational ensemble generated using normal mode analysis. *J. Mol. Graph. Model.* **2008**, *26*, 1259–1268.
47. Cavasotto, C.N.; Kovacs, J.A.; Abagyan, R.A. Representing receptor flexibility in ligand docking through relevant normal modes. *J. Am. Chem. Soc.* **2005**, *127*, 9632–9640.
48. Kovacs, J.A.; Cavasotto, C.N.; Abagyan, R. Conformational Sampling of Protein Flexibility in Generalized Coordinates: Application to Ligand Docking. *J. Comput. Theor. Nanosci.* **2005**, *2*, 354–361.
49. May, A.; Zacharias, M. Energy minimization in low-frequency normal modes to efficiently allow for global flexibility during systematic protein-protein docking. *Proteins* **2008**, *70*, 794–809.
50. Mashach, E.; Nussinov, R.; Wolfson, H.J. FiberDock: Flexible induced-fit backbone refinement in molecular docking. *Proteins* **2010**, *78*, 1503–1519.
51. Hwang, H.; Pierce, B.; Mintseris, J.; Janin, J.; Weng, Z. Protein-protein docking benchmark version 3.0. *Proteins* **2008**, *73*, 705–709.
52. Canutescu, A.A.; Shelenkov, A.A.; Dunbrack, R.L. A graph-theory algorithm for rapid protein side-chain prediction. *Protein Sci.* **2003**, *12*, 2001–2014.
53. Rotkiewicz, P.; Skolnick, J. Fast procedure for reconstruction of full-atom protein models from reduced representations. *J. Comput. Chem.* **2008**, *29*, 1460–1465.
54. Kennedy, J.; Eberhart, R.C. Particle Swarm Optimization. In Proceedings of the IEEE International Conference on Neural Networks, Perth, Australia, 1995; Volume 4, pp. 1942–1948.
55. Solis, F.J.; Wets, R.J.B. Minimization by Random Search Techniques. *Math. Oper. Res.* **1981**, *6*, 19–30.
56. Sousa, T.; Silva, A.; Neves, A. Particle swarm based Data Mining Algorithms for classification tasks. *Parallel Comput.* **2004**, *30*, 767–783.
57. Xiao, X.; Dow, E.R.; Eberhart, R.; Miled, Z.B.; Oppelt, R.J. Gene Clustering Using Self-Organizing Maps and Particle Swarm Optimization. In Proceedings of the International Parallel and Distributed Processing Symposium, Nice, France, 2003; pp. 154b.
58. Rasmussen, T.K.; Krink, T. Improved Hidden Markov Model training for multiple sequence alignment by a particle swarm optimization-evolutionary algorithm hybrid. *BioSystems* **2003**, *72*, 5–17.
59. Namasivayam, V.; Gunther, R. PSO@Autodock: A fast flexible molecular docking program based on Swarm intelligence. *Chem. Biol. Drug Des.* **2007**, *70*, 475–484.
60. Chen, H.M.; Liu, B.F.; Huang, H.L.; Hwang, S.F.; Ho, S.Y. SODOCK: Swarm optimization for highly flexible protein-ligand docking. *J. Comput. Chem.* **2007**, *28*, 612–623.
61. Janson, S.; Merkle, S.; Middendorf, M. Molecular docking with multi-objective Particle Swarm Optimization. *Appl. Soft Comput.* **2008**, *8*, 666–675.

62. Li, X.; Moal, I.H.; Bates, P.A. Detection and Refinement of Encounter Complexes for Protein-Protein Docking: Taking Account of Macromolecular Crowding. *Proteins* **2010**, doi: 10.1002/prot.22770.
63. Magnusson, U.; Chaudhuri, B.N.; Ko, J.; Park, C.; Jones, T.A.; Mowbray, S.L. Hinge-bending motion of D-allose-binding protein from Escherichia coli: Three open conformations. *J. Biol. Chem.* **2002**, *277*, 14077–14084.
64. Arnold, G.E.; Ornstein, R.L. Protein hinge bending as seen in molecular dynamics simulations of native and M61 mutant T4 lysozymes. *Biopolymers* **1997**, *41*, 533–544.
65. Boehr, D.D.; Nussinov, R.; Wright, P.E. The role of dynamic conformational ensembles in biomolecular recognition. *Nat. Chem. Biol.* **2009**, *5*, 789–796.
66. Isabet, T.; Montagnac, G.; Regazzoni, K.; Raynal, B.; El Khadali, F.; England, P.; Franco, M.; Chavrier, P.; Houdusse, A.; Menetrey, J. The structural basis of Arf effector specificity: The crystal structure of ARF6 in a complex with JIP4. *EMBO J.* **2009**, *28*, 2835–2845.
67. O’Neal, C.J.; Jobling, M.G.; Holmes, R.K.; Hol, W.G. Structural basis for the activation of cholera toxin by human ARF6-GTP. *Science* **2005**, *309*, 1093–1096.
68. Offman, M.N.; Tournier, A.L.; Bates, P.A. Alternating evolutionary pressure in a genetic algorithm facilitates protein model selection. *BMC Struct. Biol.* **2008**, *8*, 34.
69. O’Shea, E.K.; Klemm, J.D.; Kim, P.S.; Alber, T. X-ray structure of the GCN4 leucine zipper, a two-stranded, parallel coiled coil. *Science* **1991**, *254*, 539–544.
70. Katz, B.A.; Finer-Moore, J.; Mortezaei, R.; Rich, D.H.; Stroud, R.M. Episelection: Novel Ki approximately nanomolar inhibitors of serine proteases selected by binding or chemistry on an enzyme surface. *Biochemistry* **1995**, *34*, 8264–8280.
71. Kurkcuoglu, O.; Jernigan, R.L.; Doruker, P. Loop motions of triosephosphate isomerase observed with elastic networks. *Biochemistry* **2006**, *45*, 1173–1182.
72. Brooks, B.R.; Janezic, D.; Karplus, M. Harmonic analysis of large systems. I. Methodology. *J. Comp. Chem.* **1995**, *16*, 1522–1542.
73. Tama, F.; Gadea, F.X.; Marques, O.; Sanejouand, Y.H. Building-block approach for determining low-frequency normal modes of macromolecules. *Proteins* **2000**, *41*, 1–7.
74. Suhre, K.; Sanejouand, Y.H. ElNemo: a normal mode web server for protein movement analysis and the generation of templates for molecular replacement. *Nucleic Acids Res.* **2004**, *32*, W610–W614.
75. Dongarra, J. Basic Linear Algebra Subprograms Technical Forum Standard. *Int. J. High Perform. Appl. Supercomput.* **2002**, *16*, 115–199.
76. Tama, F.; Gadea, F.X.; Marques, O.; Sanejouand, Y.H. Building-block approach for determining low-frequency normal modes of macromolecules. *Proteins* **2000**, *41*, 1–7.
77. Li, G.; Cui, Q. A coarse-grained normal mode approach for macromolecules: an efficient implementation and application to Ca²⁺-ATPase. *Biophys. J.* **2002**, *83*, 2457–2474.
78. Durand, P.; Trinquier, G.; Sanejouand, Y.H. A new approach for determining low-frequency normal modes in macromolecules. *Biopolymers* **1994**, *34*, 759–771.
79. Durand, P. Direct determination of effective Hamiltonians by wave-operator methods. I. General formalism. *Phys. Rev. A* **1983**, *28*, 3184–3192.

80. Marques, O.; Sanejouand, Y.H. Hinge-bending motion in citrate synthase arising from normal mode calculations. *Proteins* **1995**, *23*, 557–560.
81. Chen, R.; Mintseris, J.; Janin, J.; Weng, Z. A protein-protein docking benchmark. *Proteins* **2003**, *52*, 88–91.
82. Mendez, R.; Leplae, R.; de Maria, L.; Wodak, S.J. Assessment of blind predictions of protein-protein interactions: current status of docking methods. *Proteins* **2003**, *52*, 51–67.
83. MacKerell, A.D.; Bashford, D.; Bellott.; Dunbrack, R.L.; Evanseck, J.D.; Field, M.J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F.T.K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D.T.; Prodhom, B.; Reiher, W.E.; Roux, B.; Schlenkrich, M.; Smith, J.C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.

© 2010 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license ([http://creativecommons.org/licenses/by/3.0/.](http://creativecommons.org/licenses/by/3.0/))