JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

Efficient Relaxation of Protein−Protein Interfaces by Discrete Molecular Dynamics Simulations

Agusti Emperador,[†,§] Albert Solernou,[‡,§] Pedro Sfriso,[†,§] Carles Pons,[‡,§] Josep Lluis Gelpi,[‡,§,∥] Juan Fernandez-Recio,*[,‡,§] and Modesto Orozco*[,†,‡,§,∥]

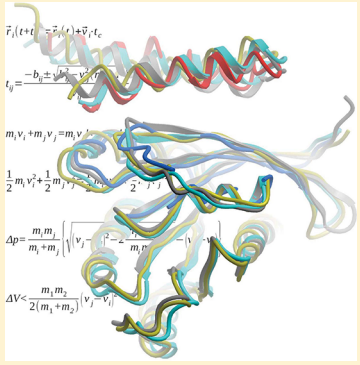
[†]Institute for Research in Biomedicine (IRB Barcelona), Baldiri i Reixac 10, Barcelona 08028, Spain

[‡]Barcelona Supercomputing Center, Jordi Girona 29, Barcelona 08034, Spain

[§]Joint BSC-IRB Research Program in Computational Biology, Barcelona, Spain

[∥]Departament de Bioquímica, Facultat de Biologia, Avgda Diagonal 645, Barcelona 08028, Spain

ABSTRACT: Protein−protein interactions are responsible for the transfer of information inside the cell and represent one of the most interesting research fields in structural biology. Unfortunately, after decades of intense research, experimental approaches still have difficulties in providing 3D structures for the hundreds of thousands of interactions formed between the different proteins in a living organism. The use of theoretical approaches like docking aims to complement experimental efforts to represent the structure of the protein interactome. However, we cannot ignore that current methods have limitations due to problems of sampling of the protein−protein conformational space and the lack of accuracy of available force fields. Cases that are especially difficult for prediction are those in which complex formation implies a non-negligible change in the conformation of the interacting proteins, i.e., those cases where protein flexibility plays a key role in protein−protein docking. In this work, we present a new approach to treat flexibility in docking by global structural relaxation based on ultrafast discrete molecular dynamics. On a standard benchmark of protein complexes, the method provides a general improvement over the results obtained by rigid docking. The method is especially efficient in cases with large conformational changes upon binding, in which structure relaxation with discrete molecular dynamics leads to a predictive success rate double that obtained with state-of-the-art rigid-body docking.



INTRODUCTION

Proteins are social molecules that exert their biological action through the interaction with other molecules. In particular, most signal transduction mechanisms in the cell are mediated by protein−protein interactions, which define complex interaction networks. Massive proteomics studies outline the components of different complexes in the cell, providing the first pictures of cellular interactomes. This provides knowledge of which molecular partners are participating in a given protein−protein interaction, but in order to understand its function at the molecular level or to interfere in the complex formation by small compounds, detailed information on the three-dimensional structure of the complexes is required.

Atomic resolution experimental techniques, especially X-ray crystallography and NMR, are providing an increasing amount of information on the three-dimensional structures of protein−protein complexes. However, the number of deposited complex structures in the Protein Data Bank[1] is currently less than 10% of the number of known binary interactions between human proteins[2] and a minuscule fraction of the estimated number of total interactions including transient complexes.[3] This has encouraged the use of computational methods, especially protein docking algorithms, which in the absence of atomic resolution structural data can provide useful information on protein complexes in the context of systems and network biology.

Many different protein−protein docking procedures have been reported to provide reduced sets of docking poses ideally enriched in near-native conformations and selected out of thousands or millions of alternative poses.[4,5] To achieve computational efficiency, and to reduce false positives, protein monomers are considered in most cases as rigid entities during the process. This rigid-body approach gives excellent results for those cases in which proteins show very little flexibility upon binding (e.g., average receptor and ligand unbound−bound backbone RMSD < 0.5 Å)[6] and therefore seem to associate following the "lock-and-key" mechanism.[7] However, for the majority of cases, complex formation involves larger conformational rearrangement,[8] and therefore the rigid docking approach yields poorer results.[6] Interestingly, the problematic cases for docking can be reasonably predicted based only on the intrinsic flexibility of the unbound state of the protein,[9] so treatment of flexibility is currently one of the major challenges in protein docking.

Different techniques to account for protein flexibility, typically at later stages of the docking procedure, have been developed.[10,11] In one of the first refinement methods, side-chain conformational optimization showed improvement in the docking predictions in specific cases, e.g., in those with a few

Received: July 23, 2012

   dx.doi.org/10.1021/ct301039e | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

Table 1. Docking Results before and after DMD Relaxation

PDB	interface RMSD[a]	pyDock rank[b]	DMD rank[b]	PDB	interface RMSD[a]	pyDock rank[b]	DMD rank[b]
1AY7	0.54	17	10	1R6Q	1.67	40	14
1AZS	0.72	12	81	1RV6	1.09	6	2
1B6C	1.96	1	1	1T6B	0.62	56	23
1BUH	0.75	65	51	1TMQ	0.86	1	44
1BVK	1.24	52	69	1UDI	0.9	1	4
1BVN	0.87	2	8	1XD3	1.24	2	5
1CLV	0.86	1	1	1XQS	1.77	13	50
1E6E	1.33	3	10	1XU1	1.3	18	75
1E6J	1.05	35	20	1Z0K	0.53	7	1
1E96	0.71	1	1	1Z5Y	1.23	79	12
1EWY	0.8	28	66	1ZHI	0.68	2	5
1F51	0.74	36	59	2ABZ	0.9	18	8
1FFW	1.43	74	25	2AYO	1.39	24	72
1FLE	1.02	3	24	2B42	0.72	1	1
1FSK	0.45	3	16	2BTF	0.75	33	39
1GLA	0.98	50	51	2FD6	1.07	17	2
1GPW	0.65	1	23	2G77	1.08	13	5
1H9D	1.32	26	51	2HLE	1.4	2	2
1IQD	0.48	8	1	2HQS	1.14	15	18
1J2J	0.63	19	11	2HRK	2.03	16	8
1JTG	0.49	1	2	2I25	1.21	40	76
1K74	0.8	14	6	2JEL	0.17	42	34
1KKL	2.2	81	3	2O8V	1.37	60	84
1M10	2.1	81	2	2PCC	0.39	49	83
1MAH	0.61	19	8	2SIC	0.36	6	1
1N8O	0.94	53	35	2SNI	0.35	3	1
1NW9	1.97	30	4	2VDB	0.47	5	6
1OPH	1.21	14	1	3SGQ	0.39	98	36
1OYV	0.47	84	95	4CPA	0.62	10	14
1PPE	0.44	6	2	7CEI	0.7	11	1
1R0R	0.45	3	64				

[a]Interface C_α RMSD between unbound and bound molecules (in Å). [b]Best rank of any near-native docking conformation.

problematic interface residues,[12,13] but its applicability was found to be more limited for difficult cases with large interfaces.[14] Good predictions were also obtained with backbone refinement when experimental restraints were used, as evaluated on a limited number of test cases.[15] Another approach based on the deformation of the proteins along the low-energy normal modes[16] was applied to a test set of 10 complexes with a significant difference between unbound and bound structures, giving good results for some of the complexes.[17] Backbone flexibility was also included via Monte Carlo refinement[18] based on the Rosetta method.[19] This procedure was applied to a choice of 25 out of 49 complexes of the first Weng benchmark[20] with successful results. The above methods, typically quite expensive from a computational point of view, have shown overall good results as part of a general docking strategy in the international CAPRI assessment of docking predictions (http://www.ebi.ac.uk/msd-srv/capri/), but more efforts need to be done to assess the real improvement of refinement with respect to rigid-body docking in large sets of complexes.

In this work, we present a new method to treat protein flexibility in docking computations using discrete molecular dynamics (DMD).[21−25] DMD is very efficient from a computational point of view and allows a very simple change of granularity, from atomistic to coarse grained levels and a complete control on the sampling space, from the side-chain only to the entire protein. The technique, which has been used with success to study protein aggregation[26,27] and conformational transitions,[28,29] assumes that particles move at constant velocity (ballistic regime) from collision to collision, avoiding then the time-consuming integration of Newton's equations of motion every femtosecond and reducing dramatically the cost of calculations, without a dramatic loss of accuracy with respect to standard atomistic simulation techniques.[30,31] The method developed here has been tested with very good success on Weng's protein−protein docking benchmark 4.0,[32] for which the structures of both the complex and the unbound partners are experimentally known.

OUTLINE OF THE METHOD

Protein Benchmark. We have used Weng's protein docking benchmark 4.0,[32] with known X-ray structures for both the unbound and the bound subunits. For test purposes, we selected only the 61 cases (Table 1) in which rigid-body docking with pyDock finds at least one near-native docking orientation (see below) among the 100 top-ranked conformations. Near-native docking solutions were defined (according to one of the criteria used in CAPRI) as those with interface RMSD < 4 Å RMSD with respect to the reference complex structure, being the interface RMSD calculated for the C_α atoms in such an interface. The success rates of the predictions are defined as the percentage of test cases in which at least one near-native docking solution was found within the top N solutions. We classified the benchmark cases according to the flexibility upon binding, for which we used the interface C_α RMSD between

dx.doi.org/10.1021/ct301039e | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

bound and unbound molecules, as defined in the protein benchmark 4.0.[32]

Rigid-Body Sampling and Scoring. We generated a set of 10 000 docking models for each protein−protein complex by using FTDOCK,[33] based on geometrical complementarity, using a grid resolution of 0.7 Å for the representation of proteins, and including an electrostatics term. These models were evaluated using pyDock,[34] a physics-based docking scoring function with excellent results in standard benchmarks as well as in CAPRI.[35,36] PyDock calculates the binding energy between the interacting proteins based on (i) a truncated and linearly screened electrostatic term, (ii) a truncated and weighted van der Waals term, and (iii) an accessible surface area (ASA)-based desolvation energy term with atomic parameters previously optimized for protein docking.[37] We selected for DMD relaxation the subset of the 100 best scored poses according to pyDock.

General Discrete Molecular Dynamics (DMD) Formalism. DMD is based on the use of discontinuous, stepwise potentials, which guarantee that particles move in the ballistic regime, following a linear movement with constant velocity until they reach a potential step:[22]

$$\vec{r}_i(t + t_c) = \vec{r}_i(t) + \vec{v}_i \cdot t_c \tag{1}$$

where and $\vec{r}_i$ and $\vec{v}_i$ stand for positions and velocities and t_c is the minimum among the collision times t_{ij} between each pair of particles i and j:

$$t_{ij} = \frac{-b_{ij} \pm \sqrt{b_{ij}^2 - v_{ij}^2(r_{ij}^2 - d^2)}}{v_{ij}^2} \tag{2}$$

where r_{ij} is the magnitude of $\vec{r}_{ij} = \vec{r}_j - \vec{r}_i$, v_{ij} is the magnitude of $\vec{v}_{ij} = \vec{v}_j - \vec{v}_i$, $b_{ij} = \vec{r}_{ij}\,\vec{v}_{ij}$, and d is the distance corresponding to the wall of the square well.

The collision happens when the particle distance is that corresponding to a potential step. Then there is a transfer of linear momentum into the direction of the vector $\vec{r}_{ij}$:

$$\begin{aligned} m_i\vec{v}_i &= m_i\vec{v}_i' + \Delta\vec{p} \\ m_j\vec{v}_j + \Delta\vec{p} &= m_j\vec{v}_j' \end{aligned} \tag{3}$$

where the prime indices denote the velocities after the collision.

The changes in velocities upon collision are derived by applying conservation of energy and momentum:

$$m_i v_i + m_j v_j = m_i v_i' + m_j v_j' \tag{4}$$

$$\frac{1}{2}m_i v_i^2 + \frac{1}{2}m_j v_j^2 = \frac{1}{2}m_i v_i'^2 + \frac{1}{2}m_j v_j'^2 + \Delta V \tag{5}$$

where ΔV stands for the size of the potential energy step.

The transferred momentum can be easily determined from

$$\Delta p = \frac{m_i m_j}{m_i + m_j}\left\{\sqrt{(v_j - v_i)^2 - 2\frac{m_i + m_j}{m_i m_j}\Delta V} - (v_j - v_i)\right\} \tag{6}$$

Note that, in case that $\Delta V > 0$, the two particles can overcome the step as long as

$$\Delta V < \frac{m_1 m_2}{2(m_1 + m_2)}(v_j - v_i)^2 \tag{7}$$

Otherwise, if the particles remain in the potential well, eq 6 reduces to

$$\Delta p = \frac{m_i m_j}{m_i + m_j}\{\sqrt{(v_j - v_i)^2} - (v_j - v_i)\} \tag{8}$$

which taking the negative solution of the root leads to

$$\Delta p = \frac{2m_i m_j}{m_i + m_j}(v_i - v_j) \tag{9}$$

To obtain good computational performance, we used here a simple implementation of DMD that uses a force-field containing "bonded" and "non-bonded" terms. The first ones include stretchings, bendings, and torsions (all represented by bond or pseudobond lengths) involving double or conjugated bonds, which are represented by means of infinite square wells with a width (derived from analysis of large database of atomistic MD simulations[22]) equal to 1% of the bond/pseudobond distance.

Regarding the nonbonded part, our force field includes solvation, van der Waals, and Coulomb electrostatic terms

$$V = V_{solv} + V_{VdW} + V_{Coul} \tag{10}$$

The van der Waals (V_{vdW}) and Coulomb (V_{Coul}) terms are the DMD version of the Lennard-Jones and Coulomb potentials, while solvation (V_{solv}) was accounted by the DMD version[22] of the Lazaridis and Karplus EEF1 effective energy function.[38] Intramolecular hydrogen bonds were restrained by square wells which guarantee the maintenance of secondary structure elements during the DMD simulations.

DMD Implementation for Protein−Protein Interactions. To increase computational efficiency, we used a multiscale representation of the proteins, where the level of detail and the allowed flexibility of the residues decrease as the distance to the protein interface increases. Thus, residues in the protein−protein interface (residues with at least one atom less than 8 Å from any atom of the other protein in the rigid docking pose) were defined at the all-heavy-atoms level, keeping them completely flexible. A second layer was defined by residues located at 8−12 Å from any atom of the other protein, where all-heavy-atoms representation was used, but atom positions were restrained by Go-like square wells. The rest of the protein was represented at the C_α level, using Go-like square wells to restrain their movements. DMD trajectories were long enough to ensure equilibration of the docking conformations (see Results and Discussion), and we used as the scoring function the average of the potential energy over the last 15% of the trajectory for each conformation.

Analysis of Native Contacts along Simulation. We considered that two residues defined a native contact if any of their atoms were interacting in the experimental complex structure, i.e., if one atom from the first residue and another atom from the other residue were within the interaction distance defined in the DMD potentials (that were stepwise functions with a finite range; see above).

Statistical Significance of DMD Improvement in Flexible Cases. We have applied a Wilcoxon rank-sum test to evaluate whether improvement of the results after DMD relaxation with respect to rigid-body docking is more evident for flexible cases. We built two groups of cases, one formed by 12 cases in which the DMD relaxation *significantly improved* with respect to rigid-body docking (best near-native rank went from >10 to <10) and the other one formed by six cases in which DMD relaxation *significantly worsened* with respect to rigid-body docking (best near-native rank went from <10 to >10).

dx.doi.org/10.1021/ct301039e | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

We considered that having or not a near native within the top 10 docking solutions is a reasonable criterion in order to evaluate the success of a docking result (as used in CAPRI). For higher ranks, a small improvement or worsening is irrelevant for performance assessment purposes. The Wilcoxon rank-sum test proved that the distributions of unbound−bound RMSD values in these two groups of cases differed significantly (Mann−Whitney $U = 13$, $n_1 = 12$, $n_2 = 6$, $p < 0.05$ two-tailed).

Estimating Binding Flexibility from Normal Modes. The extent of the conformational motion of a protein due to thermal fluctuations can be estimated within the normal-mode analysis (NMA) framework.[16] NMA assumes that the protein structure is a system of coupled harmonic oscillators connecting the C_α atoms. The dynamics of such a system is constituted by the normal modes, each one involving all the particles of the system. We estimated the total deformation of a protein as the average of the amplitudes due to each mode (eq 11):

$$\mathrm{RMSD}_{\mathrm{predicted}} = \sqrt{\frac{1}{N}\sum \lambda} \tag{11}$$

where the sum runs over all the normal modes $(3N - 6)$, and N is the number of C_α atoms. At a given temperature, the amplitude λ of the motion due to a normal mode is estimated as in eq 12:[39]

$$\lambda = \frac{k_B T}{k} \tag{12}$$

where k is the stiffness associated to the mode; therefore the relevant conformational changes will be produced by the softest normal modes (those with the lowest stiffness). We have assigned the predicted deformation of each complex as the average value of the predicted deformation of ligand and receptor. To classify the complexes as (allegedly) rigid and flexible, we have used a $\mathrm{RMSD}_{\mathrm{predicted}}$ cutoff value of 0.43 Å, which is the average of the predicted deformations over the complete benchmark. With this cutoff value, 60% of the complexes in the benchmark were classified as flexible. This coincides with the fraction that is considered flexible when taking the experimental interface deformations upon binding (using 1 Å as the interface RMSD cutoff value), but some complexes were classified in different groups when using each criterion.

RESULTS AND DISCUSSION

DMD Improves Docking Predictions. We performed rigid docking using our pyDock approach on the cases of Weng's protein docking benchmark 4.0,[32] for which both bound and unbound conformations of the interacting proteins are experimentally known. After a sequential scoring procedure (see Outline of the Method), we selected the 61 complexes for which the rigid docking procedure provided at least one near-native conformation within the 100 top scored docking poses (we limited the number of analyzed docking poses to 100 in order to perform a systematic benchmark in a reasonable time). This yielded a total of 6100 docking poses that were subjected to multiscale DMD simulations. It is worth noting that contrary to other protocols, such as atomistic MD simulation, which would demand simulation times orders of magnitude higher,[40] our DMD procedure allowed us to complete the process in less than 10 h on our 512 core Xeon Cluster (6100 trajectories for 61 complexes), making the method compatible with high-throughput procedures.

We found that DMD generally improves the docking energy landscapes, decreasing more effectively the energy of docking poses that are closer to the crystallographic complex structure. This is for example the case of 1Z0K, a complex that undergoes a small conformational change upon binding (0.53 Å interface RMSD between unbound and bound forms). Figure 1A shows

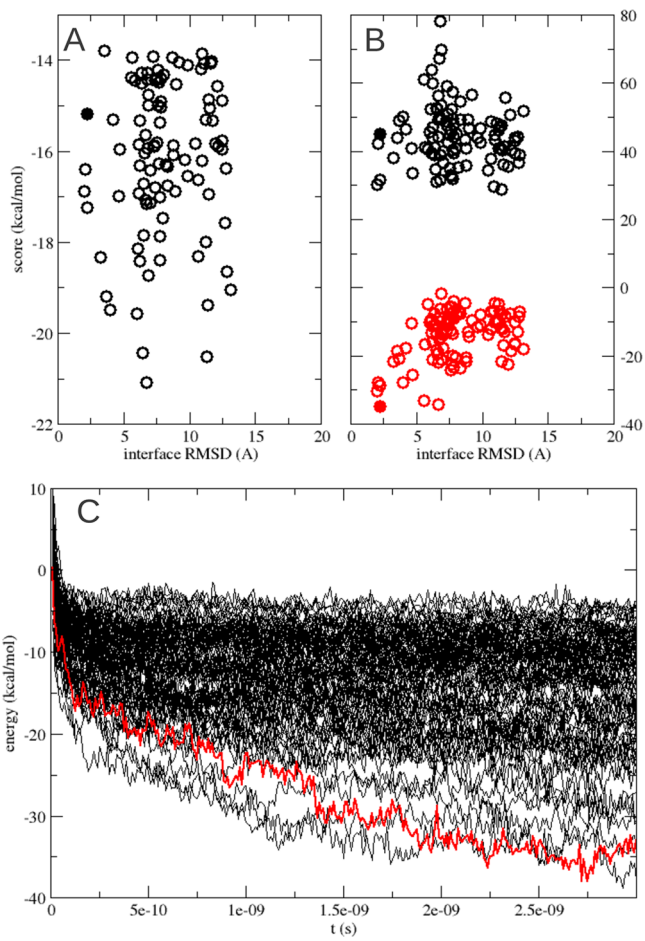


Figure 1. Rigid-body docking and DMD relaxation for complex 1Z0K. (A) pyDock score vs interface RMSD for each rigid-body docking conformation of the complex 1Z0K is shown. (B) DMD score vs interface RMSD for each conformation, before the simulations (rigid-body docking structures; in black) and after the DMD relaxation (in red). Filled circles represent the best-scoring near-native solution after DMD relaxation, shown also before DMD for comparison. (C) Evolution of the energies of the 100 docking conformations for the 1Z0K complex during the DMD simulations. The first ranked near-native conformation after relaxation has been highlighted in red.

the pyDock scoring of the rigid-body docking poses distributed according to interface RMSD with respect to the crystallographic complex structure. Figure 1B shows the distribution of these docking solutions according to DMD scoring before and after 3 ns DMD relaxation. Remarkably, the rank 1 docking solution after DMD relaxation (filled circle) is one of the best near-native rigid docking poses before relaxation in structural terms (2.2 Å interface RMSD from the experimental structure) but is not so good in terms of rigid-body energy (rank 67 and 57 by pyDock and DMD scoring before relaxation, respectively). The effect of DMD relaxation in docking refinement is clear in Figure 1C, which shows the evolution of the DMD potential energies of each docking pose for complex 1Z0K during the

dx.doi.org/10.1021/ct301039e | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

simulation, with the red line corresponding to the rank 1 solution after DMD relaxation. We found that DMD trajectories of 3 ns were sufficient to reach equilibrium. Clearly, the DMD procedure leads to a decrease in the energy of all conformations, but near-native conformations experience a deeper energy improvement, populating the best-scoring docking poses after DMD relaxation. In terms of predictive results, this case was not that bad for rigid-body pyDock, since there was another near-native solution with rank 7 (Table 1), albeit with higher RMSD (3.7 Å), but DMD relaxation was more efficient in the near-native solutions closer to the bound state and yielded an overall improvement in the docking energy landscapes and in the predictive results.

The results of the described docking and DMD procedure on the entire data set of complexes (the largest benchmark available) demonstrate that DMD relaxation significantly improves the success rate of the docking predictions, by systematically improving the ranking of the near-native docking poses. For instance, the success rate for the top 10 scoring structures (which is a reasonable number of docking models in a realistic scenario, e.g, the number of models submitted to CAPRI) increased from 39% for the rigid-body procedure to 50% after 3 ns DMD relaxation (Figure 2A). Interestingly, the improvement in success rates was already evident from the very beginning of the DMD trajectory. After only 200 ps, the general results were very similar to those after 3 ns, suggesting that the fast side-chain movements have a fundamental influence on the improvement in energy of the docking conformations (as discussed below).

Improvement of Predictions after DMD Relaxation Is More Evident for Flexible Cases. In order to check which types of complexes benefit the most from DMD relaxation, we reanalyzed our results by grouping cases depending on the extent of the binding-induced geometrical changes upon interaction. Results summarized in Figure 2C indicate that 3 ns DMD leads to a very significant improvement in the ranking of docking poses in the case of complexes showing large conformational changes upon binding (interface deformation above 1.0 Å RMSD). For these cases, the success rates obtained with DMD for the top 1 and top 10 docking poses were 8% and 48%, respectively, i.e., twice those obtained with rigid-body docking. Among these flexible cases, there were actually seven cases in which the results significantly improved (best near-native rank went from >10 to <10 after DMD relaxation), while only one successful docking case became significantly worse after DMD relaxation. In the case of complexes with low conformational change upon binding, the success rates obtained with DMD

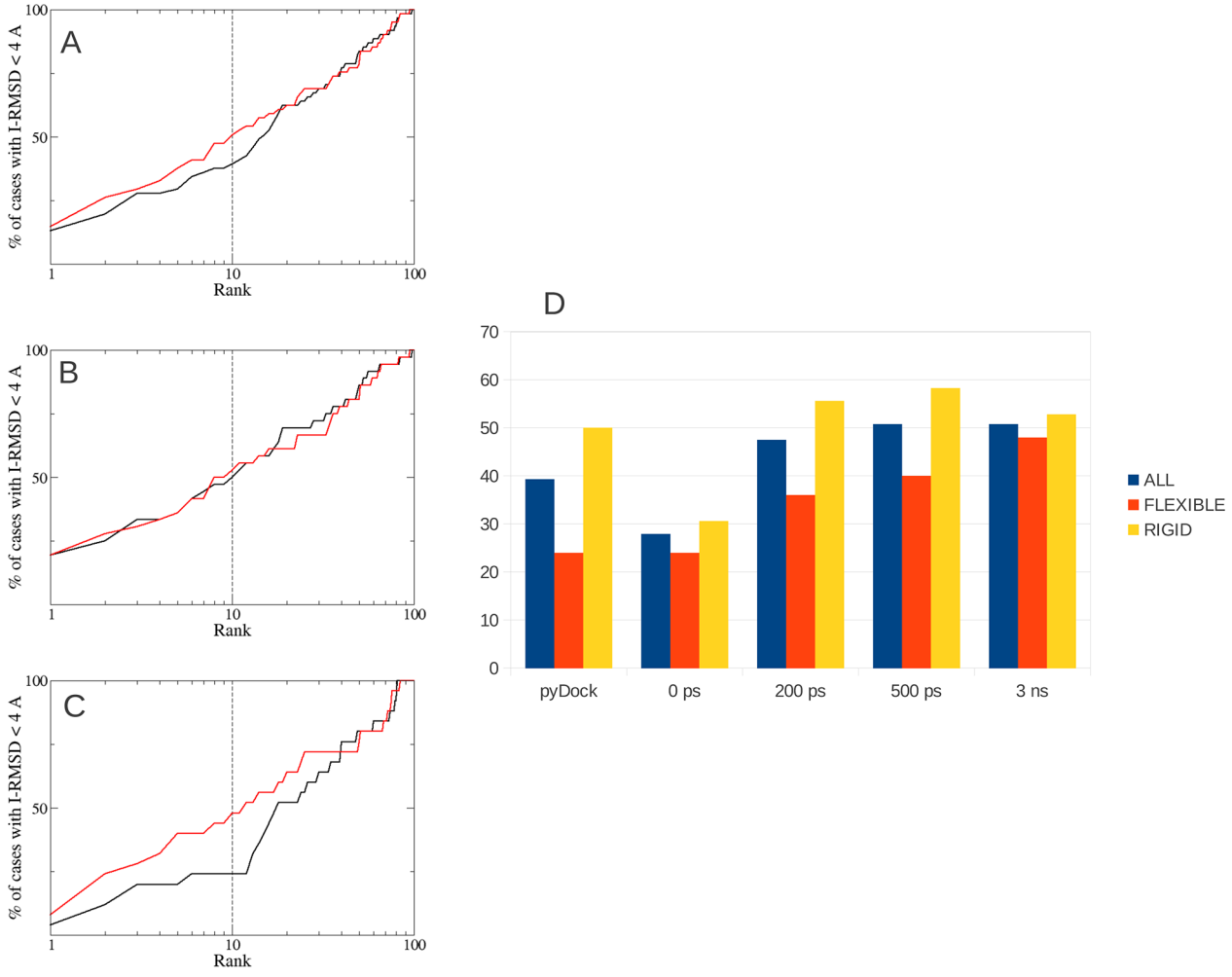


Figure 2. Docking success rates computed for the data set of protein−protein complexes used here. Black line, rigid-body docking ranked with the pyDock scoring function; red lines, docking structures relaxed after 3 ns DMD simulations. (A) Success rates for the entire data set. (B) Success rates for complexes with small conformational changes upon binding (interface RMSD unbound−bound <1 Å). (C) Success rates for complexes with large conformational changes upon binding (interface RMSD unbound−bound >1 Å). (D) Success rates for top 10 docking solutions at different times of the simulation, for rigid and flexible cases.

dx.doi.org/10.1021/ct301039e | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

were similarly high to those obtained with rigid docking (Figure 2B). There were five cases that significantly improved and another five that significantly worsened. A Wilcoxon rank-sum test proved that the improvement in the results after DMD relaxation is significantly more evident for flexible cases (see Outline of the Method). This is in an important result, since a flexible refinement should not only improve the results in flexible cases but also not degrade the quality of docking models where binding-induced conformational changes are negligible. It has been previously reported that optimization of non-native interfaces may increase the number of false positives,[12] which does not seem to happen here.

The pace of improvement of the success rates along the DMD simulation is also different for the rigid and flexible cases. While for the rigid cases (Figure 2B) the optimal predictive results were achieved from the very beginning (at 200 ps, success rates are very similar to those after 3 ns), for the flexible cases (Figure 2C) the improvement of the predictive rates with respect to rigid-body docking is more progressive and actually shows the best results at 3 ns. For the purpose of clarity, Figure 2D shows the evolution of the top 10 success rates for rigid and flexible cases at different stages of the DMD trajectories. This confirms that DMD relaxation does not further improve the already good docking success rates in rigid cases, while flexible cases need some conformational rearrangement before improving docking success rates to achieve values more similar to the rigid cases. We will explore in the next section the nature of these conformational rearrangements.

Structural and Energetic Determinants of DMD Docking Success. The DMD relaxation procedure leads to non-negligible changes in the complex conformations obtained by rigid-body docking, with extensive modification of the whole interface (involving not only side chain but also backbone movements). In many cases, the deformation of both ligand and receptor along DMD relaxation, as it occurs in any standard molecular dynamics simulations,[40] exceeds the typical conformational changes upon binding (around 1−2 Å RMSD). In some successful cases, like 1Z0K, the relaxation makes the majority of the interface of a docking solution get closer to the crystallographic complex structure, which can explain the improvement in the predictions (Figure 3A). The conformational changes are shown to be similar to those we found making explicit solvent MD simulations with GROMACS. Indeed, the best near-native solution in this case conserved the same number of native contacts during DMD relaxation as the native complex (Figure 4A), which made this solution achieve the same energy as the native complex (Figure 4B). However, in other successful cases like 1M10, DMD relaxation cannot reproduce the whole native interface (Figure 3B; Figure 4D). In the past, it has been reported that refinement can be beneficial just by reorientation of the side chains to adopt conformations that improve the binding energy, even though overall interface conformation does not necessarily get closer to that in the native complex.[15] This seems to be the case here. We can explain the reasons for DMD improvement, in spite of the interface deformation, in terms of the type of side-chain contacts formed during simulation. Figure 4F shows the evolution of the native contacts that form attractive interactions during simulation for different docking solutions in the 1M10 case. The near-native solution shows a rapid formation of a few attractive native contacts that are ultimately responsible for its favorable energy. Although its energy value is not as good as that of the native complex (Figure 4E), it is sufficient to improve the rank of this near-native solution from 81 to 2 after DMD. Simultaneously, the number of repulsive contacts dramatically reduces in the first stages of the DMD relaxation (data not shown).

Overall, these results indicate that during the DMD simulation the structures relax to reduce the repulsive contacts that arose from incorrect rigid unbound conformations (between regions with electric charges of the same sign, steric clashes, unfavorable desolvation, etc.) and to increase the attractive native interactions found in the experimental complex structures (hydrophobic, salt bridges, etc.). This effect should be especially important in the flexible cases, in which DMD relaxation represents a greater advantage with respect to rigid approaches. As a consequence, the relaxation method is very efficient in improving the energy of many of the near-native conformations, which is reflected in the improvement of the docking predictions.

Can We Identify a Priori the Cases That Will Benefit from DMD Relaxation? We showed that proteins that undergo large conformational movements upon binding are

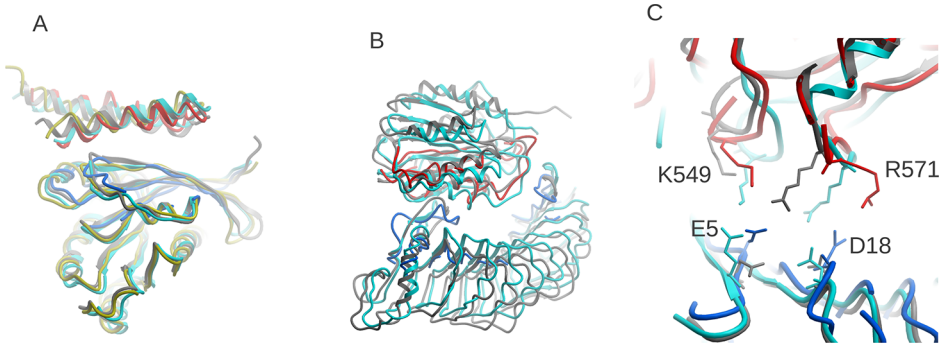


Figure 3. Conformational changes in near-native docking conformations after DMD relaxation. (A) Rank 1 near-native docking conformation of the complex 1Z0K after DMD relaxation (receptor in blue, ligand in red). The rigid docking conformation (initial structure) is shown in gray. For comparison, the experimental complex structure is shown in light blue, and the relaxed structure after simulation with GROMACS is shown in yellow. The relaxed protein−protein interface after DMD is getting closer to the experimental complex interface. (B) Rank 1 near-native docking conformation of the complex 1M10 after DMD relaxation (same color code as in A). In spite of the improvement in ranking, the relaxed protein−protein interface is not closer to the complex structure. (C) Key residue−residue native contacts for the 1M10 complex are salt-bridges between glycoprotein IB-α E5 and Von Willebrand Factor K549, and between D18 and R571 (in cyan). These were formed in the rank 2 conformation after DMD relaxation (receptor in blue, ligand in red), but they were not formed in the rigid-body docking solution before relaxation (in gray).

dx.doi.org/10.1021/ct301039e | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

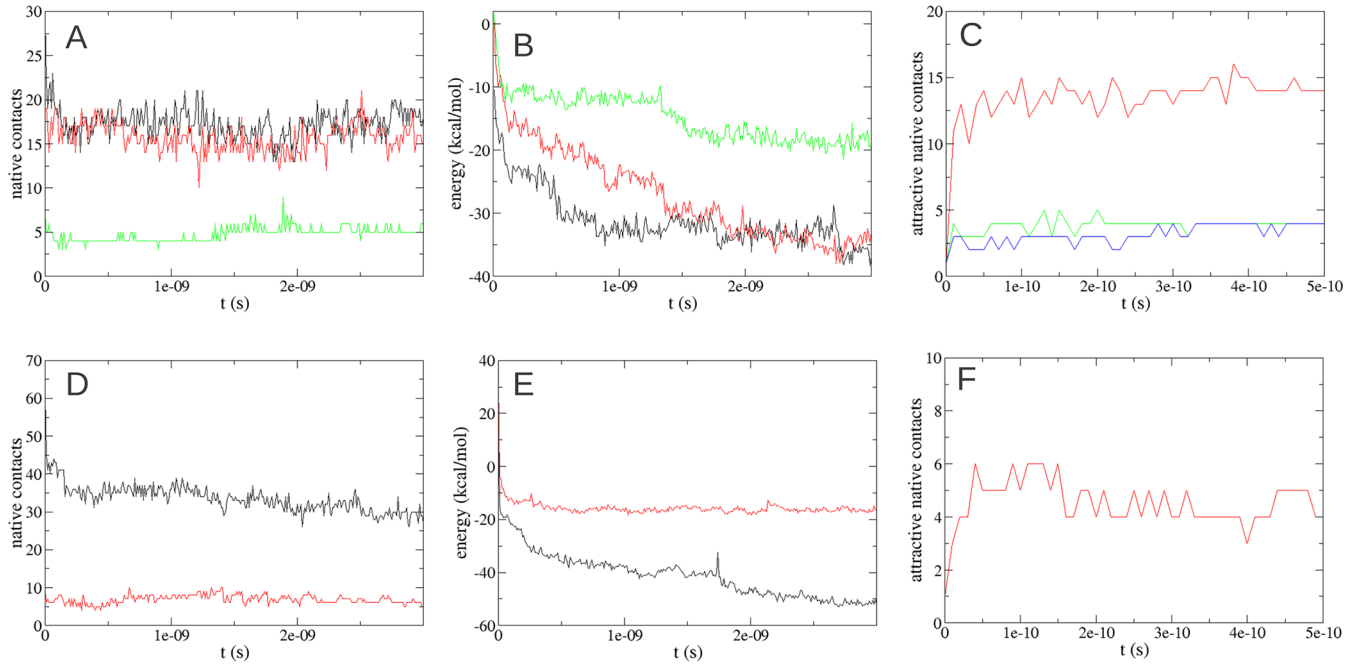


Figure 4. Analysis of DMD relaxation trajectories for two successful cases. (A) Native contacts, (B) energy, and (C) native attractive contacts along DMD relaxation for different conformations in the 1Z0K case. (D, E, F) Same for 1M10 case. The color code is as follows: experimental complex structure in black; best near-native conformation after DMD relaxation in red; other near-native conformations in green; other non-near-native conformations in blue.

the ones that most benefit from DMD relaxation. However, proteins that do not change conformation upon binding do not need DMD relaxation, as rigid-body docking already gives good predictions for them. Thus, in a realistic situation it would be interesting to know whether DMD relaxation is going to be useful for a given case or not. For that, we can assume that flexibility of proteins is related to the conformational variability upon binding; hence we would just need a method to estimate conformational flexibility of the individual proteins. Here, we have estimated the extent of the conformational motion of a protein based on NMA[16] (see Outline of the Method). It is important to notice that this method to estimate the extent of the conformational changes is valid as long as the normal mode hypothesis is valid for a given protein; therefore it might happen that a protein has a conformational change upon binding that is very different from the change predicted from NMA. In addition, global flexibility of a protein based on NMA might not need to correlate with interface deformation upon binding, which has been shown here to be critical for the performance of DMD. Figure 5 shows the top 10 success rates after DMD relaxation for the cases that were classified as rigid or flexible based on normal mode calculations. Interestingly, the improvement of success rates after DMD relaxation for the predicted flexible cases is evident, virtually the same that we obtained for the real flexible cases. This result indicates that, based on simple normal mode calculations, one could estimate whether DMD relaxation is going to be useful for a given protein−protein docking problem.

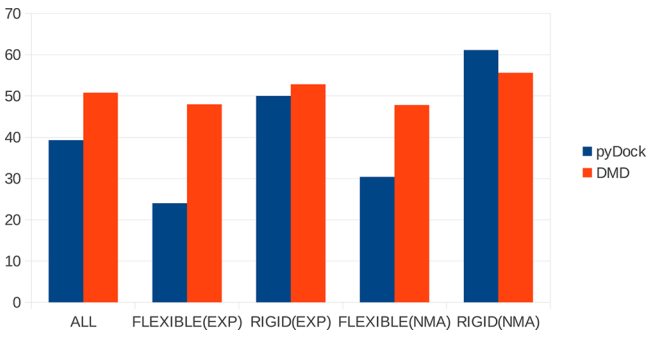


Figure 5. Top 10 success rates for the predicted rigid and flexible cases according to normal-mode analysis. For comparison, the top 10 success rates for the known rigid and flexible cases, as well as for the whole data set, are also shown.

CONCLUSIONS

We outlined here a fast, simple, and efficient protocol for protein−protein docking, which combines rigid-body orientation sampling with structural relaxation by means of discrete molecular dynamics (DMD) simulations. The procedure takes advantage of the intrinsic characteristics of DMD, such as (i) the use of nondifferentiable square potentials, typically used in docking procedures but difficult to implement in Newtonian molecular dynamics, (ii) the high computational efficiency of the algorithm to trace large scale movements, (iii) the simplicity of using the method in the context of multiresolution representations of the proteins, and (iv) the ability of the method to maintain intramolecular geometry in some parts of the monomer, while keeping intact flexibility in other regions and maintaining the global inter-residual flexibility. We have found that the predictive power of the DMD relaxation method is much less affected by the deformation upon binding of the ligand and receptor as compared to rigid-body docking methods, leading to a clear improvement in the predictive success rates over the complete benchmark. Finally, we found that it is possible to estimate a priori whether a given case is going to benefit from DMD relaxation.

dx.doi.org/10.1021/ct301039e | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

AUTHOR INFORMATION

Corresponding Author

*E-mail: modesto.orozco@irbbarcelona.org (M.O.) or juanf@bsc.es (J.F.-R.).

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We thank Laura Orellana for help with the NMA calculations. This work has been supported by grant number BIO2010-22324 (J.F.-R.) and BIO2009-10964 (M.O.) from MICINN-Spain, the European Research Council (ERC-Advanced Grant, M.O.), the Instituto Nacional de Bioinformatica (INB; M.O., J.L.G.), the Consolider E-Science Project (M.O.), and the Fundación Marcelino Botin (M.O.). P.S. is a fellow of the La Caixa doctoral program. M.O. is an ICREA Academia Fellow.

REFERENCES

(1) Bernstein, F. C.; Koetzle, T. F.; Williams, G. J. B.; Meyer, E. F.; Brice, M. D.; Rodgers, J. R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. *J. Mol. Biol.* **1977**, *112*, 535−542.

(2) Stein, A.; Mosca, R.; Aloy, P. *Curr. Opin. Struct. Biol.* **2011**, *21*, 200−208.

(3) Venkatesan, K.; Rual, J.-F.; Vazquez, A.; Stelzl, U.; Lemmens, I.; Hirozane-Kishikawa, T.; Hao, T.; Zenkner, M.; Xin, X.; Goh, K.; Yildirim, M. A.; Simonis, N.; Heinzmann, K.; Gebreab, F.; Sahalie, J. M.; Cevik, S.; Simon, C.; Smet, A.-S.; Dann, E.; Smolyar, A.; Vinayagam, A.; Yu, H.; Szeto, D.; Borick, H.; Dricot, A.; Klitgord, N.; Murray, R. R.; Lin, C.; Lalowski, M.; Timm, J.; Rau, K.; Boone, C.; Braun, P.; Cusick, M. E.; Roth, F. P.; Hill, D. E.; Tavernier, J.; Wanker, E. W.; Barabási, A.-L.; Vidal, M. *Nat. Methods* **2009**, *6*, 83−90.

(4) Halperin, I.; Ma, B.; Wolfson, H.; Nussinov, R. *Proteins* **2002**, *47*, 409−443.

(5) Ritchie, D. *Curr. Protein Pept. Sci.* **2008**, *9*, 1−15.

(6) Pons, C.; Grosdidier, S.; Solernou, A.; Pérez-Cano, L.; Fernández-Recio, J. *Proteins* **2010**, *78*, 95−108.

(7) Fischer, H. E. *Chem. Ber.* **1894**, *27*, 2985−2993.

(8) Stein., A.; Rueda, M.; Panjkovich, A.; Orozco, M.; Aloy, P. *Structure* **2011**, *19*, 881−889.

(9) Pons, C.; D'Abramo, M.; Svergun, D. I.; Orozco, M.; Bernadó, P.; Fernández-Recio, J. *J. Mol. Biol.* **2010**, *403*, 217−230.

(10) Zacharias, M. *Curr. Opin. Struct. Biol.* **2010**, *20*, 180−186.

(11) Bonvin, A. M. J. J. *Curr. Opin. Struct. Biol.* **2006**, *16*, 194−200.

(12) Fernández-Recio, J.; Totrov, M.; Abagyan, R. *Protein Sci.* **2002**, *11*, 280−291.

(13) Fernández-Recio, J.; Totrov, M.; Abagyan, R. *Proteins* **2003**, *52*, 113−117.

(14) Fernández-Recio, J.; Abagyan, R.; Totrov, M. *Proteins* **2005**, *60*, 308−313.

(15) Dominguez, C.; Boelens, R.; Bonvin, A. M. J. J. *J. Am. Chem. Soc.* **2003**, *125*, 1731−1737.

(16) Dobbins, S. E.; Lesk, V. I.; Sternberg, M. J. E. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 10390−10395.

(17) Zacharias, M. *Proteins* **2004**, *54*, 759−767.

(18) Wang, C.; Bradley, P.; Baker, D. *J. Mol. Biol.* **2007**, *373*, 503−519.

(19) Gray, J. J.; Moughon, S.; Wang, C.; Schueler-Furman, O.; Kuhlman, B.; Rohl, C. A.; Baker, D. *J. Mol. Biol.* **2003**, *331*, 281−299.

(20) Chen, R.; Mintseris, J.; Janin, J.; Weng, Z. *Proteins* **2003**, *52*, 88−91.

(21) Zhou, Y. Q.; Karplus, M. *Nature* **1999**, *401*, 400−403.

(22) Emperador, A.; Meyer, T.; Orozco, M. *Proteins* **2010**, *78*, 83−94.

(23) Urbanc, B.; Cruz, L.; Yun, S.; Buldyrev, S. V.; Bitan, G.; Teplow, D. B.; Stanley, H. E. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 17345−17350.

(24) Ding, F.; Buldyrev, S. V.; Dokholyan, N. V. *Biophys. J.* **2005**, *88*, 147−155.

(25) Ding, F.; Sharma, S.; Chalasani, P.; Demidov, V. V.; Broude, N. E.; Dokholyan, N. V. *RNA* **2008**, *14*, 1164−1173.

(26) Ding, F.; LaRoque, J. J.; Dokholyan, N. V. *J. Biol. Chem.* **2005**, *280*, 40235−40240.

(27) Nguyen, H.; Hall, C. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 16180−16185.

(28) Ding, F.; Borreguero, J. M.; Buldyrev, S. V.; Stanley, H. E.; Dokholyan, N. V. *Proteins* **2003**, *53*, 220−228.

(29) Sfriso, P.; Emperador, A.; Orellana, L.; Hospital, A.; Gelpi, J. L.; Orozco, M. *J. Chem. Theory Comput.* **2012**, *8*, 4707−4718.

(30) Alder, B. J.; Wainwright, T. E. *J. Chem. Phys.* **1959**, *31*, 459−466.

(31) Smith, W. S.; Hall, C. K.; Freeman, B. D. *J. Comput. Phys.* **1997**, *134*, 16−30.

(32) Hwang, H.; Vreven, T.; Janin, J.; Weng, Z. *Proteins* **2010**, *78*, 3111−3114.

(33) Gabb, H. A.; Jackson, R. M.; Sternberg, M. J. E. *J. Mol. Biol.* **1997**, *272*, 106−120.

(34) Cheng, T.; Blundell, T. L.; Fernández-Recio, J. *Proteins* **2007**, *68*, 503−515.

(35) Grosdidier, S.; Pons, C.; Solernou, A.; Fernández-Recio, J. *Proteins* **2007**, *69*, 852−858.

(36) Pons, C.; Solernou, A.; Pérez-Cano, L.; Grosdidier, S.; Fernández-Recio, J. *Proteins* **2010**, *78*, 3182−3188.

(37) Fernández-Recio, J.; Totrov, M.; Abagyan, R. *J. Mol. Biol.* **2004**, *335*, 843−865.

(38) Lazaridis, T.; Karplus, M. *Proteins* **1999**, *35*, 133−152.

(39) Rueda, M.; Chachon, P.; Orozco, M. *Structure* **2007**, *15*, 565−575.

(40) Rueda, M.; Ferrer-Costa, C.; Meyer, T.; Pérez, A.; Camps, J.; Hospital, A.; Gelpí, J. L.; Orozco, M. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 796−801.