Consistent Improvement of Cross-Docking Results Using Binding Site Ensembles Generated with Elastic Network Normal Modes

Manuel Rueda,[†] Giovanni Bottegoni,[†] and Ruben Abagyan*[,†,‡]

Department of Molecular Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, Mail TPC-28, La Jolla, California 92037, and Molsoft, LLC, 3366 North Torrey Pines Court, Suite 300, La Jolla, California 92037

Received October 8, 2008

The representation of protein flexibility is still a challenge for the state-of-the-art flexible ligand docking protocols. In this article, we use a large and diverse benchmark to prove that is possible to improve consistently the cross-docking performance against a single receptor conformation, using an equilibrium ensemble of binding site conformers. The benchmark contained 28 proteins, and our method predicted the top-ranked near native ligand poses 20% more efficiently than using a single receptor. The multiple conformations were derived from the collective variable space defined by all heavy-atom elastic network normal modes, including backbone and side chains. We have found that the binding site displacements for best positioning of the ligand seem rather independent from the global collective motions of the protein. We also found that the number of binding site conformations needed to represent nonredundant flexibility was <100. The ensemble of receptor conformations can be generated at our Web site at http://abagyan.scripps.edu/MRC.

INTRODUCTION

Molecular docking is routinely used within the first stages of Structural-Based Drug Design (SBDD). Unfortunately, despite ongoing methodological advances and some success histories, state-of-the-art algorithms predict an incorrect binding pose for ∼50%−70% of all ligands when a single receptor conformation is used.[1]

The need for including the receptor flexibility in docking comes after much experimental and theoretical evidence showing that proteins move at room (or physiological) temperature and rearrange in response to binding.[2−4] Inspection of structures from the Protein Data Bank clearly shows that different binding site conformations of the same protein exist for different ligands.[5] Such compelling arguments led to the replacement of the *lock and key* paradigm (where the protein is assumed to be rigid) by a more-realistic view where the ligands interact with an ensemble of differently populated conformational states in equilibrium.[6,7] A ligand could bind with different affinities to one (or some) of such *pre-existing equilibrium* conformations, inducing the protein to populate infrequent states by shifting their population distribution, namely, *induced fit*.

Recent advances in the description of protein flexibility in the context of docking have been published elsewhere,[1,5−9] and it is beyond of the scope of this paper to give an extensive revision of them. Among the proposed protocols, one of the most promising alternative is the use of ensembles of structures, namely Multiple Receptor Conformations (MRCs), where the ligand is docked not to just one receptor conformation, but rather to a set of receptor conformations.[6] The conformational ensemble can be provided by experimental methods such as X-ray crystallography or nuclear magnetic resonance (NMR) spectroscopy;[10−15] however, this information may only be partially available or not at all.

Probably the most rigorous approach to date for the atomistic description of flexibility is molecular dynamics (MD) simulation, insofar as it represents the dynamics of the molecules in physiological environments using rigorous physical potentials.[16,17] MD has been successfully employed in the context of docking;[18−26] however, high requirements, both in terms of computational resources and human expertise, strongly limit its routine application in the field. Also, a certain level of concern exists regarding the selection of representative structures for docking among the vast pool of generated conformations. In this regard, it is possible to isolate functional motions from purely thermal noise from the MD trajectories using principal component analysis (PCA) of the covariance of atomic fluctuations, which is usually called essential dynamics.[27] This collective space is very stable and highly dependent on the molecule topology,[28,29] so that similar results can be obtained with different approximations. This is the case of motions obtained from anharmonic MD simulation, relative to simplistic methods for exploring the equilibrium dynamics such as harmonic Normal Mode Analysis (NMA),[29−37] which is widely used as a tool to describe flexibility and has been already applied in the context of docking.[38−43] The fact that binding sites in globular proteins are rigid,[44−46] packed as tightly as a solid,[47] along with its simplicity and speed, makes NMA especially suitable for docking, representing a viable alternative to the costly MD simulations.

Here, we propose a fast and general in silico method to represent the equilibrium dynamics of the receptor, where a small number of MRC conformations are generated within all heavy atom (backbone and side chains) anisotropic Elastic

* Author to whom correspondence should be addressed. E-mail address: abagyan@scripps.edu.
[†] Department of Molecular Biology, The Scripps Research Institute.
[‡] Molsoft, LLC.

10.1021/ci8003732 CCC: $40.75 © 2009 American Chemical Society
Published on Web 02/16/2009

Table 1. List of Crystallographic Structures Included in the Cross-Docking Benchmark

			Ligand Information	
protein	receptor PDB chain	resolution (Å)	name	function
Cyclin dependent kinase-2	1AQ1.a	2.0	Staurosporine	Inhibitor
	1DM2.a	2.1	Hymenialdisine	Inhibitor
Cyclooxygenase-2	1CX2.a	3.0	SC-588	Inhibitor
	3PGH.a	2.5	Flurbiprofen	Inhibitor
Estrogen receptor	1ERR.a	2.6	Raloxifene	SERM[a]
	3ERT.a	1.9	4-Hydroxytamoxifene	SERM[a]
Factor Xa	1KSN.a	2.1	FXV673	Inhibitor
	1XKA.c	2.3	FX2212-A	Inhibitor
Glycogen synthase kinase-3 β	1Q4L.a	2.7	I-5	Inhibitor
	1UV5.a	2.8	6-Bromoindirubin-3′-Oxime	Inhibitor
Human immunodeficiency virus-1 reverse transcriptase	1C1C.a	2.5	TNK-6123	Inhibitor
	1RTH.a	2.2	1051U91	Inhibitor
c-Jun N-terminal kinase-3	1PMN.a	2.2	Compound 1	Inhibitor
	1PMV.a	2.5	Compound 4	Inhibitor
Liver X receptor-β ligand binding domain	1P8D.a	2.8	24(S), 25-Epoxycholesterol	Agonist
	1PQ6.b	2.4	GW3965	Agonist
Neuraminidase	1A4Q.a	1.9	Dihydropyran-phenethyl-propyl-carboxamide 5d	Inhibitor
	1NSC.a	1.7	O-Sialic Acid	Inhibitor
P38 mitogen activated protein kinase	1BMK.a	2.4	SB218655	Inhibitor
	1DI9.a	2.6	4-anilinoquinazoline 3	Inhibitor
Protein kinase A	1STC.e	2.3	Staurosporine	Inhibitor
	1YDS.e	2.2	H8	Inhibitor
Peroxisome proliferator activated receptor-γ	1FM9.d	2.1	G1262570	Agonist
	2PRG.a	2.3	Rosiglitazone	Agonist
Timidine kinase	1KI4.a	2.3	5-Bromothienydeoxyuridine	Inhibitor
	1KIM.a	2.1	Deoxythymidine	Endogenous Substrate
Trypsin	1PPC.e	1.8	NAPAP	Inhibitor
	1PPH.e	1.9	3-TAPAP	Inhibitor

[a] Selective estrogen receptor modulator.

Network Normal Modes Analysis (EN·NMA) space.[28] In contrast to previous results,[41,42] the method does not require a priori knowledge of the region to sample and can be applied to any experimental structure without further refinement steps,[39] significantly reducing the computational time. The performance of the method has been evaluated with ICM[48,49] in a complete cross-docking benchmark containing 28 holoconformations, corresponding to 14 targets of interest in biomedical applications[50] (see Table 1) and compared with docking to single experimental receptor conformation (SRC). The algorithm concentrates near-native cross-ligand geometries (at a distance of 2 Å (root mean square deviation (rsmd)) from the crystallographic pose) as the first ranked solutions 20% more efficiently than a single experimental conformation and provides stronger binding score energies, which might be interesting from a predictive point of view.[51] Three of the 28 conformations (1P8D, 1PMV, and 2PRG) present strong memory of the interactions and placement of their cognate ligand. For these cases, the sum of harmonic potentials seems inappropriate to model the fluctuations that are needed to accommodate new ligands properly. Based on our results, we can speculate that, for such extreme induced-fit changes, neither conformations derived from EN·NMA nor state-of-the-art MD done without the cross-ligand in the pocket will provide optimum solutions in cross-docking experiments. Difficult cases may still be solvable using more time-consuming methods that explicitly include the exogenous ligand in the refinement of the protein structure, such as, for example, the procedure of Sherman et al.,[52] SCARE,[50] longer MD simulations,[19] or enhanced sampling methods.[53]

RESULTS AND DISCUSSION

The proposed method consists of three major consecutive steps (see Figure 1): (i) the generation of the multiple receptor conformations, (ii) the docking of the ligand to the representative protein conformations, and (iii) the final ICM scoring of the poses. The protocol is built on top of a well-established and optimized ICM docking and scoring algorithm to a single static receptor conformation (SRC).[48,49] The generation of the binding site conformations is described in detail below.

1. Generation of Multiple Receptor Conformations through Normal Modes Ensemble. Previous studies[39,41] had already shown the capacity of coarse-grained normal-mode analysis with elastic potentials for the representation of the trace flexibility of protein kinases (Cα) in docking context, especially when the region of interest is known. The use of the Cα as descriptors has the advantage of speed, because it drastically reduces the time needed for the diagonalization of the Hessian (i.e., $3N \times 3N$ matrix, where N is the number of atoms) of a standard protein (e.g., 300 residues). However, the concerted fluctuations of the atoms from backbone and side chains are not explicitly represented.

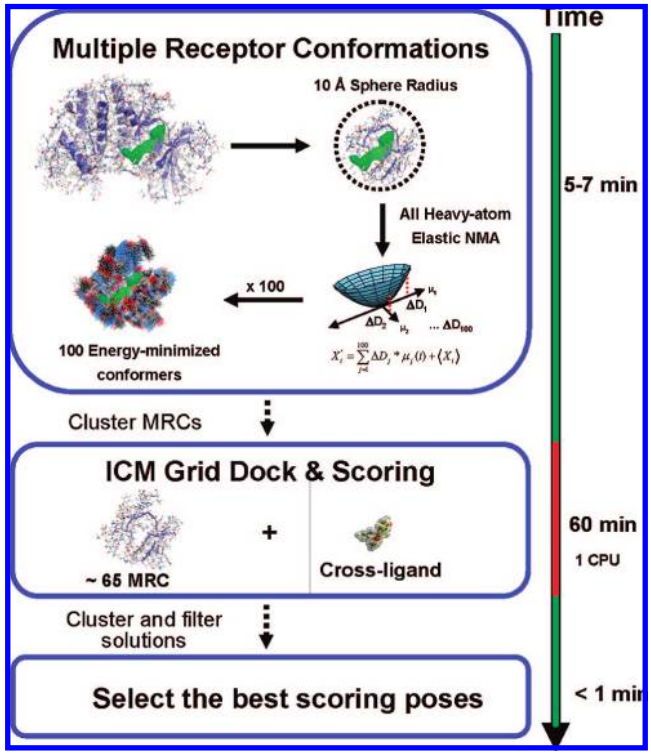

Figure 1. Schematic outline of the Normal Mode Analysis Multiple Receptor Conformation (NMA MRC) algorithm.

Here, we address the same problem from a different perspective, assuming that the initial specificity of the ligand is dependent on local fluctuations of all the atoms that define the binding site.

1.1. Effect of the Sphere Radius. The first issue that we explored here was the optimum radius of the sphere for the residues to be included in EN·NMA. The need for a cut-off radius has two clear implications: one that is practical, because the computational speed and memory requirements for the diagonalization of the Hessian increases exponentially with the number of atoms (see Figure S1 in the Supporting Information), and second, that each coordinate adds one dimension to the orthonormal space, diluting the local binding site fluctuations into the global protein movement.

For that purpose, we selected the heavy atoms from residues within a distance of 10, 15, and 20 Å from the cognate ligand, leaving a reasonable space between the residues implied in docking and the boundaries (see Figure 2a), performed EN·NMA (see Methods for details), and generated 100 different conformers for each distance. For each MRC, independent docking calculations were performed where the bounding box was derived from residues within 5 Å from the cognate ligand. As can be observed in Figure S2 in the Supporting Information, the best results are obtained globally when a radius of 10 Å is chosen, which implies the use of 63−125 amino acids, including 482−1017 heavy atoms (see Table 2), depending on the location of the binding site (surface, buried) and the packing of the protein.

To quantify the effect of the remainder of the protein in the docking performance, we repeated the experiments using (i) a multilevel representation of the protein,[54] where we used an all heavy atom definition for the binding site and Cα for the remainder, and (ii) all the heavy atoms for the complete protein. According to our results, the inclusion of more atoms in the EN·NMA did not improve the docking performance in our benchmark, resulting in near-native solution (NNs) ranks that are similar to those obtained with a radius of >10 Å. Therefore, it seems unnecessary to include them in the EN·NMA calculation unless a structural argument exists for the coupling of global and local movements (e.g., hinge movements, domain movement or coupled motions). These results are consistent with those obtained for bacteriorhodopsin, which suggested that the dynamics of the active site were different from the global protein movement.[55,56]

1.2. Calibration of the Elastic Network Heavy Atom Potential. After the residues are chosen, the ligand is removed and an EN·NMA is performed on the receptor (see Methods section), where the structure is chosen as a minimum and an inverse exponential function is adopted to account for the distance dependence force constants.[57] This approximation shows better correlation with atomistic MD, with respect to a classical pure cutoff method,[58] where all the atoms within a cutoff have the same elastic constant.

The relative amplitude of the displacements in an elastic network is empirical and must be defined by the value of the spring constants between the atoms. To generate realistic fluctuations, the springs were calibrated so that the binding pocket residues fluctuate in a similar manner, not only qualitatively but also quantitatively, to those obtained with atomistic MD simulation for the protein Aldose Reductase (from the MoDEL database;[59] PDB Code: 1AH0). (See Figure S3a in the Supporting Information.) After the minimization, the average heavy atom rmsd of the ensembles against the X-ray (see Table 2) was ∼1 Å (0.8 Å for backbone), which is in the same range of the rmsd between each pair of conformations.[50] As seen previously,[20,60] such a slight fluctuation dramatically improves the sterical interactions between the receptor and the ligand (see Figure 2). Note that other qualitative and quantitative descriptions of the topology of the protein can be easily implemented, as for example, the calibration of the elastic constants according to X-ray densities. The method will also benefit from more-accurate potential energy functions that contain bonded and nonbonded terms, such as those used in standard NMA.[37]

1.3. Definition of the "Important Subspace". The next step is to generate Cartesian MRC structures along the collective EN·NMA space that can be used in docking simulations. The diagonalization of the Hessian provides a set of $3N - 6$ eigenvectors, ranked according to the corresponding eigenvalues. The eigenvectors are the collective directive of motions of atoms, and the eigenvalues give the energy cost of deforming the system by one unit length along the eigenvectors. When using a spherical radius of 10 Å, we involve ∼675 heavy atoms on average, resulting in 2010 eigenvectors (or, in other words, 2010 collective directions to derive movements). However, not all the modes contribute equally to the variance and, in fact, only a subspace provides significant movement (see dimensionality[61] in Table 2). In an attempt to establish the size of such subspace, we calculated the correlation between MD and EN·NMA spaces for the Aldose Reductase protein (see Figure S3b in the Supporting Information). The dot product of both orthonormal spaces increases up to ∼100 eigenvectors, indicating that a subspace that contains 5% of the vectors has a correlation of 60%, which is surprisingly high (∼250

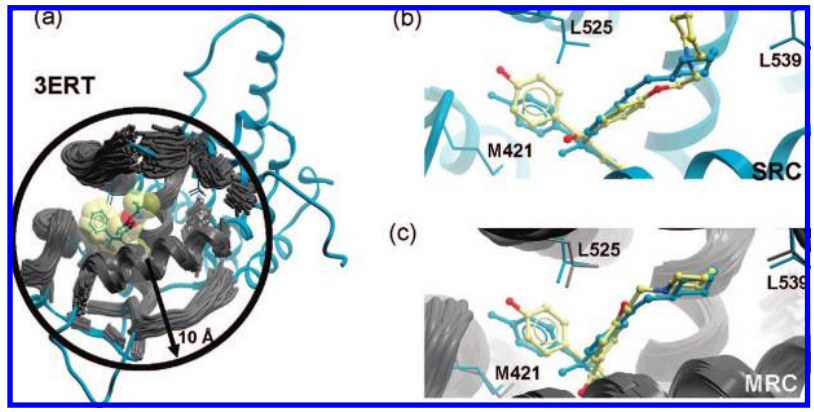

Figure 2. MRC methodology applied to the estrogen receptor protein, PDBid: 3ERT: (a) example of the residues included in the elastic network normal-mode analysis (EN•NMA) with a spherical radius of 10 Å around 4-hydroxytamoxifene; (b) cross-docking of raloxifene in the pocket of conformation of 3ERT (mainly due to the positions of M421, L523 and L539, the best result that can be achieved using a single receptor is a high-energy pose at 1.7 Å (rmsd) to the crystallographic position); and (c) using MRC, is it possible to obtain a very-low-energy ligand pose at 1.2 Å (rmsd). In all panels, the experimental structures are reported in blue, the multiple receptor conformations are gray, and the cross-docking ligand poses are colored based on atom type.

Table 2. Structural and Flexibility Descriptors Associated to the Proteins Included in the Benchmark

protein	PDB	BS volume ($Å^3$)[a]	#HA 8 Å from BS	#HA 10 Å from SL	dimensionality[b]	variance (A^2)[c]	rmsd to X-ray (Å)[d]	number of clusters[e]
CDK2	1AQ1	428.6	575	613	92	634.6	1.0 (0.8)	43
	1DM2	423.7	560	535	83	604.4	1.0 (0.8)	28
COX-2	1CX2	226.0	659	783	99	553.3	0.9 (0.7)	77
	3PGH	198.2	593	774	100	515.1	0.9 (0.7)	58
Estrogen receptor	1ERR	300.2	654	780	84	518.1	0.9 (0.7)	88
	3ERT	363.7	651	741	108	831.5	1.0 (0.8)	58
factor Xa	1KSN	392.5	557	657	81	585.5	0.9 (0.6)	95
	1XKA	286.6	583	664	68	420.6	0.8 (0.6)	79
GSK-3 β	1Q4L	491.2	544	516	74	544.4	1.0 (0.8)	80
	1UV5	530.2	651	549	71	536.7	1.0 (0.8)	92
HIV-1 RT	1C1C	481.3	545	569	79	538.8	1.0 (0.8)	82
	1RTH	475.8	515	535	82	653.9	1.1 (0.9)	78
JNK3	1PMN	829.2	965	655	108	820.4	1.1 (0.9)	80
	1PMV	822.6	845	482	75	577.6	1.1 (0.9)	86
LXR β LBD	1P8D	572.4	960	908	109	537.4	0.9 (0.6)	35
	1PQ6	508.7	843	934	133	660.1	0.8 (0.6)	12
Neuraminidase	1A4Q	235.8	572	695	73	469.6	0.9 (0.7)	86
	1NSC	241.1	533	653	59	370.0	0.8 (0.7)	86
P38 MAP kinase	1BMK	444.8	748	529	81	557.3	1.1 (0.9)	98
	1DI9	137.1	416	600	98	758.4	1.1 (0.9)	56
PKA	1STC	640.9	898	713	103	658.9	0.9 (0.7)	72
	1YDS	381.5	757	605	80	483.1	0.9 (0.7)	16
PPARγ	1FM9	1145.2	1138	1017	126	621.0	0.9 (0.6)	27
	2PRG	896.2	1006	793	122	697.2	1.0 (0.8)	86
TK	1KI4	342.0	606	770	81	408.4	0.8 (0.6)	56
	1KIM	294.1	613	735	78	432.9	0.9 (0.6)	58
Trypsin	1PPC	212.9	440	586	71	512.6	1.1 (0.8)	82
	1PPH	206.9	438	560	78	620.0	1.0 (0.8)	52

[a] Computed with ICM. [b] The dimensionality is defined as the number of normal modes with oscillations of >2 $Å^2$. [c] The variance described by the first 100 modes. [d] Average all heavy atom rmsd from MRC to X-ray structures. The numbers in parenthesis are for rmsd done including only backbone atoms. [e] Number of clusters (from 100 MRC) obtained with a 0.6 Å vicinity. Legend: BS, binding site; HA, heavy atoms; and SL, self-ligand.

standard deviations above random) and similar to the correlation between two parts of the same MD trajectory. The collectivities[62] (i.e., the percentage of atoms displaced in each mode) of the first 100 modes is high (see Figure S3c in the Supporting Information), indicating that, at this level, the force that is opposing oscillations stems from a combined effect of sufficiently numerous interacting pairs, still being within the theoretical limit of the elastic network model.[28] According to that, and after several tests, we generalized one hundred eigenvectors as the definition of the important subspace for all the 28 proteins, enclosing the majority of the relevant movement (see dimensionality in Table 2). Other roomy definitions (e.g., vectors including 75% of variance) provide similar results for the entire dataset. However, although the excess of normal modes does not decrease the global performance, the lack of them (e.g., using only 10−25 modes) consistently reduces the number of correct ligand geometries scored first in 10%−15% of the benchmark.

1.4. Generation of the Cartesian Ensemble and Choice of the Number of Conformations. For docking purposes, the ensemble should be not only diverse but also relevant, so a

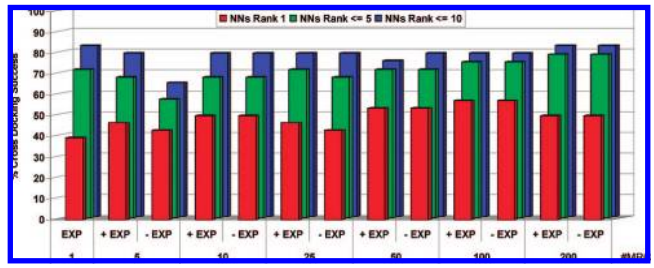

Figure 3. Success rate for cross-docking experiments using 1, 5, 10, 25, 50, and 200 protein conformations. A value of 100% indicates success in the 28 proteins. The colors show if the rank of the first near native solution (NNs) has the lowest energy score, is within the first 5 or within the first 10. Note that for MRC, the results are shown with the experimental protein included (+EXP) and excluded (−EXP).

Figure 4. Effect of the geometric clustering (side chains) on the global success in cross-docking experiments. A value of 100% indicates success in the 28 proteins. The colors indicate if the rank of the first near-native solution (NNs) has the lowest energy score, is within the first 5, or within the first 10. Note that the convergence is obtained with a vicinity of 0.6 Å.

balance between computational efficiency and accuracy is necessary. Our goal is to determine the average number of equilibrium structures that globally provides the best docking results (28 proteins). For that purpose, we generated 5, 10, 25, 50, 100, and 200 independent conformations for the binding site (see the Monte Carlo procedure that is described in the Methods section), which were tested in independent docking runs. As shown in Figure 3, when using a single receptor conformation (SRC), we found a NNs ranked in the top scoring position in 39% of the cases, in 71% of the cases within the first five scored poses, and in 82% within the first ten scored poses, which indicates that ICM is performing well, with regard to finding the NN geometries, but the associated scoring energies are not within the top ranks.

As can be observed in Figure 3, it is possible to increase the number of correct ligand geometries scored first only by using 5 or 10 MRCs. The results show also that the inclusion of the experimental structure among the MRC is important when the number of conformers is <25. The best results are obtained when 100 MRCs are used, having a good compromise between near native solutions ranked first, or within the first 5 or 10 lowest energy scores. Interestingly, when more conformations (200) are added, the increase of false positives reduces the performance, as seen previously with experimental MRC.[10]

1.5. Geometrical Clustering To Eliminate Protein Structural Redundancy. Although 100 protein conformations provide the best results, some structural redundancy may exist. To eliminate it, we performed a geometric clustering (rmsd-based) on the side-chains heavy atoms (at a distance of 5 Å from the cognate ligand), using thresholds in the range of 1−0.5 Å (see Figure 4). Our results suggest that a threshold of 0.6 Å captures the variability needed to obtain the same performance as the original ensemble but reduces the number of conformations by 40% on average.

At this point, we investigated whether or not it is possible to reduce the number of structures being applied in, for example, a virtual ligand screening experiment (VLS). Cheng et al.[18] recently reported good results using only the representative structures from the most populated clusters from a MD simulation. Based on our results, we have not found any clear relation between the populations of the NMA clusters and the scoring energy of the ligand poses. (See Figure S5 in the Supporting Information.) The data shows that, in multiple cases, the most populated conformations of our ensembles are not the ones that provide the best complementarities for the ligand. Thus, although the exclusion of less-populated conformations could improve the speed, it could also affect the quality of the results.

After analyzing our results globally, we can conclude that, with our protocol, we need, on average, 66 ± 24 conformations to improve the cross-docking results, in terms of geometry and scoring against a single receptor conformation. This implies using only ∼2% of the computational time needed for other refinement-based protocols, such as SCARE.[50] The inclusion of the experimental structure of the protein enhances the global ratio of success when using <25 conformations, being less noticeable as the conformational space increases.

2. Benchmark Results. The presented benchmark comprises protein conformations that correspond to 14 targets of interest in biomedical applications, with each of them displaying two different conformations when co-crystallized with different ligands (see Table 1).

For single receptor conformation docking (see Figure S4 and Table S1 in the Supporting Information), it is possible to successfully self-dock the cognate ligand in 27 of 28 cases. In 25 of 28 cases, the correct geometry is predicted as the best scored solution, and, in 27 of 28 cases, within the first two. We found only in one of the conformations (COX-2;

Table 3. Cross-Docking Results Obtained When Using 100 (EN·NMA-Based) Multiple Receptor Conformations (MRCs) against Single Receptor Conformation[a]

protein	PDB	rank	root mean square deviation, rmsd (Å)	energy (kcal/mol)[b]
CDK2	1AQ1	2	0.5	−25.0
		1	*0.7*	*−34.3*
	1DM2	1	0.5	−8.4
		1	*0.5*	*−18.1*
COX-2	1CX2	1	1.9	−23.0
		1	*1.8*	*−28.7*
	3PGH	2	1.9	−24.7
		5	*1.8*	*−22.7*
Estrogen receptor	1ERR	1	1.4	−32.3
		1	*1.4*	*−35.8*
	3ERT	6	1.7	−30.0
		2	*1.2*	*−41.4*
Factor Xa	1KSN	1	1.9	−44.1
		1	*1.4*	*−49.2*
	1XKA	6	1.9	−28.9
		1	*1.3*	*−60.0*
GSK-3 β	1Q4L	1	0.9	−26.7
		1	*0.6*	*−32.6*
	1UV5	LEP[c]	5.2	−31.9
		20	*1.6*	*−28.7*
HIV-1 RT	1C1C	1	1.2	−21.4
		1	*1.0*	*−23.2*
	1RTH	2	1.0	−23.7
		1	*0.9*	*−26.7*
JNK3	1PMN	1	0.5	−26.3
		1	*1.3*	*−34.4*
	1PMV	LEP[c]	9.0	−6.9
		LEP[c]	*7.6*	*−15.1*
LXR β LBD	1P8D	LEP[c]	4.0	5.8
		LEP[c]	*4.5*	*−3.4*
	1PQ6	2	1.0	4.1
		2	*0.9*	*7.2*
Neuraminidase	1A4Q	1	1.0	−33.1
		1	*0.5*	*−30.8*
	1NSC	2	1.6	−20.9
		1	*1.4*	*−25.1*
P38 MAP Kinase	1BMK	2	1.3	−19.0
		2	*1.1*	*−21.3*
	1DI9	1	1.6	−29.2
		1	*1.8*	*−37.8*
PKA	1STC	2	1.4	−23.4
		1	*1.5*	*−30.1*
	1YDS	9	0.7	41.1
		51	*0.9*	*18.1*
PPARγ	1FM9	1	1.5	−30.1
		1	*1.9*	*−34.7*
	2PRG	LEP[c]	9.7	−15.1
		LEP[c]	*10.4*	*−34.9*
TK	1KI4	3	1.4	−20.9
		8	*1.4*	*−20.7*
	1KIM	4	1.2	−5.3
		11	*1.2*	*−14.5*
Trypsin	1PPC	1	1.1	−29.1
		1	*0.9*	*−34.3*
	1PPH	LEP[c]	6.2	−35.4
		3	*1.6*	*−37.2*

[a] The first line in each row corresponds to SRC and the second (with italic font) corresponds to MRC. The experimental protein structure was not included among the MRC. [b] According to ICM scoring function. [c] Lowest energy pose.

1CX2 with ligand SC-588) that the geometry of the ligand had an rmsd of >2 Å, relative to the experimental position, but this still had the correct orientation and was predicted to be the lowest-energy solution. For cross-docking (see Figure 3 and Table 3), 11 of 28 cases have a NNs as the lowest-energy position, 20 of 28 cases within the first five, and 23 of 28 cases within the first ten.

When sufficient MRCs are used in cross docking, there is a global increase in the number of NNs that are ranked as the lowest-energy solution. Also, the associated binding scores of correctly predicted geometries are generally stronger (see Table 3, e.g., 1XKA). Interestingly, a small improvement in the energies appears in some self-docking solutions (see Table S1 in the Supporting Information). Currently, we are analyzing if this improvement in the energies leads to better enrichment factors in Virtual Library Screening experiments.

There are still a couple of cases (1PQ6, 1YD6) where the binding energies are positive. With 1PQ6 it is still possible to capture an NNs within the first two lowest-energy solutions, but still some steric clashes exist, making this conformation not optimum for cross docking. 1YD6 neither provides good solutions for SRC nor MRC within the first positions, because of high steric clashes with the ligand. Moreover, the increase in the number of ligand poses that are used in the MRCs is reflected in the rank of the first NNs (MRC rank = 51; SRC rank = 9). However, the energy of the first NNs found with MRC is still 20 kcal/mol better than that for SRC.

Of special interest are the three proteins JNK3, LXR β LBD, and PPARγ (note that the first two are not included in Sherman's original dataset), having strong induced fit in one of the conformations. Thus, although 1PMN, 1PQ6, and 1FM9 self-dock and cross-dock easily, 1PMV, 1P8D, and 2PRG do not cross-dock either in SRC or MRC. An inspection of the structures showed that they have clear clashing residues facing the binding site. Also, the co-crystallized ligands are smaller in size than those in their partner, and this is reflected in the number of heavy atoms selected for the EN·NMA (see Table 2). With the only exception of crystal structure 1P8D, the difference in size persisted after the ligand was removed and the binding pocket boundaries were predicted by the Pocketome Gaussian Convolution algorithm.[63] Interestingly, despite the smaller number of atoms used in the EN·NMA, and less dimensionality, the problematic proteins needed more clusters than their partners (see Table 2). A deeper investigation is being conducted into the importance of flexibility descriptors in the selection of more-suitable structures for VLS.

The proposed method is intended to be a fast and simple alternative to MD simulations. Nevertheless, even MD structures from holo-form simulations may not perform satisfactorily in cross-docking experiments.[9] The results from the more-challenging pairs are consistent with MD simulation[59] of the porcine holo-Aldose Reductase (1AH0), where conformations generated including the self-ligand Sorbinil could never accommodate the ligand Tolrestat, whose native binding mode require a peculiar protein rearrangement that could not be sampled during the simulation (co-crystal 1AH3). For such a protein, neither MD conformations (coming from a multiminima conformational energy surface) nor EN·NMA conformations (coming from a harmonic potential surface determined at a single energy minimum) were able to represent the fluctuations needed for the correct placement of the cross ligand. When the positions of the side chains are strongly dependent on the foreign ligand (e.g., PPARγ: 2PRG), other methodologies (including refinement) provide better ligand geometries, such as Sherman's original dataset[52] or SCARE,[50] but with an associated important increase in computational time.

In contrast to other protocols with similar objectives, our approach allows the movement of both backbone and side chains. Using all heavy-atom EN·NMA, we achieved a performance very similar to that of the Molecular Mechanics-based method that was proposed very recently by Koska et al.[64] The full potential of the method is now being tested in proteins where wider backbone displacements are needed (such as, for example, those with some protein kinases or G-protein-coupled receptors). A variant of the method that could be easily implemented is the use of the collective motions from the combination of different ensembles, such as experimental conformations (X-ray or NMR) or MD. Concerted motions from such ensembles can be determined using the PCA method,[27,65] and the important subspace can be used for the generation of displacements.

Although we used ICM for the docking procedure, the proposed MRC methodology is intended to be program-independent. For that reason, we have developed a web server where the user can obtain an ensemble of protein conformers in various formats, including PDB, which can be used sequentially with other docking software.

MATERIALS AND METHODS

1. Benchmark. The proposed benchmark was already used to test the performance of the SCARE cross-docking methodology.[50] The set of complexes is mainly based on the dataset of highly challenging test cases compiled by Sherman and colleagues,[52] which was further expanded to include several structures from other cross-docking exercises recently reported in the literature.[66−70] No more than two structures of the same protein (identical sequence) were collected; if possible, holo structures were preferred over apo structures and complexes with small organic molecules over complexes with endogenous binders, peptides, or peptidomimetic drugs. Metalloproteins were excluded from the selection, because of the unfeasibility of correctly assessing the receptor flexibility without manually excluding the side chains that coordinate the metal ion(s). Finally, structures where the conformational differences could be ascribed to the poor quality of the crystals, rather than to a genuine induced-fit effect, were also excluded.

The benchmark encompassed only structure pairs where a limited number of side chains and backbone regions displayed significant rearrangements. No structure pairs where only subtle (if any) differences where detectable or, conversely, where large scale movements or domain motions took place were included. Twenty-eight structures (2 conformations for each of 14 different proteins) were selected (see Table 1). The set of structures can be considered diverse and consistently challenging for a cross-docking protocol, being representative of several important targets of interest in medicinal chemistry. The atomic coordinates were retrieved from the Protein Data Bank (PDB).[71]

1.1. Single Receptor Conformation Setup. Chains that did not define the binding site were deleted. When multiple copies of the biological entity were included in the asymmetric unit, the most complete one was chosen. Water molecules were deliberately removed and only HET groups within the EN·NMA boundaries were kept. After the assignment of the correct atom types, disulfide bonds, hydrogens, and missing heavy atoms were added to the structure. For the side chains whose occupancy was <1, the best energy conformation was selected. If available, histidines were assigned the tautomeric state reported; otherwise, the energetically most favorable state was assigned. Positions of polar hydrogen atoms, asparagines, and glutamines adjacent to the binding site were optimized to maximize hydrogen bonding with the remainder of the protein and the cognate ligand. Finally, the ligand was deleted from the holo structures and the binding pocket regions of the two conformations of each protein were superimposed[72] using backbone atoms.

1.2. Ligand Setup. Coordinates of the ligands were also taken from the crystallographic complexes. Bond order, tautomeric forms, stereochemistry, hydrogen atoms, and the protonation state were assigned based on the primary literature description. Each ligand was assigned the MMFF[73] force field atom types and charges. Ligand molecules were prepared for docking by energy minimization in the absence of the receptor, and the lowest-energy conformations were used as starting points for docking simulations.

1.3. Multiple Receptor Conformation Setup. The heavy atoms from residues within a distance of 10 Å from the cognate ligand were selected as initial coordinates for the NMA.

2. Normal Mode Analysis. In this methodology, it is assumed that the displacement of an atom from its equilibrium position is small and that the potential energy in the vicinity of the equilibrium position can be approximated as a sum of terms that are quadratic in the atomic displacements.[29] In its purest form, it uses the same all-atom force field from a MD simulation, implying a prior minimization in vacuo.[37] Tirion[74] proposed a simplified model where the interaction between two atoms was described by Hookean pairwise potential where the distances are taken to be at the energy minimum, avoiding the minimization

$$E(r_i, r_j) = K_{ij}(r_{ij} - r_{ij}^0)^2 \quad (1)$$

where r_i denotes the coordinates of atom i, and $r_{ij} = ri - rj$. The superscript zero indicates the coordinate at the original conformation. The strength of the potential K_{ij} is assumed to be the same for all interacting pairs, and the total potential is

$$E = \sum_{i,j|r_{i,j}|\leq R}^{N,N} E(r_i, r_j) \quad (2)$$

where R is the cut-off parameter, so that the interactions are limited to pairs separated by R (8−12 Å).

This simplified potential function commonly is referred as elastic network (EN·NMA), because the macromolecule is represented as a network, which includes all interactions within a distance. In this paper, we used a continuum function for the spring constant K_{ij} that assumes an inverse exponential relationship between the distance and the force constant and avoids the use of the cutoff.[57] The constants were quantitatively adjusted to describe fluctuations in the range of a 10-ns MD trajectory:[59]

$$K_{ij} = \begin{cases} 350\ \text{kcal/(mol Å}^2) & (\text{for } j = i+1) \\ 0.25\ \text{kcal/(mol Å}^2) \times (d_{ij}^0/d_{ij})^6 & (\text{for } j \neq i+1) \end{cases} \quad (3)$$

where d^0 is a fitted constant, taken as the mean Cα−Cα distance between consecutive residues. The masses of the

heavy atoms where set to 1. The force constant matrix of the system is described by the Hessian matrix ($3N \times 3N$ matrix, where N the number of heavy atoms), obtained as the partial second derivatives of the potential with respect to the coordinates. Complete diagonalization of the Hessian yields $3N - 6$ eigenvectors, ranked according to their corresponding eigenvalues.

The EN·NMA space is used for the generation of Cartesian displacements. Here, we have chosen a Metropolis Monte Carlo algorithm with a Hamiltonian that reproduced the original sampling fluctuations in MD trajectories[36] (see eq 4):

$$E = \sum_{i=1}^{n} \frac{1}{2} k_i \Delta D_i^2 \quad (4)$$

where n is the number of modes ($n = 100$) and ΔD_i represents the displacement along a given mode i. The parameter k_i is the stiffness constant associated with harmonic deformation on mode i (i.e., the eigenvalue). In each iteration, 1 of the 100 modes is displaced randomly, following a Metropolis criteria. This generates allowed displacements (at a temperature of 300 K) on each of the modes that can be projected onto the Cartesian space. It is possible to obtain bigger displacements by increasing the temperature of the Monte Carlo procedure.

The ICM receptor topology of the SRC (see above) was tethered to the coordinates of each MRC. The chemical distortions produced by displacements were corrected with 25 steps of Cartesian minimization, making the inclusion of the hydrogen atoms in the EN·NMA calculation unnecessary.

ICM Docking and Scoring. ICM addresses the docking issue as a global optimization problem, implementing a biased probability Monte Carlo (BPMC) global stochastic optimizer.[48] Because the BPMC method has been previously reported and thoroughly described, it is summarized here only briefly. During docking, one of the ligand torsional or roto-translational variables is randomly changed. A local refinement is performed on the analytically differentiable terms by a conjugate-gradient minimization. The complete energy is calculated by summing the contributions of the solvation energy and those of the conformational entropy. The new conformation is accepted or rejected according to the Metropolis criteria. A new random change is introduced and the entire procedure is repeated all over until a previously set number of steps are performed.

The molecular system was described using internal coordinate variables. Protein atomic types and parameters were taken from a modified version of the ECEPP/3 force field.[75] The bounding box was derived from the residues at a distance of 5 Å from the self-ligand. The binding pocket was described as a combination of five potentials that accounted for two van der Waals boundaries, electrostatics, hydrophobicity, and hydrogen bonding. Each potential was expressed on a precalculated grid spacing of 0.5 Å. Because the standard 6−12 kcal/mol van der Waals potential was considered too sensitive to steric clashes, for the purpose of the simulations, a truncated soft version was introduced and the other potentials were rescaled accordingly, where the van der Waals potentials were set to 1.0 kcal/mol. The ligand was docked into the grid representation of the pocket using the BPMC method described previously.

The global optimization provided a stack of geometrically diverse ligand poses, and the best 10 were assigned a docking score by the standard ICM empirical scoring function.[76,77] In MRC docking, an independent run was performed for each receptor conformation and the best five poses of each run were scored and combined in a stack consisting of $5 \times M$ (where M is the number of MRC) poses. For selected cases, we checked that the different number of scored poses between SRC and MRC was not representative of their performance (i.e., combining $5 \times M'$ poses, where M' is the number of SRC docking experiments). In all cases, the redundancy was eliminated from the stack and the poses were sorted according to their ICM binding score energies. An NNs was considered when the ligand heavy atom rmsd, relative to the crystallographic coordinates, was ≤2.0 Å.

The receptors and ligands preparation, the docking simulations and the energy evaluations were all carried out by ICM 3.6 (Molsoft L.L.C., La Jolla, CA) coupled to in-house programs for the MRC generation.

Calculation Time. The MRC method, including NMA, generation of a Cartesian ensemble of 100 conformations, tethering, minimization, and geometrical clustering required ∼5 min per protein (see Figure 1). Each single rigid receptor docking run required ∼30−90 s, depending on the ligand size. The approximate total time per ligand is 1 h on a single computer processing unit (CPU). Because rigid dockings are independent, the calculation could be easily split between different CPUs.

The docking simulations ran on an Intel Intel Core 2 Quad workstation (2.4 GHz with 3 GB of RAM memory). The MRC methodology is implemented in a web server, where the user can either choose a PDB code or upload a file that contains a PDB structure and obtain an ensemble of conformations (http://abagyan.scripps.edu/MRC).

CONCLUSIONS

The inclusion of the flexibility of the protein is, in many cases, crucial to accurately predict the orientation and interactions of ligands within the binding pocket. One of the big challenges is to routinely incorporate flexibility considerations into structure-based drug discovery in an affordable computing time. Here, we have shown that using <100 conformations it is possible to improve the performance of top scoring cross docking against a single experimental conformation. The multiple conformation ensembles were derived from the essential motions of the residues surrounding the binding site, including heavy atoms from backbone and side chains. The needed fluctuations of pocket residues seem local and rather independent of the global collective motions of the protein. The results presented here could be useful to the docking-simulation community, and improvements are expected from the use of the collective motions extracted from mixed ensembles from MD, X-ray or NMR structures. The use of the essential motions of the binding site could enhance the conformational space needed for docking.

Note Added in Proof. A manuscript from Mischino et al. "Community-wide blind assessment of methods for GPCR structure modeling and docking" has been submitted to *Nature Reviews Drug Discovery*. The results suggested

significant improvements in accuracy of the adenosine A2a receptor modelling when multiple A2a pocket conformations were generated using our NMA methodology.

ACKNOWLEDGMENT

The authors thank Alberto Perez, Carles Ferrer-Costa, Michael S. Fish, and George Nicola for reviewing the manuscript. M.R. is supported by a Spanish MEC Postdoctoral fellowship. This work was funded in part by NIH grants 5-R01-GM071872 and 1-R01-GM074832.

Supporting Information Available: Figures S1−S5 and Table S1, as described in the text. (PDF) This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES AND NOTES

(1) Totrov, M.; Abagyan, R. Flexible ligand docking to multiple receptor conformations: A practical alternative. *Curr. Opin. Struct. Biol.* **2008**, *18*, 178–184.
(2) Henzler-Wildman, K. A.; Lei, M.; Thai, V.; Kerns, S. J.; Karplus, M.; Kern, D. A hierarchy of timescales in protein dynamics is linked to enzyme catalysis. *Nature* **2007**, *450*, 913–916.
(3) Ishima, R.; Torchia, D. A. Protein dynamics from NMR. *Nat. Struct. Biol.* **2000**, *7*, 740–743.
(4) Lange, O. F.; Lakomek, N. A.; Fares, C.; Schroder, G. F.; Walter, K. F.; Becker, S.; Meiler, J.; Grubmuller, H.; Griesinger, C.; de Groot, B. L. Recognition dynamics up to microseconds revealed from an RDC-derived ubiquitin ensemble in solution. *Science* **2008**, *320*, 1471–1475.
(5) Teague, S. J. Implications of protein flexibility for drug discovery. *Nat. Rev. Drug Discovery* **2003**, *2*, 527–541.
(6) Sousa, S. F.; Fernandes, P. A.; Ramos, M. J. Protein−ligand docking: Current status and future challenges. *Proteins−Struct. Funct. Bioinf.* **2006**, *65*, 15–26.
(7) Carlson, H. A.; McCammon, J. A. Accommodating protein flexibility in computational drug design. *Mol. Pharmacol.* **2000**, *57*, 213–218.
(8) Teodoro, M. L.; Kavraki, L. E. Conformational flexibility models for the receptor in structure based drug design. *Curr. Pharm. Des.* **2003**, *9*, 1635–1648.
(9) Cozzini, P.; Kellogg, G. E.; Spyrakis, F.; Abraham, D. J.; Costantino, G.; Emerson, A.; Fanelli, F.; Gohlke, H.; Kuhn, L. A.; Morris, G. M.; Orozco, M.; Pertinhez, T. A.; Rizzi, M.; Sotriffer, C. A. Target flexibility: an emerging consideration in drug discovery and design. *J. Med. Chem.* **2008**, *51*, 6237–6255.
(10) Barril, X.; Morley, S. Unveiling the full potential of flexible receptor docking using multiple crystallographic structures. *J. Med. Chem.* **2005**, *48*, 4432–4443.
(11) Sheridan, R. P.; McGaughey, G. B.; Cornell, W. D. Multiple protein structures and multiple ligands: effects on the apparent goodness of virtual screening results. *J. Comput.-Aided. Mol. Des.* **2008**, *22*, 257–265.
(12) Rockey, W. M.; Elcock, A. H. Structure selection for protein kinase docking and virtual screening: homology models or crystal structures. *Curr. Protein Pept. Sci.* **2006**, *7*, 437–457.
(13) Thomas, M. P.; McInnes, C.; Fischer, P. M. Protein structures in virtual screening: a case study with CDK2. *J. Med. Chem.* **2006**, *49*, 92–104.
(14) Huang, S. Y.; Zou, X. Ensemble docking of multiple protein structures: considering protein structural variations in molecular docking. *Proteins* **2007**, *66*, 399–421.
(15) Damm, K. L.; Carlson, H. A. Exploring experimental sources of multiple protein conformations in structure-based drug design. *J. Am. Chem. Soc.* **2007**, *129*, 8225–8235.
(16) Karplus, M.; Kuriyan, J. Molecular dynamics and protein function. *Proc. Natl. Acad. Sci., U.S.A.* **2005**, *102*, 6679–6685.
(17) Karplus, M. Molecular dynamics of biological macromolecules: A brief history and perspective. *Biopolymers* **2003**, *68*, 350–358.
(18) Cheng, L. S.; Amaro, R. E.; Xu, D.; Li, W. W.; Arzberger, P. W.; McCammon, J. A. Ensemble-based virtual screening reveals potential novel antiviral compounds for avian influenza neuraminidase. *J. Med. Chem.* **2008**, *51*, 3878–3894.
(19) Wang, Y.; Tajkhorshid, E. Electrostatic funneling of substrate in mitochondrial inner membrane carriers. *Proc. Natl. Acad. Sci., U.S.A.* **2008**, *105*, 9598–9603.
(20) Zacharias, M. Rapid protein-ligand docking using soft modes from molecular dynamics simulations to account for protein deformability: binding of FK506 to FKBP. *Proteins* **2004**, *54*, 759–767.
(21) Xu, Y.; Colletier, J. P.; Jiang, H.; Silman, I.; Sussman, J. L.; Weik, M. Induced-fit or preexisting equilibrium dynamics? Lessons from protein crystallography and MD simulations on acetylcholinesterase and implications for structure-based drug design. *Protein Sci.* **2008**, *17*, 601–605.
(22) Marco, E.; Gago, F. Overcoming the inadequacies or limitations of experimental structures as drug targets by using computational modeling tools and molecular dynamics simulations. *Chem. Med. Chem.* **2007**, *2*, 1338–1401.
(23) Sotriffer, C. A.; Kramer, O.; Klebe, G. Probing flexibility and "induced-fit" phenomena in aldose reductase by comparative crystal structure analysis and molecular dynamics simulations. *Proteins* **2004**, *56*, 52–66.
(24) Meagher, K. L.; Carlson, H. A. Incorporating protein flexibility in structure-based drug discovery: Using HIV-1 protease as a test case. *J. Am. Chem. Soc.* **2004**, *126*, 13276–13281.
(25) Soliva, R.; Gelpi, J. L.; Almansa, C.; Virgili, M.; Orozco, M. Dissection of the recognition properties of p38 MAP kinase. Determination of the binding mode of a new pyridinyl-heterocycle inhibitor family. *J. Med. Chem.* **2007**, *50*, 283–293.
(26) Amaro, R. E.; Baron, R.; McCammon, J. A. An improved relaxed complex scheme for receptor flexibility in computer-aided drug design. *J. Comput.-Aided. Mol. Des.* **2008**, *22*, 693–705.
(27) Amadei, A.; Linssen, A. B.; Berendsen, H. J. Essential dynamics of proteins. *Proteins* **1993**, *17*, 412–425.
(28) Tirion, M. M. Large Amplitude Elastic Motions in Proteins from a Single-Parameter, Atomic Analysis. *Phys. Rev. Lett.* **1996**, *77*, 1905–1908.
(29) Cui, Q.; Bahar, I. *Normal Mode Analysis: Theory and Applications to Biological and Chemical Systems*; CRC Press: Boca Raton, FL, 2006; p 406.
(30) Go, N.; Scheraga, H. A. On the Use of Classical Statistical Mechanics in the Treatment of Polymer Chain Conformation. *Macromolecules* **1976**, *9*, 535–542.
(31) Levy, R. M.; Karplus, M. Vibrational approach to the dynamics of an α-helix. *Biopolymers* **1979**, *18*, 2465–2495.
(32) Go, N.; Noguti, T.; Nishikawa, T. Dynamics of a small globular protein in terms of low-frequency vibrational modes. *Proc. Natl. Acad. Sci., U.S.A.* **1983**, *80*, 3696–3700.
(33) Brooks, B.; Karplus, M. Harmonic dynamics of proteins: normal modes and fluctuations in bovine pancreatic trypsin inhibitor. *Proc. Natl. Acad. Sci., U.S.A.* **1983**, *80*, 6571–6575.
(34) Hayward, S.; Kitao, A.; Go, N. Harmonic and Anharmonic Aspects in the Dynamics of BPTI−A Normal-Mode Analysis and Principal Component Analysis. *Protein Sci.* **1994**, *3*, 936–943.
(35) Hayward, S.; Kitao, A.; Go, N. Harmonicity and Anharmonicity in Protein Dynamics−A Normal-Mode Analysis and Principal Component Analysis. *Proteins−Struct. Funct. Genet.* **1995**, *23*, 177–186.
(36) Rueda, M.; Chacon, P.; Orozco, M. Thorough Validation of Protein Normal Mode Analysis: A Comparative Study with Essential Dynamics. *Structure* **2007**, *15*, 565–575.
(37) Hayward, S.; de Groot, B. L. Normal modes and essential dynamics. *Methods Mol. Biol.* **2008**, *443*, 89–106.
(38) 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. Comput. Chem.* **1999**, *20*, 287–300.
(39) 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.
(40) Sandera, T.; Liljeforsa, 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.
(41) Cavassotto, C.; Kovacs, J.; Abagyan, R. Representing receptor flexibility in ligand docking through relevant normal modes. *J. Am. Chem. Soc.* **2005**, *127*, 9632–9640.
(42) 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.
(43) 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.
(44) Bartlett, G. J.; Porter, C. T.; Borkakoti, N.; Thornton, J. M. Analysis of catalytic residues in enzyme active sites. *J. Mol. Biol.* **2002**, *324*, 105–121.
(45) Yuan, Z.; Zhao, J.; Wang, Z. X. Flexibility analysis of enzyme active sites by crystallographic temperature factors. *Protein Eng.* **2003**, *16*, 109–114.
(46) Sacquin-Mora, S.; Laforet, E.; Lavery, R. Locating the active sites of enzymes using mechanical properties. *Proteins* **2007**, *67*, 350–359.

(47) Zhou, Y.; Vitkup, D.; Karplus, M. Native proteins are surface-molten solids: application of the Lindemann criterion for the solid versus liquid state. *J. Mol. Biol.* **1999**, *285*, 1371–1375.
(48) Abagyan, R.; Totrov, M. Biased probability Monte Carlo conformational searches and electrostatic calculations for peptides and proteins. *J. Mol. Biol.* **1994**, *235*, 983–1002.
(49) Totrov, M.; Abagyan, R. Detailed ab initio prediction of lysozyme-antibody complex with 1.6 A accuracy. *Nat. Struct. Biol.* **1994**, *1*, 259–263.
(50) Bottegoni, G.; Kufareva, I.; Totrov, M.; Abagyan, R. A new method for ligand docking to flexible receptors by dual alanine scanning and refinement (SCARE). *J. Comput.-Aided. Mol. Des.* **2008**, *22*, 311–325.
(51) Brem, R.; Dill, K. A. The effect of multiple binding modes on empirical modeling of ligand docking to proteins. *Protein Sci.* **1999**, *8*, 1134–1143.
(52) Sherman, W.; Day, T.; Jacobson, M. P.; Friesner, R. A.; Farid, R. Novel procedure for modeling ligand/receptor induced fit effects. *J. Med. Chem.* **2006**, *49*, 534–553.
(53) Gervasio, F. L.; Laio, A.; Parrinello, M. Flexible docking in solution using metadynamics. *J. Am. Chem. Soc.* **2005**, *127*, 2600–2607.
(54) Kurkcuoglua, O.; Jernigan, R. L.; Dorukera, P. Collective Dynamics of Large Proteins from Mixed Coarse-Grained Elastic Network Model. *QSAR Comb. Sci.* **2004**, *24*, 443–448.
(55) Daniel, R. M.; Dumm, R. V.; Finney, J. L.; Smith, J. C. The role of dynamics in enzyme activity. *Annu. Rev. Biophys. Biomol. Struct.* **2003**, *32*, 69–92.
(56) Reat, V.; Patzelt, H.; Ferrand, M.; Pfister, C.; Oesterhelt, D.; Zaccai, G. Dynamics of different functional parts of bacteriorhodopsin: H-2H labeling and neutron scattering. *Proc. Natl. Acad. Sci., U.S.A.* **1998**, *95*, 4970–4975.
(57) Kovacs, J. A.; Chacon, P.; Abagyan, R. Predictions of protein flexibility: first-order measures. *Proteins* **2004**, *56*, 661–668.
(58) Tirion, M. M.; Ben-Avraham, D. Normal mode analysis of G-Actin. *J. Mol. Biol.* **1993**, *230*, 186–195.
(59) Rueda, M.; Ferrer-Costa, C.; Meyer, T.; Perez, A.; Camps, J.; Hospital, A.; Gelpi, J. L.; Orozco, M. A consensus view of protein dynamics. *Proc. Natl. Acad. Sci., U.S.A.* **2007**, *104*, 796–801.
(60) Gutteridge, A.; Thornton, J. Conformational changes observed in enzyme crystal structures upon substrate binding. *J. Mol. Biol.* **2005**, *346*, 21–28.
(61) Perez, A.; Blas, J. R.; Rueda, M.; López-Bes, J. M.; de La Cruz, X.; Luque, F. J.; Orozco, M. Exploring the essential dynamics of B.DNA. *J. Chem. Theor. Comput.* **2005**, *1*, 790–800.
(62) Brüschweiler, R. Collective protein dynamics and nuclear spin relaxation. *J. Chem. Phys.* **1995**, *102*, 3396–3403.
(63) An, J.; Totrov, M.; Abagyan, R. Pocketome via comprehensive identification and classification of ligand binding envelopes. *Mol. Cell Protein* **2005**, *4*, 752–761.
(64) Koska, J.; Spassov, V. Z.; Maynard, A. J.; Yan, L.; Austin, N.; Flook, P. K.; Venkatachalam, C. M. Fully Automated Molecular Mechanics Based Induced Fit Protein−Ligand Docking Method. *J. Chem. Inf. Model.* **2008**, *48*, 1965–1973.
(65) Mustard, D.; Ritchie, D. W. Docking essential dynamics eigenstructures. *Proteins* **2005**, *60*, 269–274.
(66) Cavasotto, C. N.; Abagyan, R. A. Protein flexibility in ligand docking and virtual screening to protein kinases. *J. Mol. Biol.* **2004**, *337*, 209–225.
(67) Meiler, J.; Baker, D. ROSETTALIGAND: protein−small molecule docking with full side-chain flexibility. *Proteins* **2006**, *65*, 538–548.
(68) Polgar, T.; Baki, A.; Szendrei, G. I.; Keseru, G. M. Comparative virtual and experimental high-throughput screening for glycogen synthase kinase-3 beta inhibitors. *J. Med. Chem.* **2005**, *48*, 7946–7959.
(69) Polgar, T.; Keseru, G. M. Ensemble docking into flexible active sites. Critical evaluation of FlexE against JNK-3 and beta-secretase. *J. Chem. Inf. Model.* **2006**, *46*, 1795–1805.
(70) Sheng-You Huang, X. Z. Ensemble docking of multiple protein structures: Considering protein structural variations in molecular docking. *Proteins* **2007**, *66*, 399–421.
(71) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucl. Acid. Res.* **2000**, *28*, 235–242.
(72) Damm, K. L.; Carlson, H. A. Gaussian-weighted RMSD superposition of proteins: a structural comparison for flexible proteins and predicted protein structures. *Biophys. J.* **2006**, *90*, 4558–4573.
(73) Halgren, T. A. Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94. *J. Comput. Chem.* **1996**, *17*, 490–519.
(74) Tirion, M. M.; Ben-Avraham, D.; Lorenz, M.; Holmes, K. C. Normal modes as refinement parameters for the F-Actin model. *Biophys. J.* **1995**, *68*, 5–12.
(75) Nemethy, G.; Gibson, K. D.; Palmer, K. A.; Yoon, C. N.; Paterlini, G.; Zagari, A.; Rumsey, S.; Scheraga, H. A. Energy parameters in polypeptides. 10. Improved geometrical parameters and nonbonded interactions for use in the ECEPP(SLASH)3 algorithm, with application to proline-containing peptides. *J. Chem. Phys.* **1992**, *96*, 6472.
(76) Totrov, M.; Abagyan, R. In *Derivation of sensitive discrimination potential for virtual ligand screening*; Istrail, S., Pevzner, P., Waterman, M., Eds.;RECOMB'99: Proceedings of the Third Annual International Conference on Computational Molecular Biology, Lyon, France, 1999; Association for Computer Machinery: New York, 1999; pp 312−320.
(77) Totrov, M.; Abagyan, R., Protein−ligand docking as an energy optimization problem. In *Drug-Receptor Thermodynamics: Introduction and Experimental Applications*, Raffa R. B., Ed.; Wiley: New York, 2001; p 603.

CI8003732