

Interdomain Conformations in the Full-Length MMP-2 Enzyme Explored by Protein–Protein Docking Calculations Using pyDock

Haydee Valdés,[†] Natalia Díaz,[†] Dimas Suárez,^{*,†} and Juan Fernández-Recio[‡]

Dpto. Química Física y Analítica, Universidad de Oviedo, C/Julián Clavería, 8,
33006, Oviedo (Asturias), Spain and Barcelona Supercomputing Center,
C/Jordi Girona 29, E-08034 Barcelona, Spain

Received February 18, 2010

Abstract: Current understanding of the collagenolytic activity performed by the matrix metalloproteinases (MMPs) assumes some degree of relative motion between their catalytic and hemopexin-like domains, according to evidence from low-resolution techniques for some of the MMP family members. Herein, we employ protein–protein docking calculations to investigate the structure in aqueous solution of the full-length MMP-2 enzyme in its active form, for which there is not yet experimental evidence of interdomain movement. After docking the domains as free rigid-body subunits, the linker region connecting the catalytic and hemopexin-like domains is taken into account *a posteriori* by merely adding an empiric energy term computed from expected end-to-end distance to the scoring function. Finally, full-length MMP-2 structures are generated by model building the linker residues in the most stable docking poses. The results add support to the hypothesis that the interdomain dynamics of a single MMP-2 molecule in aqueous solution can result in a manifold of conformations, with some preferred orientations. Globally, this structural information could be helpful in future experimental or computational studies aimed to elucidate the dynamical behavior of the MMP-2 enzyme in solution.

Introduction

Matrix metalloproteinases (MMPs) are an important family of zinc- and calcium-dependent peptidases involved in the proteolytic processing of the pericellular environment. The MMPs can cleave virtually all structural matrix proteins (collagen, aggrecan, laminin, etc.), but they also process adhesion molecules (integrins) and biologically active molecules like growth factors, cytokines, and growth factor receptors, contributing thus to the regulation of cellular behavior.^{1,2} Accordingly, they play a central role in different physiological processes, and their expression is also known to increase in various inflammatory, malignant, and degenerative diseases.^{3,4}

All of the MMPs share a significant sequence homology and, in most cases, a common multidomain structure formed by an N-terminal prodomain, a catalytic domain, and a C-terminal hemopexin-like domain joined to the catalytic domain through a linker region (**LK**) of variable length (14–68 residues).^{5,6} The N-terminal pro-peptide blocks the access to the active site cleft in the catalytic domain and is removed upon activation. The catalytic domain (**CAT**, about 170 residues), which holds the proteolytic activity, displays a twisted five-stranded β sheet, three α helices, and several bridging loops. In the gelatinases (MMP-2 and MMP-9), this domain has an additional 175 amino acid residue insert comprising three fibronectin-related type II modules (**FIB1–3**) conferring gelatin and collagen binding properties (see Figure S1 in the Supporting Information). The C-terminal hemopexin-like (**HPX**) domain, which is important in regulating the MMP activation, localization, and inhibition, presents a four-bladed propeller structure. The MMPs also contain a catalytic zinc ion (Zn1), a second zinc ion (Zn2),

* Author to whom correspondence should be addressed. Fax: 34-985-103125. E-mail: dimas@uniovi.es.

[†] Universidad de Oviedo.

[‡] Barcelona Supercomputing Center.

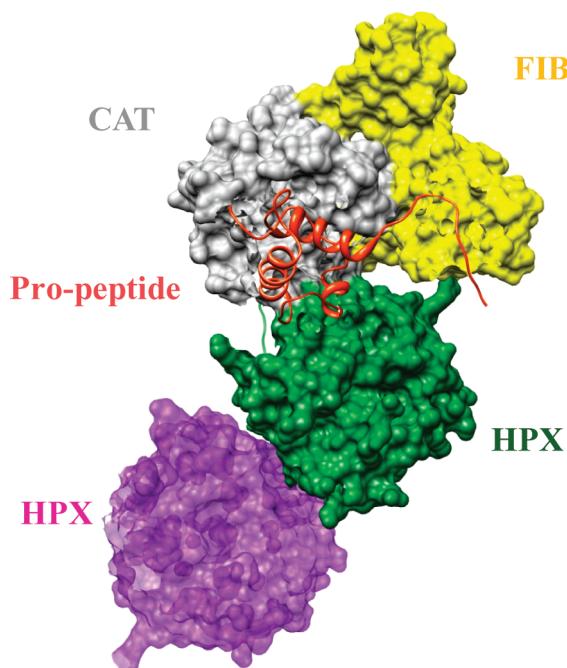


Figure 1. Intermolecular interactions between **HPX** domains of two different molecules in the 1CK7 structure. The second molecule (purple) is only partially represented by the **HPX** domain.

and a number of calcium ions that, according to previous studies, play a structural role in stabilizing several loop regions and fine-tuning the access to the binding site cleft.⁷

To date, X-ray structures of the full-length enzymes have been only reported for MMP-1,^{8,9} MMP-2,^{10,11} and MMP-12.¹² The structures of MMP-1 and MMP-2 display a similar compact arrangement of the catalytic and hemopexin domains, which may suggest the presence of stable interdomain contacts. In contrast, the solid state structure of MMP-12 displays the two domains in a different orientation characterized by a less compact arrangement and a smaller interdomain contact area. Interestingly, some degree of relative motion between the **CAT** and the **HPX** domains of the MMPs has been invoked to explain the collagenolytic activity performed by these enzymes.¹³ This hypothesis has been experimentally confirmed for MMP-1, MMP-9, and MMP-12. Thus, small-angle X-ray scattering and atomic force microscopy experiments have shown that significant interdomain motions occur for MMP-9 in solution.¹⁴ In addition, nuclear magnetic resonance measurements performed for the MMP-1 and MMP-12 enzymes have also revealed that the **CAT** and **HPX** domains experience conformational freedom with respect to each other on the nanosecond time scale.^{12,15}

For MMP-2, an important enzyme in angiogenesis, there is no experimental evidence of interdomain movement yet. The tridimensional structures currently available correspond to the latent pro-enzyme (PDB codes: 1CK7 and 1GXD), where a number of contacts have been observed between the different domains.¹⁰ Thus, the propeptide contacts the **CAT** and **FIB3** domains while remaining close to the first propeller blade of **HPX** (see Figure 1). A larger patch of molecular surface, which amounts to 310 Å² in terms of the reduction of the solvent excluded molecular surface of the

two domains, is covered by the packing of 12 linker residues against the Ω loop of the catalytic domain, resulting in several interdomain H bonds (e.g., Gln₄₃₅—NH···Pro₄₆₃—O, Asp₄₁₆—Oδ···Thr₄₆₅—OγH) and hydrophobic interactions (e.g., Pro₄₁₇···Thr₄₆₅). Finally, the smallest interdomain contact region is formed between the second blade of the **HPX** domain and the first of the three existing **FIB** domains. The only interaction existing in this particular region is a solvent-exposed salt-bridge (Glu₂₄₃···Arg₅₅₀). Similarly, in the 1CK7 structure, the propeptide domain only interacts with **HPX** through a salt bridge (Glu₉₅···Arg₄₉₅). In addition to these interdomain contacts, which are present in every single MMP-2 molecule in the crystal state, the unit cell of the X-ray structure contains six MMP-2 molecules among which several *intermolecular* contacts between the **HPX** domains of different MMP-2 molecules exist. Interestingly, the most important intermolecular interaction involves the hydrophobic clustering of residues situated at the fourth blade propeller of two different **HPX** domains (Tyr₆₃₆, Leu₆₃₈, Val₆₄₈, Phe₆₅₀, ...), covering 453 Å² of solvent-excluded molecular surface, which is clearly larger than the **CAT**–**HPX** interdomain contact (see Figure 1). In other words, the **HPX** domain contains different areas (patches), some of which are involved in intramolecular interactions and others in intermolecular protein–protein interactions. On the other hand, we have found in a previous computational work that the **CAT**–**HPX** interdomain contact observed in the initial X-ray structure is lost after a 100 ns MD simulation of a fully hydrated MMP-2 molecule, resulting in a quite remarkable rearrangement of the **HPX** domain with respect to the **CAT** and **FIB** domains that adopt an extended conformation during the simulation.¹⁶

Assuming that relatively ample interdomain motions can occur in solution and, as seen above, the **HPX** domain exhibits different regions favorable for protein–protein interactions, it would not be misconceived to think that along with the X-ray structure other conformations could be accessible for a single MMP-2 molecule in solution. Thus, our aim in the present work has been to explore the conformational landscape of the full-length MMP-2 in its active form (i.e., without the pro-peptide), aiming to find feasible conformations that are an alternative to the solid state structure reported experimentally. Clearly, an intensive conformational search using unbiased MD simulations in explicit solvent is computationally too expensive in the case of multidomain proteins like MMP-2, and therefore, other computational techniques should be considered. In this sense, there are several reported computational protein–protein docking methods that are able to efficiently sample alternative orientations between interacting proteins, with the goal of predicting the binding mode of the association.¹⁷ One of the most successful rigid-body docking and scoring schemes, pyDock,¹⁸ has been recently adapted to predict the binding mode of two domains joined by a flexible linker with the addition of end-to-end linker distance restraints, implemented in the module pyDockTET.¹⁹ As an example, the structure of a two-domain protein was proposed to be predicted from homologues of each individual domain in the blind test CAPRI (target 35; <http://www.ebi.ac.uk/msd-srv/capri/>), and

the only successful prediction among all participants was generated by pyDockTET.

Hence, we have used the docking program pyDock, with the module pyDockTET¹⁹ specifically developed to study domain–domain interactions, which has been customized for the present study in order to obtain new models of the MMP-2 enzyme in which the **HPX** domain explores alternative contact regions with the **CAT** and **FIB** domains.

Methods and Computational Details

pyDock Methodology. The first stage of our computational study applied the pyDock docking protocol,¹⁸ which is written in python and uses the MMTK set of python libraries²⁰ for parsing PDB files, calculating united atom AMBER 94 charges, and for other geometry manipulation tasks.

The pyDock docking protocol consists of four steps. In the first one, all the PDB files containing the macromolecules to be docked are preprocessed. This means that only the ATOM records of the 20 standard residues are kept, whereas everything else is removed. In other words, cofactors, hydrogens, and OXT along with HET records are systematically erased. These output PDB files become then the input files for the FTdock²¹ algorithm executed in the second step. FTdock is an algorithm within the 3D-Dock suite of programs designed to enable computational prediction of protein–protein conformations. The FTdock algorithm is based on that of Katchalski-Katzir et al.²² It discretizes the two molecules onto orthogonal grids and performs a global scan of translational and rotational space of possible positions of the two molecules, limited by surface complementarity and an electrostatic filter (optional). The latter is mainly used to discriminate, according to electrostatic favorability, between those complexes (poses) that have similar surface complementarity. FTdock 2.0 was used here, including the electrostatics filter to generate a total of 10 000 rigid-body docking orientations. Those poses are stored in terms of translational coordinates (x, y, z), expressed as integer grid cell displacements of the mobile molecule's center from the center of the static molecule, and rotational angles ($z_{\text{twist}}, \theta, \Phi$) expressed in degrees. In the third step, each individual geometry of the previously generated 10 000 poses undergoes a coordinate transformation into a suitable format for their use in the fourth and final step (i.e., a rotation and translation matrix for each pose is generated). Here, protein–protein docking poses are scored in order to predict their preferred binding geometry according to the following equation:

$$E = E_{\text{ele}} + E_{\text{desol}} + E_{\text{vdW}} \quad (1)$$

The first term of eq 1 stands for the Coulombic electrostatics where the distance-dependent dielectric constant ($\epsilon = 4r_{ij}$) has been explicitly calculated for all intermolecular atom pairs, with q atomic charges from the AMBER 94²³ force field in elementary charge units and pairwise interaction energy values truncated to a maximum and minimum of +1.0 and -1.0 kcal/mol, respectively, in order to avoid errors from incorrect geometries from the rigid-body approach. E_{desol} represents the effective water-to-interface desolvation

energy,^{24,25} and E_{vdW} is the Lennard-Jones van der Waals energy, also limited to a maximum of +1.0 kcal/mol to allow some interatomic clashes. Typically, E_{vdW} is weighted by a factor of $\omega_{\text{vdW}} = 0.1$ since the van der Waals term is somehow already implicitly included during the FTdock generation of docking poses.

Rigid-body docking poses of multiple domain proteins can be scored by a pseudoenergy term based on restraints derived from linker end-to-end distances.¹⁹ In this method, named pyDockTET (tethered-docking), the scoring function uses the average end-to-end distance, X_m , for a particular linker length (previously derived from a structural database¹⁹) as a restraint to select the correct docking poses. Then, the X_m value and its corresponding standard deviation (σ) are used to develop a function, E_{linker} , which is further incorporated (just by summing it) into the pyDock energy function for the final rescoring of domain–domain poses (for more details on the calculation of E_{linker} , see Figure 8 in ref 19). Essentially, pyDockTET evaluates the linker end-to-end distance for each independent pose, compares it with X_m , and introduces the corresponding energetic penalization according to it.

The conformation of the backbone chains in the rigid-body docking solutions is not refined, as the pyDock protocol has been extensively benchmarked for protein–protein interaction predictions both using internal tests and through the blind competition CAPRI, and the results have given top success rates without needing any refinement. While there have been successful attempts for backbone refinement,²⁶ and the PyDock developing team is actually working on the development of new refinement methods, their current protocol does not seem to significantly improve with respect to the rigid-body scoring.

Two additional tools used in the present work are the Optimal Docking Area (ODA) and the normalized interface propensity (NIP) analyses.^{24,25} The former is a method which enables the examination of any protein surface looking for areas with favorable energy change when buried upon protein–protein association. For that purpose, surface patches with optimal desolvation energy are identified. Such desolvation energy is based on atomic solvation parameters, derived from octanol/water transfer experiments,²⁴ adjusted for protein–protein docking. The NIP method analyzes the 100 lowest-energy solutions (higher-energy solutions do not have an impact on these results) to identify the residues that are most often involved in the docking interfaces, and which are probably involved in protein–protein interactions.

Setting up the pyDock Calculations on the MMP-2 System. Using the pyDockTET method, which as commented above has been specifically developed to study multiple domain proteins, implies that the multidomain molecule has to be split into two subdomains that are expected to be rigid. In the case of the MMPs, this rigid-body approximation is supported by the fact that the secondary structure of the **CAT** and **HPX** domains is very similar in the different X-ray structures regardless of their actual interdomain orientation. Thus, the RMS deviations of the backbone atoms of the **CAT** domain in the 1SU3 (MMP-1), 2CLT (MMP-1), and 3BA0 (MMP-12) structures with respect to that in the 2CLT structure (MMP-2) are 0.62, 0.71,

and 0.92 Å, respectively. The architecture of the **HPX** domain is also well conserved in the same set of structures, the corresponding RMSD values being 1.40, 1.34, and 1.34 Å. Therefore, we decided to formally divide the MMP-2 protein into *molecule A*, composed of the **CAT** and **FIB** domains, and *molecule B*, corresponding to the **HPX** domain. These two *molecules* need to be treated independently as if they are two different proteins. At this stage, the linker region was considered implicitly by adding the pyDockTET energy term to the scoring function.

Another modification was carried out before processing the MMP-2 PDB files. As discussed before, the MMP-2 system contains metallic ions that are already disregarded in the first step of the pyDock protocol. Obviously, this introduces a modification in the overall charge of the molecules to be docked and may, in principle, affect the final results. Indeed, the absence of metallic ions leaves free charged residues in the active site region which turn out to behave as “*strong attractors*”. Consequently, the docking solutions can be biased to conformations where the **HPX** domain is mostly interacting with the active site of the **CAT** domain (see Figure S2 in the Supporting Information) but that do not correspond to the active form of the MMP-2 enzyme. Then, we decided to modify the charge of some of those critical residues in the **CAT** domain in order to overcome this inconvenient. Thus, we mutated, where necessary, glutamic acid into glutamine, histidine into protonated histidine, and aspartic acid into asparagines. In particular, we modified the following residues: (a) Glu₄₀₄ and His₄₁₃, which are coordinated to Zn1; (b) Asp₁₈₀ and His₁₉₃ coordinated to Zn2; (c) Asp₁₈₅ and Glu₂₁₁ coordinated to Ca1; and (d) Glu₁₆₆ and Asp₂₀₄ coordinated to Ca2. Finally, 10 000 rigid-body docking poses were generated by FTdock, and they were further scored according to the pyDock and pyDockTET equations, as above-described.

Building of the MMP-2 Linker. To estimate the influence of the actual linker residues on the relative stability of the most stable poses, as well as to further discriminate among the family of complexes generated by the protein–protein docking calculations, we decided to generate the complete model from the docking solutions by rebuilding the linker region comprising the Asp₄₅₀–Ile₄₆₈ residues that are not taken into account during the pyDock calculations. In addition, bad contacts among the **CAT–FIB** and **HPX** interacting residues were also relaxed. The details of the computational procedure, which are also summarized in Figure S7 in the Supporting Information, are as follows:

1. First, we generated an ensemble of linker structures by means of restrained MD simulations of the following peptide sequence: Ace–Asp–Ile–Asp–Leu–Gly–Thr–Gly–Pro–Thr–Pro–Leu–Gly–Pro–Val–Thr–Pro–Glu–Ile–Nme. The AMBER03 force field,²⁷ which has been used in our previous simulations of the full-length MMP-2 enzyme,¹⁶ was coupled with the Hawkins–Cramer–Truhlar pairwise Generalized-Born (GB) solvent model²⁸ to carry out the MD simulations using the SANDER program included in the AMBER9 suite of programs.²⁹ We defined an end-to-end distance (X_m) as the distance between the C α carbons of the terminal Ace and Nme residues. The value of X_m was

restrained to a specified value using a harmonic biasing potential with a force constant of 10 kcal/mol/Å. We carried out a series of simulation windows beginning at an extended form which corresponds to $X_m = 36.0$ Å. The restrained end-to-end distance was then reduced by 0.25 Å steps down to 5.0 Å (125 windows). The end point of each window was used as a starting point for the next, and each window consisted of 200 ps of equilibration followed by 1.8 ns of production dynamics. The value of the reaction coordinate X_m was saved every 1.0 ps. The biased samplings obtained were used to derive potentials of mean force (PMF) for the end-to-end elongation of the peptide using the Weighted Histogram Analysis method (WHAM).³⁰

2. From each of the most stable pyDock poses (a total of 30 docking solutions), we built a family of full-length MMP-2 structures. To this end, we extracted a set of 40 equally spaced snapshots from the MD simulation of the isolated linker peptide, which were chosen in such way that their reference distance X_m matched the actual C α atoms of the last/first residues of the **CAT/HPX** domains in the corresponding pyDock structures. Each one of the linker structures is connected with the MMP-2 model by superposing the Ace/Nme heavy atoms of the linker peptide onto their counterpart atoms in the terminal **CAT/HPX** residues, and then removing the Ace/Nme coordinates. We also note that in these model building operations, we employed a full atomic representation of the **CAT–FIB** and **HPX** domains including H atoms and metallic ions. At this stage, a total of $30 \times 40 = 1200$ full-length complexes were generated.

3. For every single full-length MMP-2 model, steric clashes between the linker and the **CAT/HPX** atoms, or between the **CAT–FIB** and **HPX** domains, were iteratively identified and relaxed in the following manner. First, the SCWRL4 program³¹ for prediction of protein side-chain conformations was employed to rebuild the side chains of the residues involved in the corresponding steric clash. Then, the coordinates of the same residues were relaxed by carrying out 1000 conjugate gradient steps followed by 25 ps of MD using the AMBER03 force field and a distance dependent dielectric constant ($\epsilon = 4r_{ij}$). A high temperature value (500 K) was used in the restricted MD simulations in order to promote uphill moves of bulky side chains that can be important for properly relaxing some steric collapses. Once the loop over all the steric clashes was completed, the coordinates of the linker atoms and those of the **CAT–HPX–FIB** residues involved in the steric clashes were simultaneously optimized.

4. In principle, the total energy of the partially relaxed full-length MMP-2 models is not useful in obtaining a compensated energetic description because the number and identity of the MMP-2 residues that are structurally relaxed is different in each model. Hence, we combined the inter-domain interaction energies (**CAT–FIB**···**HPX** and **CAT–FIB–HPX**···**LK**) and the intrinsic stability of the **LK** region (which is fully relaxed) to assess the stability of the models, defining thus the following scoring function:

$$E = \Delta E_{\text{int}}^{\text{CAT–FIB} \cdots \text{HPX}} + \Delta E_{\text{int}}^{\text{CAT–FIB–HPX} \cdots \text{LK}} + E^{\text{LK}} \quad (2)$$

The required energies were obtained by using the Molecular Mechanical (MM) Poisson–Boltzmann Surface Area (MM-PBSA) approach, which has been applied to perform many classes of approximate binding energy calculations, including protein–protein complexes.³² Hence, we performed single-point MM-PBSA energy calculations using the SANDER program on the whole MMP-2 molecules and on the separated **CAT–FIB**, **HPX**, and **LK** domains (capped by Ace/Nme residues; the C-and N-terminal residues of **CAT** and **HPX** are removed). These calculations were performed for all 1200 models using the typical MM-PBSA settings in AMBER9.

Finally, the structural quality information of the most stable full-length MMP-2 models was analyzed by means of the WHAT_CHECK program.³³

Results and Discussion

We applied the pyDock protocol on two different sets of MMP-2 coordinates. On one hand, we used the X-ray (1CK7) structure reported experimentally.¹⁰ For this structure (**XR_MMP-2**), we deleted the coordinates of the propeptide residues, as we want to model the MMP-2 enzyme in its active form (Pro₃₁–Asn₁₀₉). For the same reasons, the N-terminal Tyr₁₁₀ ammonium group was placed interacting with the Asp₄₃₆ as in the so-called “superactivated” form.³⁴ On the other hand, we selected one MD snapshot from our previous MD study (**MD_MMP-2**), which corresponds to an extended configuration accessible for the full-length MMP-2 enzyme in solution.¹⁶ By using two different sets of coordinates, we can assess the sensitivity of the pyDock poses with respect to minor changes in the placement of the residue side chains and in the secondary structure of the protein domains.

For the two structures, **XR_MMP-2** and **MD_MMP-2**, we derived three series of docking solutions differing in the value of the end-to-end average distance (X_m) that is used in pyDockTET. Thus, the family of docking solutions labeled with the **Ink_10** suffix correspond to $X_m \sim 10 \pm 3.2 \text{ \AA}$. Similarly, **Ink_21** and **Ink_25** stand for $X_m \sim 21 \pm 5.0 \text{ \AA}$ and $X_m \sim 25 \pm 10 \text{ \AA}$, respectively. These average values and standard deviations were selected from the relation between the length and frequency of linkers in a database of 542 linker structures considered for the parametrization of the pyDockTET treatment. Although the length (number of residues) of the MMP-2 linker is fixed, we note that by using three different end-to-end X_m values, the **HPX** domain is allowed to adopt many more poses beyond those that are compatible with the relatively extended conformation of the linker region in the crystallographic 1CK7 structure ($X_m \sim 36 \text{ \AA}$). For example, the MMP-12 linker, which is only shorter by about three residues than the MMP-2 one, is folded in a more compact arrangement characterized by $X_m \sim 10 \text{ \AA}$.

For each individual combination of X_m distance and initial geometry (e.g., **MD_MMP-2_Ink21**), we analyzed in detail the most stable docking solutions falling within an energy interval of $\sim 10 \text{ kcal/mol}$. Typically, 10 different poses fall within such an energy interval. Figure 2 collects the 10 most stable docking solutions (according to the pyDockTET scoring function) obtained for each individual linker restric-

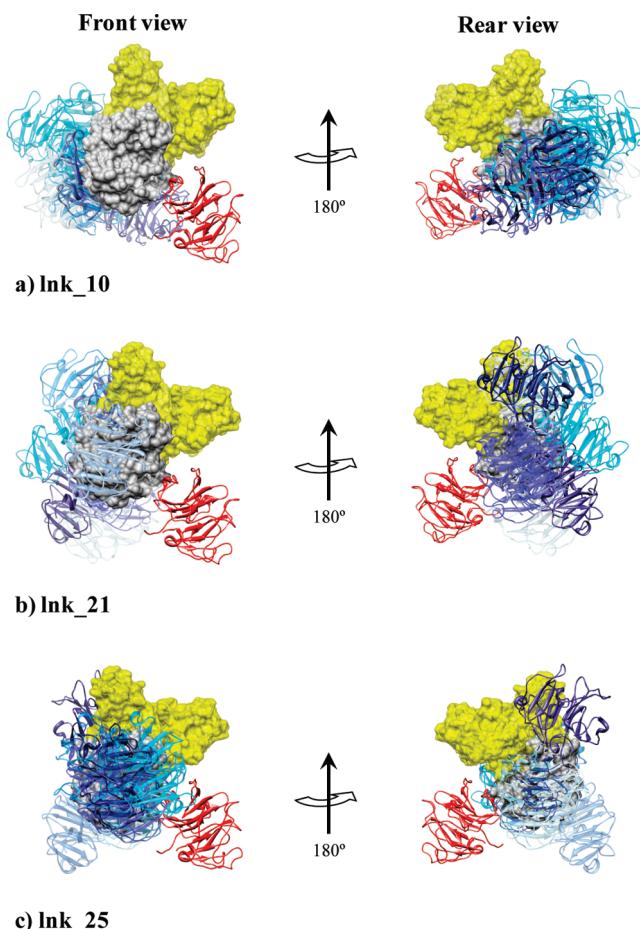


Figure 2. The 10 most stable **HPX** orientations (blue) according to the pyDockTET scoring function (**MD_MMP2**). The position of **HPX** in the X-ray structure is colored in red. Gray and yellow represent the **CAT** and **FIBs** domains, respectively. The gradient of blue colors reflects the different stability (the darker, the most stable) of the various docking solutions. *Front view* refers to the standard orientation of the MMP-2, whereas *rear view* results from rotating the frontal view 180° along the y axis.

tion imposed on the **MD_MMP-2** system. Interestingly, quite similar results were obtained for **XR_MMP-2** (see Figure S3 in the Supporting Information), and therefore, we concluded that the pyDock protocol is quite robust and predicts the same kinds of solutions for the MMP-2 system regardless of the actual protein side chain conformations. In what follows, we will present only the results obtained for the **MD_MMP-2** geometry.

Protein–Protein Contacts Favored by the pyDock Calculations. The first observation we can make about the results shown in Figure 2 is that none of the most stable docking solutions predicts an orientation for the **HPX** domain (structures shown in blue) similar to that observed in the crystallographic structure (shown in red). Hence, there is a chance that the **CAT–FIB** domains may have another patch suitable for a favorable interaction with **HPX**; i.e., there may exist alternative conformations of the activated and solvated MMP-2 to that reported experimentally for the proenzyme in the solid state. Such affirmation needs, however, a deeper analysis of the data obtained. Inspection of Figure 2 also shows that depending on the formal end-to-end distance of

the linker, the docking solutions explore two different areas of the **CAT–FIB** domains: (a) the active site groove, particularly in case of **Ink_25**, and (b) the rear of the **CAT** and **FIB** domains (mostly for **Ink_10** and **Ink_21**). No docking solutions are placed in the front part of the **FIB** domains.

Solutions where the **HPX** domain is interacting with the active site cleft could be, in a sense, reasonable results provided by the docking calculations given that the MMP-2 active site region actually binds and hydrolyzes other protein systems. Certainly, if the linker adopts an extended conformation with average end-to-end X_m values of 21–25 Å, then the **HPX** domain may be placed in front of the active site (circumstance not possible with a more compact conformation of the linker), but those solutions should be disregarded in our analyses since we are mainly interested in analyzing the active form of the enzyme. Focusing now on the solutions placed in the rear of the **FIB** and **CAT** domains, the following can be observed. In case of large X_m values (**Ink_21** and **Ink_25**), few solutions are placed in a position where the **HPX** is interacting with the rear of the **FIB** domains. More specifically, there are three solutions in the case of **Ink_21** and one solution in the case of **Ink_25**, suggesting that such a region is a potential patch for protein–protein interactions. The remaining solutions are all placed in the rear of the **CAT** domain. This area constitutes then another patch for protein–protein interactions, though it is hard to be precise with a region within the rear of the **CAT** domain where the interaction with **HPX** would be more favorable. Notice, however, that as a general observation we can affirm that the more compact the linker is, the greater the number of structures are interacting with the $\beta 1$ and $\beta 2$ strands of the **CAT** domain. Finally, in terms of abundance, we can establish the following ranking: interactions **HPX–FIB** (H–F) < interactions **HPX–active site** (H–A) < interactions **HPX–CAT** (H–C), where the symbol < implies that a smaller amount of docking solutions show this type of interaction.

As commented upon in the Introduction, the X-ray MMP-12 structure (3BAO) is characterized by a less compact arrangement (in comparison with, e.g., 1CK7) where the **HPX** domain is oriented toward the rear of the **CAT** domain (see Figure S4, Supporting Information). Moreover, in MMP-12, the fourth blade of the **HPX** domain is oriented toward the **CAT** domain as in the case of the most stable docking solutions reported here (see below). Thus, we checked if the orientations of the **HPX** domains of any of the docking solutions collected in Figure 2 matched the orientation of the **HPX** domain in MMP-12. Interestingly, a certain degree of parallelism in the relative orientation of the **HPX** and **CAT** exists between the MMP-12 structure and the MMP-2 docking solutions (see Figure S5, Supporting Information).

Relative Stability of the pyDock Poses. Unfortunately, no systematic behavior exists in terms of energy, meaning that no type of interaction is significantly more stable than any other (see Table 1). For instance, the pose obtained as the most stable solution in the **Ink_25** set, but also that is ranked number eight in **Ink_21**, is of the H–A type. One thing that can be easily seen when analyzing Table 1 is that

Table 1. Relative Energies (kcal/mol) for the 10 Most Stable Docking Solutions in Each of the Three Series of pyDockTET Docking Calculations with Different Linker Restrictions^a

Ink_10	interaction	Ink_21	interaction	Ink_25	interaction
-64.51	H–C	-78.56	H–F	-84.32	H–A
-62.71	H–C	-69.56	H–C	-78.56	H–F
-62.16	H–C	-69.50	H–C	-77.03	H–A
-61.02	H–C	-69.37	H–C	-75.38	H–A
-60.75	H–C	-68.56	H–F	-71.20	H–A
-60.28	H–C	-67.21	H–C	-69.83	H–A
-59.45	H–C	-66.44	H–C	-69.83	H–A
-57.62	H–C	-66.41	H–A	-69.56	H–A
-57.06	H–C	-65.16	H–C	-69.50	H–A
-55.11	H–C	-65.10	H–F	-69.37	H–A

^a Specific domains interacting for each particular solution are indicated in the columns labeled “interaction”.

docking solutions are highly biased depending on the end-to-end distance. For example, the H–C interactions are systematically favored in the **Ink_10** solutions, whereas in the **Ink_25** set the H–A interactions are the most abundant ones. An intermediate situation (**Ink_21**) allows a more even distribution. If for the above commented reasons solutions of the H–A type are disregarded, the most stable solution (H–F type) is found in the **Ink_21** family (this same pose is also the second most stable one in the **Ink_25** ranking). In this structure, the actual value of the linker end-to-end distance measured as the C α –C α distance belonging to the last residue of **CAT–FIB** and the first residue of **HPX** is 25.6 Å, and the **HPX** domain interacts with the rear of the **FIB** domains (see the darkest blue solution in Figure 2, rear view of inset b).

The Fourth Blade in HPX as a Likely Interaction Site. In the most stable pyDock solutions, it turns out that the **HPX** domain interacts preferentially with the first and the second domains of **FIB** via its fourth blade propeller (see Figure S6 in the Supporting Information). Indeed, we have confirmed that in the rest of the 10 most stable solutions (including not only the ones corresponding to the **Ink_21** but also to the **Ink_10** and **Ink_25** sets), **HPX** interacts again via its fourth blade. Interestingly, this is in agreement with the behavior found in the intermolecular **HPX**···**HPX** contacts in the 1CK7 crystal structure (see Figure 1) and in the complex between MMP-2 and the TIMP-2 inhibitor,^{11,35} where the **HPX** domain is also interacting via its fourth blade with TIMP-2. However, the intramolecular **HPX**···**CAT** interactions in the 1CK7 structure involve mainly the first **HPX** blade. It may be interesting to note that further statistical analysis performed on 5000 docking poses showed that a large number of solutions in which the **HPX** domain is interacting via its first blade are also obtained, but they are less stable than those involving interactions with the fourth blade.

The **HPX** interaction sites can also be characterized by mapping residue NIP values onto the protein surface. We focused on those residues with NIP values between 1 (residues appear in the interface in all the docking solutions) and 0 (residues are found in the docking interfaces as frequently as randomly expected). For **HPX**, the maximum NIP value is 0.3 and corresponds to the Phe₆₅₀ residue. Then,

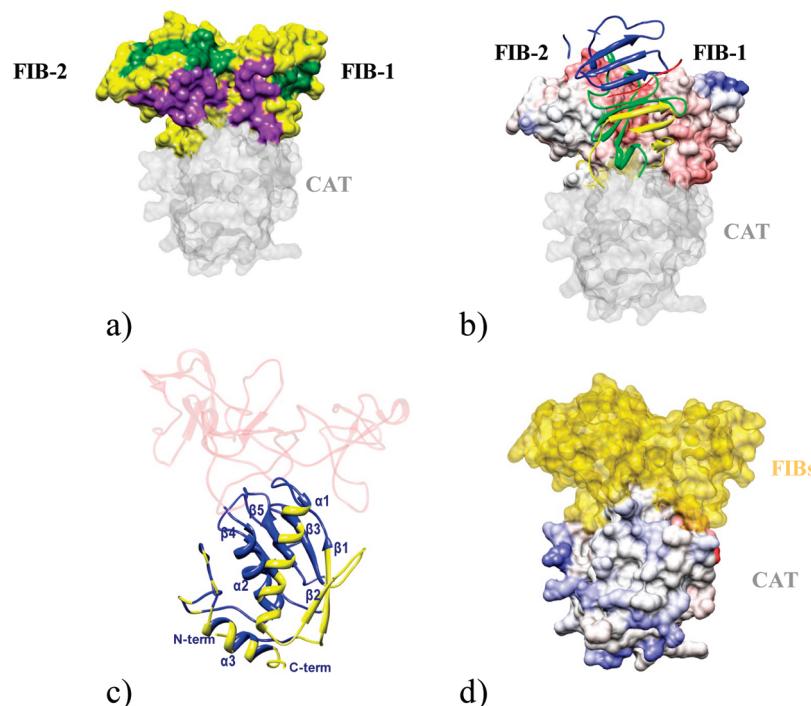


Figure 3. (a) Binding regions of the **FIB** subdomains of MMP-2 colored according to experimental (green) and theoretical (purple) predictions. Gray and yellow surfaces correspond to the **CAT** and the third **FIB** domains, respectively. (b) The **HPX** domain is represented in ribbons and its four blades colored in yellow (first blade), red (second blade), dark blue (third blade), and green (fourth blade). The first and second domains of **FIB** are colored according to ODA desolvation energies. Contrary to the blue areas, the red areas have low desolvation energy. (c) **CAT** domain of MMP-2 colored according to NIP values (yellow means values ranging within 0.4–0.0). (d) ODA representation of the rear of the **CAT** domain.

we checked whether any residue with NIP values ranging between 0 and 0.3 agree with those observed by Piccard et al.³⁵ and Morgunova et al.¹¹ in their experimental study about the interaction of MMP-2 and the TIMP-2 inhibitor. According to Piccard et al., MMP-2 recognizes the so-called GH loop of TIMP-2 by means of Ala₆₁₂, Tyr₆₃₆, Leu₆₃₈, Val₆₄₈, and most significantly, Phe₆₅₀, located at the fourth blade of the **HPX** domain. According to our calculations, those same five residues show the highest NIP values (see Table S1 in the Supporting Information), and more specifically, residue Phe₆₅₀ is playing a major role according to both the experimental and the theoretical studies. Consequently, there is a nice agreement between the theoretical predictions concerning the nature of the **HPX** interaction sites and closely related experimental data.

Protein–Protein Interaction Sites in the CAT–FIB Domains. Few experimental studies have reported information on the gelating binding regions of **FIBs**.^{36,37} Essentially, gelating binding sites have been identified in each of the three **FIB** subdomains and are formed by three conserved clusters of Phe, Trp, and Tyr residues and their surrounding region (colored in green in Figure 3a for **FIB-1** and **FIB-2**). The docking analyses show that **HPX** can interact simultaneously with **FIB-1** and **FIB-2** through a region (colored in purple in Figure 3a) quite close to the gelating binding areas of the two subdomains. For instance, residues Tyr₃₁₄ and Phe₃₃₁ that form part of the hydrophobic cluster of **FIB-2** also present positive NIP values (see Table S1, Supporting Information). Additionally, ODA calculations show that the

region of **FIB** that interacts with **HPX** is a surface patch of optimal desolvation energy (see Figure 3b).

We also looked at the NIP values of the **CAT** domains looking for information that could help to further rationalize the data obtained (see Figure 3c). A few residues belonging to the **CAT** domain show NIP values higher than 0, suggesting, thus, the existence of patches (colored in yellow in Figure 3c) in the **CAT** domain that can likely interact with other proteins. Those patches are quite spread along the β_1 – β_2 strands and the α_1 and α_3 helices (see Figure 3c), with no residues with $NIP > 0.4$; that is, no “hot spot” residues for protein interactions can be distinguished. This is in consonance with ODA analyses, indicating that the molecular surface of the rear of the **CAT** domain has neutral desolvation energy (see Figure 3d).

Rebuilding the Linker Region. The free energy profile generated with the implicit GB solvent model for the end-to-end elongation of the linker peptide shows only a minimum located at approximately 10.5 Å (see Figure 4). In fact, the PMF profile explores low energy states (<1.25 kcal/mol) within the 10–36 Å range. Although the use of explicit water models would likely improve the accuracy of the conformational free energy differences, the present PMF calculations suggest that the amino acid sequence of the MMP-2 linker region is intrinsically flexible in aqueous solution and that a broad range of end-to-end distances (10–36 Å) could be accessible at room temperature.

Besides estimating the PMF for the end-to-end elongation of the linker peptide, each MD simulation window provided an ensemble of representative structures having the appropri-

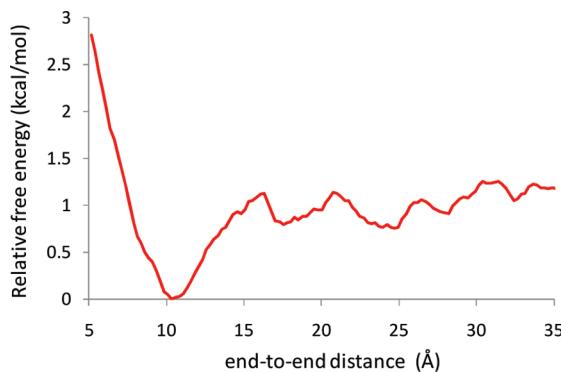


Figure 4. Potential of mean force for the isolated linker peptide.

ate end-to-end distance in order to match the corresponding $\text{C}\alpha$ – $\text{C}\alpha$ distance between the last **CAT** residue and the first **HPX** residue in the most stable pyDock complexes. In this way, and following the prescriptions detailed in the Methods and Computational Details, we were able to rebuild the full-length MMP-2 enzyme by connecting the **CAT** and **HPX** through a partially relaxed peptide chain as well as fix several bad contacts among side chains in the interaction zone between the **CAT–FIB** and **HPX** domains. In the majority of the 1200 full-length MMP-2 structures that were generated from the most stable pyDock poses, interactions between the linker residues and the nearby **CAT/HPX** residues are structurally favorable. However, we also found that several steric clashes could not be fully relaxed through the combination of restrained minimization and MD calculations of the involved residues. Moreover, the backbone of the linker region tends to adopt a strained conformation in many structures.

To translate into a scoring function the diverse structural quality of the full-length MMP-2 models, we estimated both the interaction energies between the **CAT/HPX** domains and the linker region and the intrinsic stability of the linker residues by means of the MM-PBSA method (see Figure 5). However, we note that neither the restricted relaxation of the MMP-2 models generated by the rigid docking calculations nor the limitations of the MM-PBSA methodology for predicting binding or conformational energies of large systems allow us to make clear-cut energetic predictions. Nevertheless, the pyDock structures that are best fitted to accommodate the linker residues should be well captured

by the MM-PBSA scoring function defined in eq 2 thanks to a partial cancellation of errors. Thus, we found that the second solution in the **Ink_10** pyDockTET set, which belongs to the H–C type (see above), generates full-length MMP-2 structures that are much more stable by tenths of a kilocalorie per mole than the other solutions in the same set. The best structure arising from this pyDockTET pose has an end-to-end distance of \sim 13 Å and is characterized by a rather compact arrangement of the linker chain, which is, nevertheless, well fitted to the surrounding protein environment (see Figure 5a). In the case of the **Ink_21** and **Ink_25** sets of pyDockTET solutions, their relatively large end-to-end distances are more compatible with the inclusion of the linker residues, but it turns out that placement of the linker chain stabilizes preferentially the pyDockTET solutions presenting the H–F interaction instead of the H–C one. One of these models with a $\text{C}\alpha$ – $\text{C}\alpha$ separation of 24 Å is shown in Figure 5b, in which we observe how the disposition of the **CAT–FIB** and **HPX** domains seems particularly suitable to accommodate the linker chain in a semiextended conformation, which is relatively stable according to the MM-PBSA calculations. Overall, the two full-length MMP-2 models shown in Figure 5 confirm that the interdomain contacts predicted by the pyDockTET solutions are compatible with the actual molecular structure of the linker chain.

Summary and Conclusions

The rigid-body protein–protein docking calculations reported in this work point out that, in the absence of crystallographic contacts and/or other proteins, a single MMP-2 molecule in its active form (i.e., in the absence of the pro-peptide) can adopt different conformations in aqueous solution with respect to that observed by X-ray crystallography. Further details about the actual structure and flexibility of the essential linker region connecting the MMP-2 domains are obtained through a series of model-building operations and MM calculations in which MD snapshots of the linker peptide and the most stable docking solutions are combined. On the basis of our results, we can also draw specific conclusions concerning the multidomain structure of the active form of the MMP-2 enzyme:

- The **HPX** domain tends to interact either with the rear part of the **CAT** domain (preferably with the five-stranded β sheet) or with the first and second **FIB** subdomains.

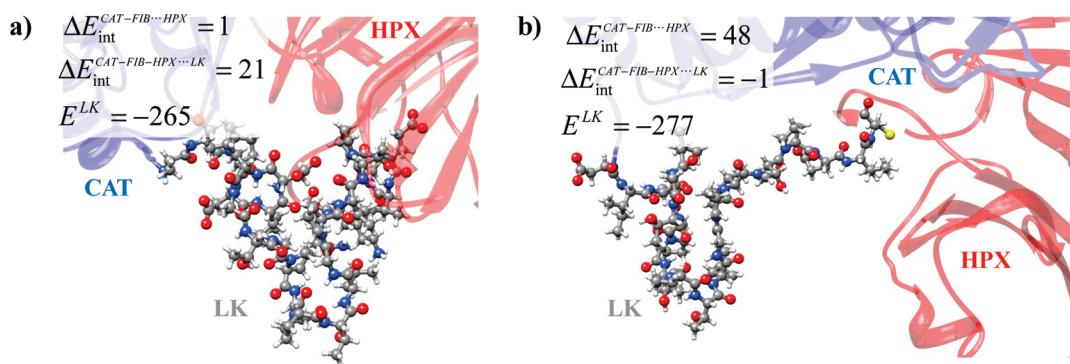


Figure 5. Ball-and-stick views of the linker region adapted to two pyDock solutions corresponding to the **Ink_10** (a) and **Ink_21** (b) sets of **MD_MMP2**. MM-PBSA interdomain interaction and conformational linker energies (kcal/mol) are also indicated.

- The most likely interaction region of the **HPX** domain is constituted by a hydrophobic cluster of residues (Ala₆₁₂, Tyr₆₃₆, Leu₆₃₈, Val₆₄₈, and Phe₆₅₀) located at its fourth blade propeller. In fact, these residues are known to interact either with the TIMP-2 inhibitor or with a second MMP-2 molecule placed in the same crystallographic unit.

- The MD and PMF calculations carried out on the isolated linker region indicate that the linker peptide can easily adopt a large range of end-to-end distances. This seems in consonance with the proposed interdomain flexibility in the MMP-2 enzyme.

- The global interdomain orientations favored by the pyDock calculations are compatible with the molecular structure of the linker residues as confirmed by our full-length MMP-2 models exhibiting a relaxed linker chain connecting their **CAT** and **HPX** domains.

From a methodological point of view, we believe that the computational protocol employed in this work, which is essentially characterized by the rigid-body Pydock calculations and the subsequent reconstruction of linker atoms and removal of bad contacts using all atom MM calculations, could be of interest for other applications. Thus, this strategy, which mixes diversity and likeness in the predicted interdomain conformations, can easily generate a pool of structures for which small-angle X-ray scattering patterns and/or NMR properties could be calculated and used for data analyses of experimental measurements.¹⁵ Similarly, the most-likely Pydock structures could be very useful for further computational studies aimed at the elucidation of the role played by water molecules and protein dynamics in the stability of interdomain conformations. In this respect, we note that the generation of reliable docking structures like those reported in this work for the MMP-2 enzyme could be seen as a prerequisite before carrying out extensive MD simulations in explicit solvent and more sophisticated free energy calculations. As a matter of fact, further computational and experimental work will be required in order to understand the specific roles played by each of the MMP-2 domains during collagenolysis.

Acknowledgment. This research was supported by Grants CTQ2007-63266 and BIO2008-02882 (MICINN, Spain). The authors thankfully acknowledge the computer resources, technical expertise, and assistance provided by the Barcelona Supercomputing Center—Centro Nacional de Supercomputación. H.V. is grateful to Carles Pons (INB—BSC) for his valuable help in using the pyDock program and also thanks MEC (Spain) for her contract and BSC for a mobility program fellowship (BCV-2009-1-0013).

Supporting Information Available: Table S1 and Figures S1–S7. This information is available free of charge via the Internet at <http://pubs.acs.org/>.

References

- Sternlicht, M. D.; Werb, Z. *Ann. Rev. Cell. Dev. Biol.* **2001**, 17, 463–516.
- McCawley, L. J.; Matrisian, L. M. *Curr. Opin. Cell Biol.* **2001**, 13, 534–540.
- Coussens, L. M.; Werb, Z. *Chem. Biol.* **1996**, 3, 895–904.
- van Meurs, J.; van Lent, P.; Holthuysen, A.; Lambrou, D.; Bayne, E.; Singer, I.; Berg, W. v. d. *J. Immunol.* **1999**, 163, 5633–5639.
- Nagase, H.; Woessner, J. F., Jr. *J. Biol. Chem.* **1999**, 274, 21491–21494.
- Maskos, K. *Biochimie* **2005**, 87, 249–263.
- Diaz, N.; Suarez, D. *Biochemistry* **2007**, 46, 8943–8952.
- Jozic, D.; Bourenkov, G.; Lim, N.-H.; Visse, R.; Nagase, H.; Bode, W.; Maskos, K. *J. Biol. Chem.* **2005**, 280, 9578–9585.
- Iyer, S.; Visse, R.; Nagase, H.; Acharya, K. R. *J. Mol. Biol.* **2006**, 362, 78–88.
- Morgunova, E.; Tuuttila, A.; Bergmann, U.; Isupov, M.; Lindqvist, Y.; Schneider, G.; Tryggvason, K. *Science* **1999**, 284, 1667–1670.
- Morgunova, E.; Tuuttila, A.; Bergmann, U.; Tryggvason, K. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, 99, 7414–7419.
- Bertini, I.; Calderone, V.; Fragai, M.; Jaiswal, R.; Luchinat, C.; Melikian, M.; Mylonas, E.; Svergun, D. *J. Am. Chem. Soc.* **2008**, 130, 7011–7021.
- Bode, W. *Structure* **1995**, 3, 527–530.
- Rosenblum, G.; Van den Steen, P. E.; Cohen, S. R.; Grossmann, J. G.; Frenkel, J.; Sertchook, R.; Slack, N.; Strange, R. W.; Opdenakker, G.; Sagi, I. *Structure* **2007**, 15, 1227–1236.
- Bertini, I.; Fragai, M.; Luchinat, C.; Melikian, M.; Mylonas, E.; Sarti, N.; Svergun, D. *I. J. Biol. Chem.* **2009**, 284, 12821–12828.
- Diaz, N.; Suarez, D.; Valdes, H. *J. Am. Chem. Soc.* **2008**, 130, 14070–14071.
- Ritchie, D. W. *Curr. Protein Pept. Sci.* **2008**, 9, 1–15.
- Cheng, T. M.; Blundell, T. L.; Fernandez-Recio, J. *Proteins* **2007**, 68, 503–515.
- MK Cheng, T.; Blundell, T. L.; Fernandez-Recio, J. *BMC Bioinf.* **2008**, 9, 441–453.
- Hinsen, K. *J. Comput. Chem.* **2000**, 21, 79–85.
- Gabb, H. A.; Jackson, R. M.; Sternberg, M. J. *J. Mol. Biol.* **1997**, 272, 106–120.
- Katchalski-Katzir, E.; Shariv, I.; Eisenstein, M.; Friesem, A. A.; Aflalo, C.; Vakser, I. A. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, 89, 2195–2199.
- Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, 117, 5179–5197.
- Fernandez-Recio, J.; Totrov, M.; Abagyan, R. *J. Mol. Biol.* **2004**, 335, 843–865.
- Fernandez-Recio, J.; Totrov, M.; Skorodumov, C.; Abagyan, R. *Proteins* **2005**, 58, 134–143.
- Mashiach, E.; Nussinov, R.; Wolfson, H. J. *Prot. Struct. Funct. Bioinf.* **2010**, 78, 1503–1519.
- Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J.; Kollman, P. A. *J. Comput. Chem.* **2003**, 14, 1999–2012.
- Srinivasan, J.; Cheatham, T. E.; Cieplak, P.; Kollman, P.; Case, P. A. *J. Am. Chem. Soc.* **1998**, 120, 9401–9409.

- (29) Case, D. A.; Darden, T. A.; Cheatham, T. E.; Simmerling, I. C. L.; Wang, J.; Duke, R. E.; Luo, R.; Merz, K. M.; Pearlman, D. A.; Crowley, M.; Walker, R. C.; Zhang, W.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Wong, K. F.; Paesani, F.; Wu, X.; Brozell, S.; Tsui, V.; Gohlke, H.; Yang, L.; Tan, C.; Mongan, J.; Hornak, V.; Cui, G.; Beroza, P.; Mathews, D. H.; Schafmeister, C.; Ross, W. S.; Kollman, P. A. *AMBER9*; University of California: San Francisco, CA, 2006.
- (30) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, D. R. H.; Kollman, P. A. *J. Comput. Chem.* **1995**, *16*, 1339–1350.
- (31) Krivov, G. I.; Shapovalov, M. L.; Dunbrack, R. L. *Prot. Struct. Funct. Bioinf.* **2009**, *77*, 778–795.
- (32) Gohlke, H.; Case, D. A. *J. Comput. Chem.* **2003**, *25*, 238–250.
- (33) Hooft, R. W. W.; Vriend, G.; Sander, C.; Abola, E. E. *Nature* **1996**, *381*, 272–272.
- (34) Díaz, N.; Suárez, D. *Proteins: Struct., Funct., Bioinf.* **2008**, *72*, 50–61.
- (35) Piccard, H.; Van den Steen, P. E.; Opdenakker, G. *J. Leukoc. Biol.* **2007**, *81*, 870–892.
- (36) Gehrmann, M. L.; Douglas, J. T.; Bányai, L.; Tordai, H.; Patthy, L.; Llinás, M. *J. Biol. Chem.* **2004**, *279*, 46921–46929.
- (37) Briknarová, K.; Grishaev, A.; Bányai, L.; Tordai, H.; Patthy, L.; Llinás, M. *Structure* **1999**, *7*, 1235–1245.

CT100097X