Journal of Computer-Aided Molecular Design **17:** 837–848, 2003.

Molecular dynamics simulation of Matrix Metalloproteinase 2: fluctuations and time evolution of recognition pockets

Mattia Falconi[a], Gioia Altobelli[a], Maria Cristina Iovino[a], Vincenzo Politi[b] and Alessandro Desideri[a,*]

[a]*INFM and Department of Biology, University of Rome 'Tor Vergata', Via della Ricerca Scientifica, I-00133 Rome, Italy;* [b]*Polifarma S.p.A., Via di Tor Sapienza 138, I-00155 Rome, Italy*

Received 23 July 2003; accepted in revised form 18 November 2003

Key words: long range communication, matrix metalloproteinases, protein flexibility, recognition pockets, solvent accessibility

Summary

We report a molecular dynamics simulation study of a zinc-protease – gelatinase A or MMP2 – which is a major target for drug design, being involved in tumor metastasis and other degenerative diseases. Two structures have been employed as starting conditions, one based on the crystal of multi-domain proMMP2, the other consisting of the catalytic domain only. The overall fold of the two models is maintained over the 1260 ps trajectory, enabling us to analyze correlations of fluctuations among domains, and to observe the presence of correlations within the catalytic domain in the multi-domain enzyme only, hence due to the presence of hemopexin and fibronectin domains. In the multi-domain protein, two cavities are conserved over the trajectory, one of them pointing to a key region, a crevice surrounding the catalytic zinc. The other one is localized across the three domains of the MMP2 metalloproteinase. These areas are partially covered by the propeptide in the crystal structure of proMMP2. We propose a model of MMP2-collagen interaction that involves both identified cavities and takes into account the inter/intra domain cross-correlations.

Introduction

Among matrix metalloproteinases – a family of more than 20 highly homologous proteins – matrix metalloproteinase 2, MMP2, has emerged as a major target for drug design, being involved in tumor metastasis and other degenerative diseases [1, 2]. MMP2 is a multidomain enzyme made of a catalytic domain, a hemopexin-like domain and a fibronectin-like domain. It is expressed as a proenzyme and tightly regulated at multiple levels. Endogenous tissue inhibitors of metalloproteinases, TIMPs, regulate the extracellular activity of MMPs, and their altered expression has been associated to many diseases [3]. TIMP2 is also involved in the activation of proMMP2 [4].

*To whom correspondence should be addressed. Fax: +39-06-72594326; E-mail: desideri@uniroma2.it

So far drug design efforts have addressed specificity of inhibition, focusing especially on the hydrophobic pocket S1′ shown by the available crystal structure of MMP2 catalytic domain [5]. The depth of pocket S1′ may act as a filter among the highly homologous MMPs [1, 6], and several inhibitors that bind S1′ with about nanomolar affinity have been developed [1]. Up to now, no synthetic inhibitor has been proved to be effective in patient treatment [2]. These shortcomings suggest that new approaches should be tried, addressing other features of the multi-domain MMP2 for drug design, including for example the unique MMP2/TIMP2 association recently elucidated [4], or a mechanism-based design by imparting or reestablishing the proenzyme structural motif [7, 8].

The crystal structure of the multidomain MMP2 in the proenzyme form has been also solved [9], and such a structure seems to be more suitable for the investigation of the catalytic properties of the enzyme,

because of the possible influence that hemopexin and fibronectin domains may have on the catalytic one. The catalytic domain, in fact, retains some catalytic activity when separated from the MMP2 [5], even though its efficiency compared to that of the full MMP2 is not actually reported in literature. Moreover, the full collagenolytic activity can be fulfilled by the multidomain enzyme only [10, 11].

The aim of our study was to examine the active form of MMP2 structure (Figure 1) as it is in aqueous solution, and thus identify regions of flexibility that might help to shed light on the functionality of MMP2. We performed two molecular dynamics simulations, one for the crystal form of the catalytic domain [5] and the other for an active multidomain MMP2 model based on the X-ray structure of inactive MMP2 [9], for assessing the role of hemopexin and fibronectin domains in modulating the function of the catalytic domain. The overall fold of the two protein systems was maintained over the 1260 ps trajectory, enabling us to analyze conformational properties of proteins and their internal correlations. We observed the dynamic conformational changes of recognition pockets and correlations of atomic fluctuations within the catalytic domain, which are due to the presence of the other domains. Our study shows that the hydrophobic pocket S1′, as much as the other crystallographic ones (S2′, S3′, S1, S2, and S3), has no permanent volume over the entire 1.26 ns molecular dynamics trajector in MMP2 protein. On the other hand, we characterize a conserved cavity near the catalytic zinc, which might be exploited for selective inhibition of MMP2, and construct a model of MMP2-collagen interaction that involves coordinated intra/inter domain movements. We discuss our findings in light of the available experimental data.

Computational methods

Model systems

Atomic coordinates of multi-domain proMMP2 at 2.8 Å resolution [9], and of its catalytic domain at 2.8 Å resolution [5], were extracted from the RCSB Protein Data Bank [12] (web address http://www.rcsb.org/pdb; entry code 1CK7 and 1QIB respectively). The package *Sybyl 6.0* by Tripos Associates (St. Louis, MO, USA) was employed for molecular modeling of the active multi-domain MMP2 and of its complex with collagen, based on X-ray crystallographic structures [9, 13, 14]. In the 1CK7 PDB file [9], the first chain (fibronectin and catalytic domains) is numbered from 31 to 449, while the second chain (hemopexin domain) is numbered from 461 to 660, for a total of 620 residues. We eliminated the pro-peptide molecule from the 1CK7 pdb file (sequence range: 31–109), and mutated Ala→ Glu at the site 404, in order to model the active form of the MMP2. The inverse mutation Glu→ Ala at this site – where Glu coordinates the catalytic zinc ion – does not affect the 3D-fold of the catalytic domain, which enables the crystallization of the complete MMP2 otherwise prevented by autoproteolysis [9]. Since the final number of residues in the multidomain protein system is 540, we renumbered the residues from 1 to 540 for the 1CK7 molecular dynamics simulation. The total number of residues in the catalytic domain system is 161 for the 1QIB simulation.

Matrix metalloproteinase MMP2 is composed of three domains, once the pro-peptide is removed (Figure 1). These are as follows: the catalytic domain, the N and C terminal domains, which are respectively referred to as *fibronectin* and *hemopexin* domains in the following. A short hinge pseudo-domain, which links the hemopexin to the catalytic domain, is also observed. The structure of the N terminal domain (*fibronectin*) is composed of three fibronectin type II modules, while four hemopexin modules form the distinctive 4-blade β-propeller structure of the C terminal domain (*hemopexin*), which contains one calcium ion [9]. The catalytic domain contains 3 α-helices and 5 β-sheets are arranged in a typical matrixin fold, and two zinc ions (named in the following *catalytic* and *structural*). There are two calcium ions in the catalytic domain of 1CK7 crystal structure (multi-domain MMP2) and three calcium ions in the 1QIB crystal structure (MMP2 catalytic domain separated). In the following, the moiety of the catalytic domain which contains calcium ions is named the *structural* moiety, as opposite to the *catalytic* moiety, which contains a zinc ion only.

In order to model the metal clusters of zinc and calcium ions, we employed the program *MOPAC 5.07mn* [15], a general-purpose semi-empirical molecular orbital package. Net charge of each cluster is +2, resulting from the sum of partial charges distributed on metal ions and selected atoms of the ligand side chains that coordinate the metal ions. The ligand atoms are as follows: for the zinc coordinating clusters, atoms of histidine imidazole rings and aspartic acid carboxyl group; for the calcium coordinating clusters, atoms

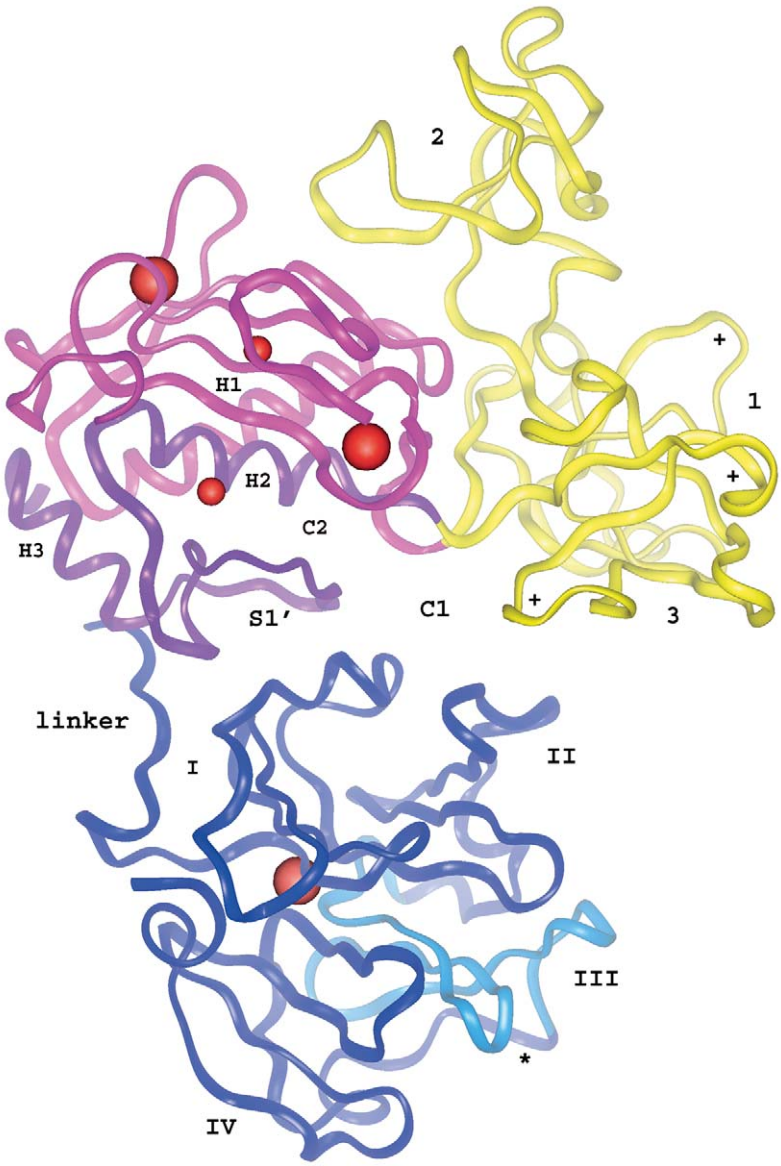

Figure 1. Ribbon model of MMP2 multi-domain protein (based on 1CK7 pdb [9]). Red spheres represent calcium (larger radius) and zinc (smaller radius) ions. Catalytic moiety of catalytic domain is indicated by purple ribbon; structural moiety by magenta. Explicitly labeled are: the α-helices H1, H2 H3, and loop S1′ in the catalytic domain; the three fibronectin modules (yellow ribbon), the four blades of hemopexin domain and the linker pseudo-domain (blue ribbon). C1 and C2 indicate the regions where dynamically identified cavities would lie in the crystal structure of MMP2 multi-domain protein. Crosses, open circle and star indicate the most mobile loops of the protein. The star also labels one of the two binding regions of the TIMP2 C-tail [4].

of aspartic and glutamic acid carboxyl groups and those of the peptide group (C, O, N) of residues that coordinate through carbonyl groups.

Molecular dynamics

A 1.26 ns trajectory of molecular dynamics simulation (MD) for both systems has been extracted from a longer trajectory of 1.6 ns. The proteins were embedded in water molecules. The number of water molecules was 17,824 in the multi-domain MMP2 system, where the total number of atoms was 58,701 (hereafter called 1CK7 simulation); and 4,105 in the catalytic domain system, where the total number of atoms was 13,857 (hereafter called 1QIB simulation).

There were at least three layers of water around the proteins. Two sodium ions were added to make the multi-domain system electroneutral, and six chloride ions were used to neutralize the catalytic system. Periodic boundary conditions [16] have been employed. The equilibrium properties of solvated systems were sampled in the isothermal-isobaric (NPT) ensemble [17]. The temperature chosen for our study was 300 K, while pressure was kept fixed at 1.0 atm.

The two systems, which dramatically differ in size (one has about one fourth the number of atoms of the other), were equilibrated with the following procedures. For the multi-domain system, it started with a 1000-step trajectory at temperature 40 K, then it proceeded by gradually increasing temperature in steps of 20 K up to 300 K – the temperature at which a 30,000-step trajectory was performed. For the catalytic domain system, it started with a 2000-step trajectory at temperature 20 K, then it proceeded by gradually increasing temperature in steps of 20 K up to 300 K, where a 35,000-step trajectory was performed.

The MD integration time step was 1.0 fs. The simulations were performed using the serial computer code *DL-PROTEIN 2.1* [18, 19]. The 1CK7 simulation was carried out on Compaq alpha server DS20E, and, excluding thermalization, took about 5760 hours of CPU time. The 1QIB simulation was carried out on sgi Origin 200, and required about 1440 hours of CPU time, excluding thermalization. Optimization of the protein structure was performed by the zero-directive of the molecular dynamics package *DL-PROTEIN 2.1* [18, 19]. We used the *GROMOS* [20] force field with the set of parameters denoted '37c' (united atoms), and the water molecules were represented by means of the SPC/E model [21]. All bond lengths were kept fixed over time by using the *SHAKE* iterative procedure [22]. Simulation box sizes were as follows: $48 \times 58 \times 56$ for catalytic and $115 \times 85 \times 65$ for multidomain MMP2. Electrostatic interactions were computed using the Ewald summation method [16] with the smooth particle mesh Ewald (SPME) method [23] to compute the reciprocal part of the Ewald sum. The Ewald method has been used with an α switching parameter of 0.32008 Å^{-1} for the catalytic domain and of 0.3581 Å^{-1} for the multi-domain protein. The SPME method has been used with the following grid points: $35\times40\times40$ for catalytic and $95\times90\times85$ for the multidomain protein, and with a spline order of eight for both systems. The α and SMPE grid values were determined manually, according to the procedure indicated by the *DL-PROTEIN 2.1* handbook [18, 19], in order to achieve the best convergence of the SMPE algorithm. All non-bonding interaction terms have been cut off beyond a distance of 9 Å, with a shifted potential van der Waals interaction further smoothed by a polynomial switching function in the range of 0.5 Å before the cutoff.

Structural analysis

To assess the reliability of the calculations, and as a measure of structural stability, we have carried out a series of standard tests on geometrical properties of the multi-domain metalloproteinase and of its catalytic domain. The secondary structure and other structural parameters have been monitored by running the *DSSP* program [24] that allows to measure, during the entire trajectory, the following features: solvent accessible surface, number of backbone hydrogen bonds, number of residues in unfavorable regions of the Ramachandran plot (i.e. with strained φ/ψ combinations), and the secondary structure assumed by each amino acid. The time dependence of the Root Mean Square Deviation (RMSD) from the starting structure has been calculated after removing the global translations and rotations [25]. The average amino acid backbone Root Mean Square Fluctuation (RMSF) has been calculated from MD simulation:

$$\text{RMSF} = \sqrt{\langle (\Delta r)^2 \rangle} \tag{1}$$

which is related to the experimental B factors, obtained from X-ray diffraction through the following equation:

$$\sqrt{\langle (\Delta r)^2 \rangle} = \sqrt{\frac{3B}{8\pi^2}} \tag{2}$$

The dynamic behaviors of the two systems in the MD simulation have been analyzed by using the dynamic cross-correlation map to yield information about possible correlated motions [26, 27]. Correlated motions may occur between residues belonging to the same or to different domains. The extent of correlated motions between residues is indicated by the magnitude of the corresponding correlation coefficient between their C_α atoms. The cross-correlation coefficient for the displacement of each pair of C_α atoms i and j, which ranges between -1 and 1, is given by:

$$Cij = \frac{\langle \Delta \mathrm{r}_i \cdot \Delta \mathrm{r}_j \rangle}{\sqrt{\langle \Delta r_i^2 \rangle \langle \Delta r_j^2 \rangle}} \tag{3}$$

where $\Delta \mathrm{r}_i$ is the displacement from the mean position of the ith atom and the symbol $\langle\rangle$ represents the

time average over the whole trajectory. The sign of the cross-correlation coefficient indicates whether the displacement is in phase (positive) or in opposition of phase (negative).

The standard Solvent Accessible Surface (SAS), defined as the area traced out by the center of a probe sphere having a radius r = 1.4 Å that represents a solvent molecule rolled over the van der Waals surface of the molecule, has been calculated for each protein atom according to the Lee and Richards algorithm [28].

Volume of the external crevices was measured using the program *SURFNET* [29]. The time evolution of the protein volume has been iteratively calculated using the program *VOLUME* [30, 31]. This program determines the volume of a polyhedron surrounding each atom in a protein when the polyhedral faces are determined by one of three procedures based on the Voronoi construction. The Voronoi procedure [32] is a geometrical construction that assigns a polyhedron, with well-defined volume and surface, to each point of a discrete set. A Voronoi polyhedron can be seen as the generalization of the Wigner-Seitz cell [33] for a non-periodic set of points. The Voronoi polyhedra, like Wigner-Seitz cells, are primitive cells in the sense that they fill all space without overlapping or leaving voids. Unlike the Wigner-Seitz cell, the Voronoi polyhedron has no translational symmetry. A subset of points can be a residue, a domain of the secondary structure, a subunit or the whole protein. Its volume is just the sum of the volumes associated to each site of the subset.

Results and discussion

Overall properties of simulated systems

A regular sampling of the proteins around their equilibrium configurations is assumed, based on the fact that the structural parameters chosen as stability test in the simulations showed constant values. RMSD of backbone atoms is set around 3.0 Å in both simulated systems, and content of α and β structures is stationary, indicating conservation of the overall 3D-fold for both multi-domain protein and catalytic domain systems. In Figure 2, we report the RMSF values for the multi-domain MMP2 from 1CK7 simulation (black circles). Experimental RMSF values for inactive multidomain protein MMP2, obtained through Eq. 2 from the B factors in the crystal structure of MMP2 (1CK7 pdb [9]), are also indicated in Figure 2 (light grey diamond).

Comparison between the two sets of values shows remarkable differences in the fibronectin domain only (Figure 2a), suggesting that propeptide removal activates this domain. The RMSF values for residues belonging to the catalytic (Figure 2a) and linker (Figure 2b, first 10 residues) domains of the MMP2 active form are rather stable. More intense fluctuations occur in fibronectin (residue range 111–285 in Figure 2a) and hemopexin (Figure 2b) domains. In these two domains, a few residues show large Root Mean Square Fluctuations (ranges 140–150, 254–264 and 270–280 in Figure 2a; and range 134–143 in Figure 2b). These mobile residues belong to four loops. Two of them are twin loops in fibronectin modules 1 and 3 (see crosses in Figure 1). The most mobile of the two (range 254–264 in Figure 2a) is localized in fibronectin module 3 and faces the hemopexin domain. In the X-ray structure of proMMP2, this loop interacts with one end of the propeptide (from Pro31 to Pro43 [9]), suggesting that removal of propeptide may set this loop free. Indeed, the entire fibronectin domain seems to gain more flexibility upon removal of propeptide, as already observed above (Figure 2a). The third most mobile loop (residue range 270–280 in Figure 2a) is situated in the fibronectin module 3, and corresponds to a loop with one helix turn in the middle (see open circle in Figure 1). The fourth most mobile loop is localized in blade III of hemopexin domain (see star in Figure 1), which is one of the two regions recognized by the tissue inhibitor TIMP2 in the binding mode that does not involve the inhibition site of TIMP2 and the catalytic site of MMP2 [4]. In detail, this is the area where the C-tail of tissue inhibitor TIMP2 binds to proMMP2, as shown by the X-ray of the TIMP2–proMMP2 complex [4].

Comparison between the experimental RMSF values, obtained through Eq. 2 from the B factors in the crystal structure of catalytic domain (1QIB pdb [5]) and the corresponding RMSF values (Eq. 1) does not show any significant difference.

Characterization of volume accessibility from 1CK7 simulation

In the X-ray diffraction of the multi-domain protein (1CK7 pdb), the catalytic domain has an accessible volume of 21,728 $Å^3$. This value slightly expands during the simulation, being its average value (22,249 ± 546 $Å^3$). Because of flexibility, some clefts and cavities observed in the X-ray structure disappear from the protein surface during molecular dynamics simu-

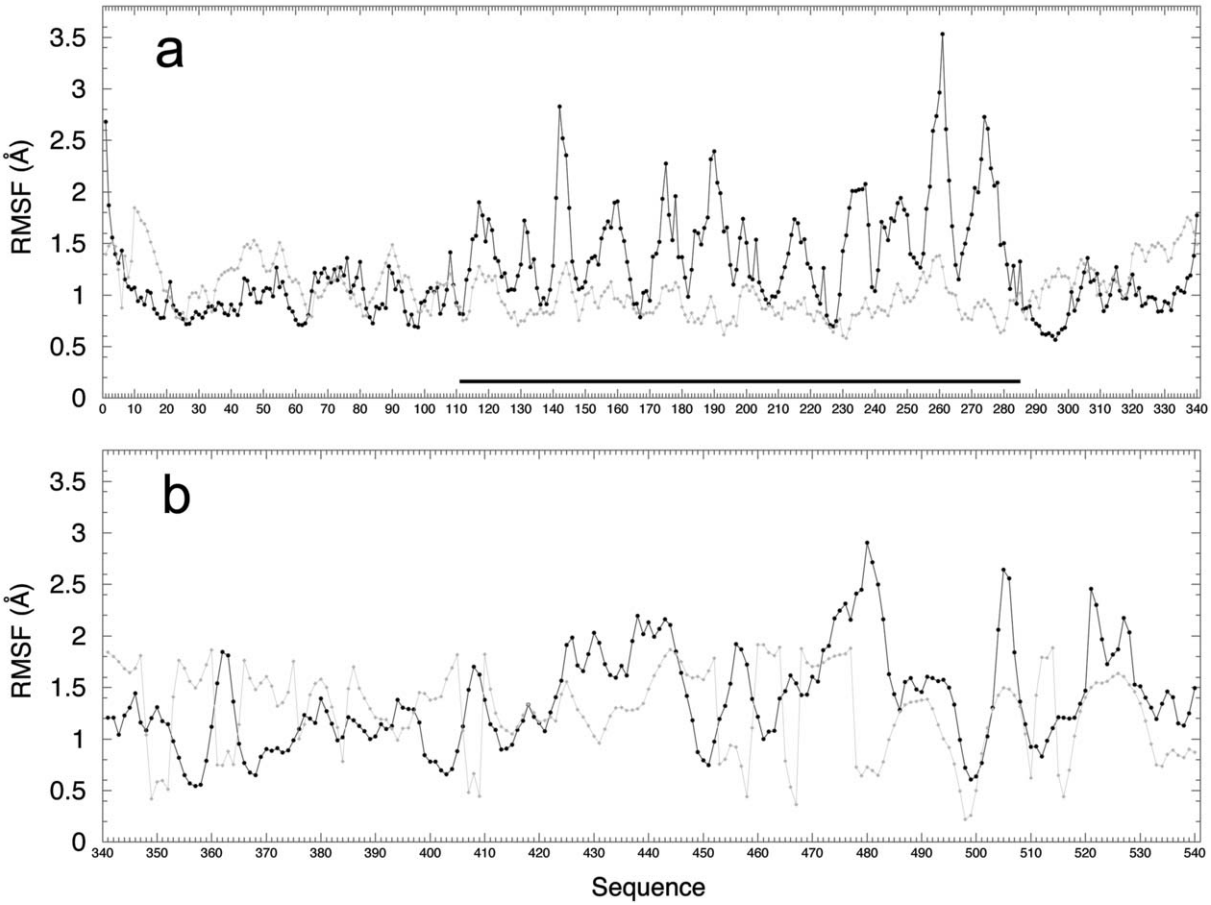

Figure 2. RMSF values for active multi-domain protein MMP2 (black circles) and experimental RMSF for inactive MMP2 (light grey diamond): (a) catalytic and fibronectin (residue range: 111–285) domains; (b) linker (first ten residues) and hemopexin domains.

lations. In the multi-domain protein, one large void (C1 in Figure 1), which is localized across catalytic, fibronectin and hemopexin domains, conserves its volume over the entire 1.26 ns trajectory. In the X-ray structure of proMMP2, this large cavity would correspond to an area of MMP2 covered by one portion of the propeptide, namely two strings of propeptide residues, from Asp40 to Thr56 and from Met97 to Arg98 [9].

A second cavity (C2 in Figure 1), smaller but stable and highly localized, can be identified in the dynamic structures of multi-domain MMP2. This hydrophobic crevice is localized across the two moieties of the catalytic domain, in a region surrounding the catalytic zinc. Its pavement is formed by one surface of α-helix H2, one end of the nearest β-sheet and part of the S1′ loop. In detail, this crevice is lined by seventeen residues identified (pdb 1CK7 sequence number [9]) as follows: Gly189-Leu190-Leu191-Ala192-His193 from a loop contiguous to the β-sheet near α-helix H2 and the first portion of the β-sheet; Tyr395, from the loop entering α-helix H2, and Leu399-Val400-His403-Glu404-His407, localized in α-helix H2; His413, in the long loop connecting α-helix H2 to the S1′ loop; Leu420-Ala421-Pro422-Ile423-Tyr424, localized in the beginning of the long S1′ loop. Among the four histidines mentioned above, His193 coordinates the structural zinc ion, while the other three coordinate the catalytic one. In the crystallographic structure, the same residues form a cavity, smaller than the one observed in the simulation, which is filled by a string of propeptide residues, from Lys99 to Pro105 [9].

Characterization of volume accessibility from 1QIB simulation

The catalytic domain has an accessible volume of 22,706 $Å^3$ in the X-ray diffraction structure

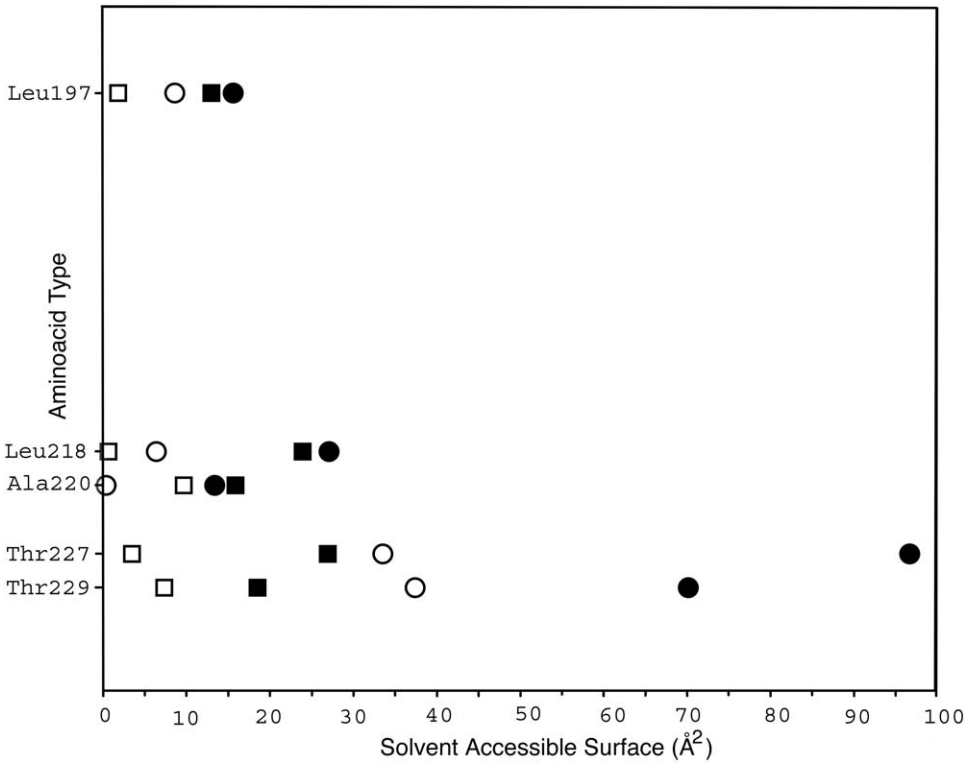

Figure 3. Solvent accessible surface (SAS) for residues shaping crystallographic pocket S1′, calculated in X-ray structures (solid symbols) and averaged over all molecular dynamics trajectories for the two systems (open symbols): multi-domain protein (squares) and catalytic domain separated (circles). Residue sequence numbers as in 1QIB pdb [5].

(1QIB pdb). During the simulation, it reaches an average accessible volume of 21,583 $\pm$ 174 Å^3. Analysis of the simulation enabled us to identify a highly localized gap that partially corresponds to the hydrophobic crevice described above (C2) for multi-domain MMP2, although here the cavity is less defined. In addition to the ones already mentioned, two more residues line its floor, Ala165 and Phe180 (pdb 1QIB sequence number [5]). This crevice is partially observed in the corresponding crystallographic structure.

Several non-localized small voids are also identified in both simulated systems. Most of them are scattered around the protein surface, and do not maintain their positions and volumes over the trajectory. On the other hand, the pockets observed in the X-ray diffraction structure of MMP2 catalytic domain close up during molecular dynamics simulations, including both the shallower ones (S2′, S3′, S1, S2, and S3) and the deep hydrophobic S1′ one considered a major determinant in recognition of substrate [1, 3].

Flexibility and crystallographic pocket S1′ from both simulations

Solvent accessible surface for residues that shape the S1′ pocket is depicted in Figure 3. Solvent accessible surfaces per residue averaged over all molecular dynamics trajectories are indicated by open squares (multi-domain protein, 1CK7 simulation) and circles (catalytic domain, 1QIB simulation), while solvent accessibilities in X-ray structures are represented by corresponding solid symbols. Threonines 227 and 229 (pdb 1QIB sequence number [5]), or Thr426 and Thr428 (pdb 1CK7 sequence number [9]), are arranged in the same way in both crystallographic structures; Thr227 is fully exposed to the solvent, while the oxydryl group of Thr229 points towards the inside of pocket S1′. However, due to the presence of hemopexin domain in the 1CK7, they are less exposed to solvent than in 1QIB. The solvent accessible surface of these threonines, which are localized near the bottom of the pocket S1′, remarkably decreases during molecular dynamics in both systems. In 1QIB simulation, Thr227 is completely buried and its oxy-

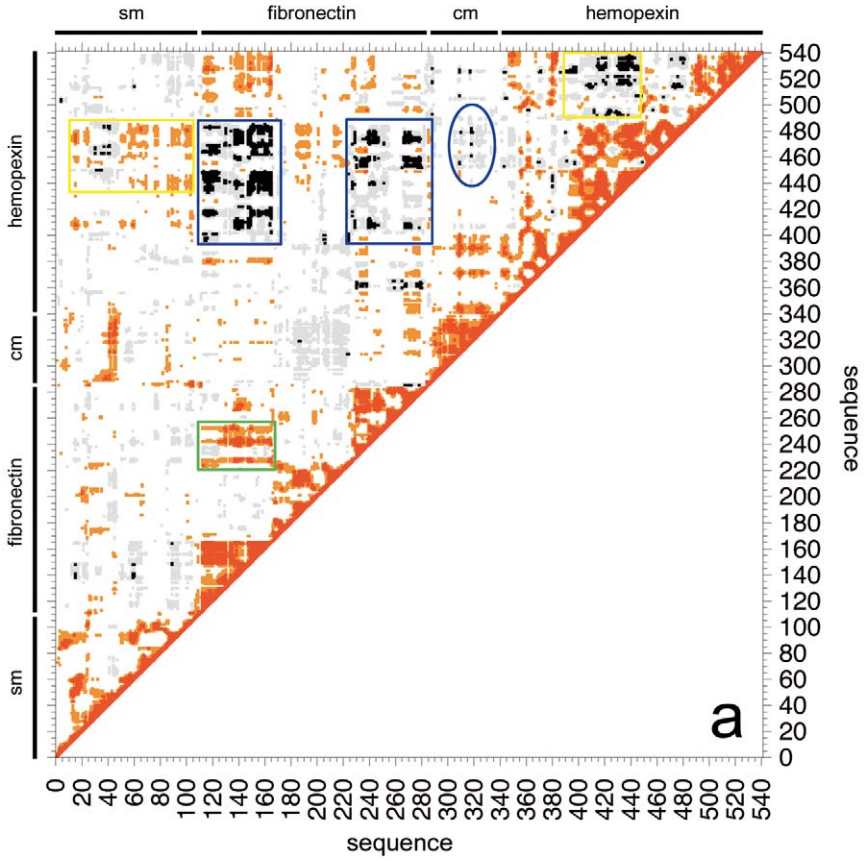

Figure 4. Dynamic cross-correlations matrix of Cα atom displacement for multi-domain MMP2 (a) and for catalytic domain (b), from 1CK7 simulation. In both maps, color code is as follows: orange and red colors represent in phase motion correlations, while gray and black colors represent the negative ones, i.e. correlation of motions in opposition of phase. Gray and orange squares represent medium correlation ($0.4 < |c_{ij}| = 0.6$); red and black squares represent strong correlations ($0.6 < |c_{ij}| = 1.0$ values). Correlations having $|c_{ij}|$ coefficients below 0.4 are not represented. Cross-correlation coefficients, c_{ij}, are defined as in Eq. 3. Thick bars and labels at the edge of both maps indicate domains and sub domains. sm stands for structural moiety of catalytic domain, residue range (1–110); cm stands for catalytic moiety of catalytic domain, residue range (286–342).

dryl group forms a hydrogen bond with the carbonyl group of Ala220; while the side chain of Thr229 is partially buried. In the 1CK7 simulation, the hydrophobic groups pack together, diminishing their solvent exposed surfaces. Other residues that line the cavity of S1′ crystallographic pocket – Leu197, Leu218 and Ala220 (pdb 1QIB sequence number [5]) – have their solvent accessibility further diminished during molecular dynamics. Among these last residues, which are near the mouth of the pocket, Leu218 is the most mobile (calculated RMSF for active MMP2: 0.84; experimental RMSF for inactive MMP2: 1.13), while Leu197, which belongs to α-helix H2, is the less mobile one (calculated RMSF value: 0.71; experimental RMSF for inactive MMP2: 0.96). Conformational changes of these two leucines affect the internal volume of the pocket S1′. Even in the crystal structures, it appears that Leu218, the most mobile one, can determine the shape of the S1′ pocket, since its CD1 and CD2 atoms point towards the pocket interior in the multi-domain structure, reducing its internal volume.

Along with the analysis of volumes, these observations indicate hydrophobic collapse of S1′, and do not support the view of a self-structured S1′ pocket. No crystal structure of S1′ inhibitor bound to the full multidomain protein MMP2 is currently available. So the extent of eventually induced conformational changes upon binding cannot yet be assessed.

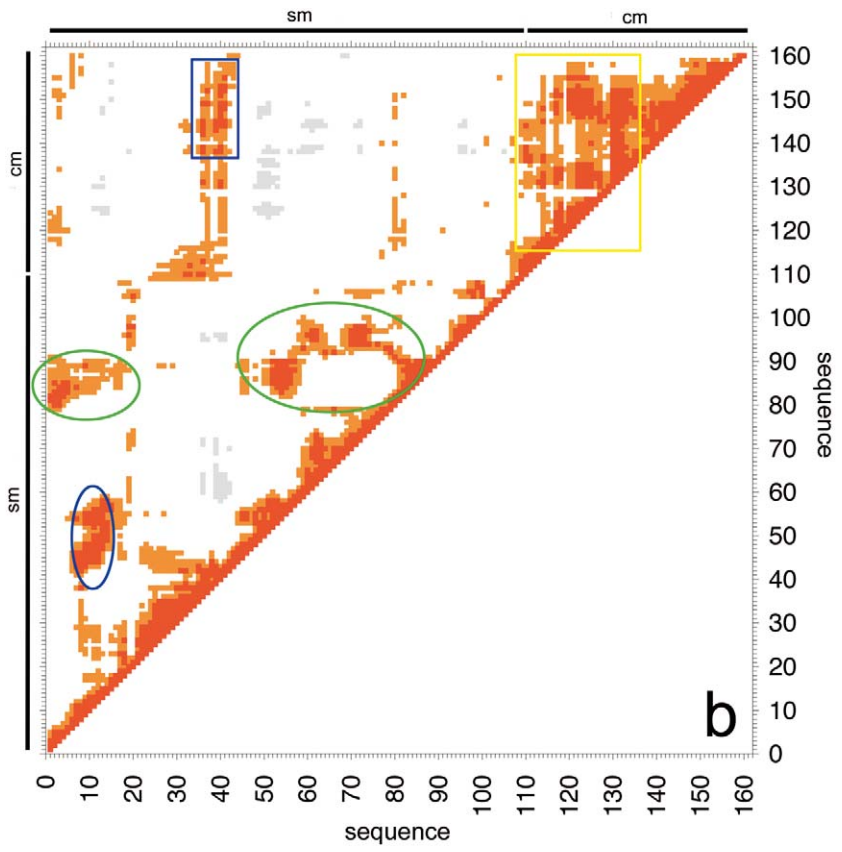

Figure 4. (Continued).

Cross-correlations maps and long range interplay from 1CK7 simulation

Internal correlations, from mild to intense, appear within each of the MMP2 domains. The most interesting feature is the occurrence of correlations within the catalytic domain in multi-domain protein, due to the presence of hemopexin and fibronectin domains. This feature will be discussed in a separate section. We analyzed the dynamic cross-correlation matrices in multi-domain MMP2 protein, reported in Figure 4a, in order to elucidate how the domains are related to each other. For a 3D-view of the structures mentioned in the following, refer to Figure 1.

Two regions of intense cross-correlation appear in Figure 4a, at the crossing of residue range (400–485) with residue ranges (111–165) and (226–285). The former corresponds to structural regions of hemopexin domain (blade II and a large portion of blade III); the latter ones to fibronectin modules 3 and 1, respectively. These areas, highlighted by two blue rectangles in Figure 4a, are spatially adjacent and their relative motion is in opposition of phase, as testified by the negative signs of cross-correlation coefficients (gray-black color in Figure 4a). Fibronectin modules 1 and 3 are also cross-correlated to each other with motions in phase (green rectangle frame in Figure 4a), suggesting that they act as one functional module. Indeed, this is the only interplay among fibronectin modules. Besides, hemopexin regions blade II and III are apparently intensely cross-correlated to hemopexin blade IV with motion in opposition of phase (Figure 4a, top right yellow rectangle).

Among less intensely cross-correlated regions, we highlight with an oval frame the ones occurring between the catalytic moiety (residue range: 286–342 in Figure 4a; purple ribbon in Figure 1) and the portion of hemopexin blade III already mentioned above (in Figure 4a, residue range: 451–485; light blue ribbon in Figure 1). In the catalytic domain, the blue oval of Figure 4a corresponds to a large portion of the catalytic moiety, starting from α-helix H3 and leading, through the long S1′ loop, to the lowest portion of α-helix H2 (see Figure 1). Correlated motions involving

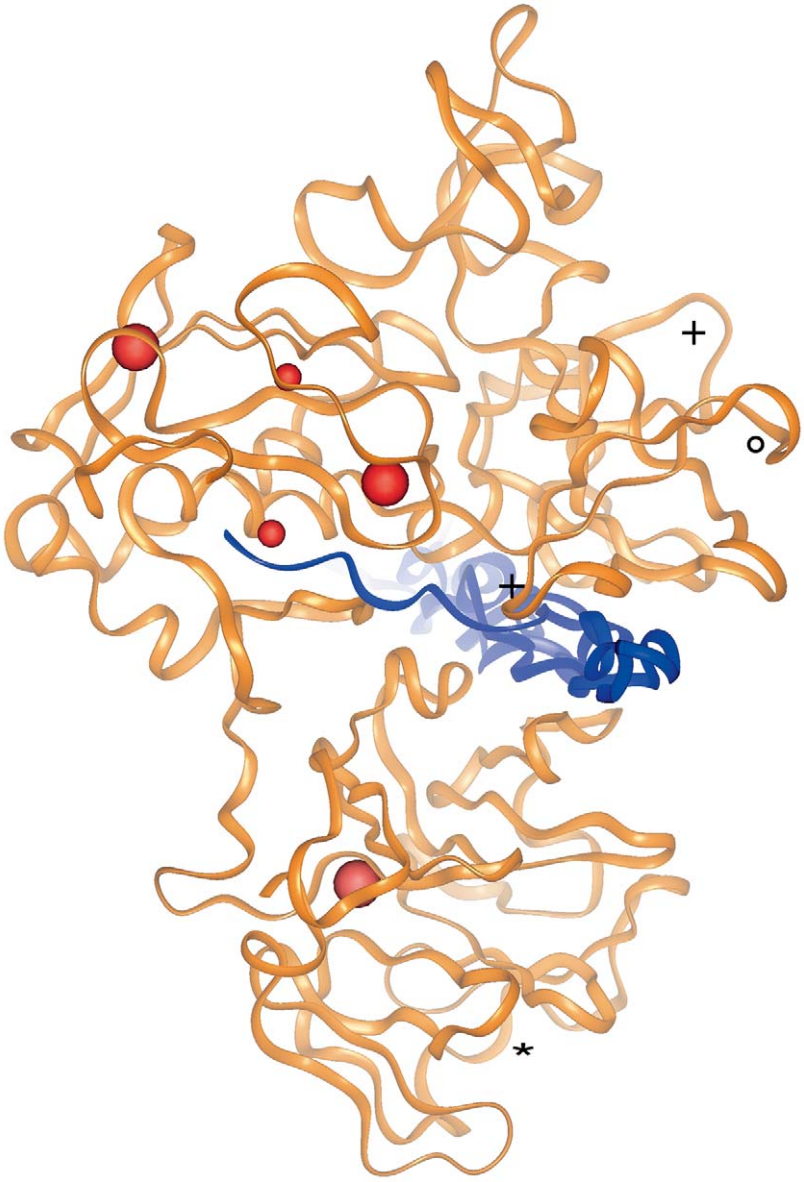

Figure 5. Ribbon model of MMP2-collagen interaction. Orientation of protein approximately as in Figure 1. Triple helix collagen fiber (blue ribbon) is accommodated in the largest cavity across domains (C1 in Figure 1) of MMP2 protein (orange ribbon), while one collagen fibril fits into the cleft near catalytic zinc (C2 in Figure 1). Red spheres of different diameter represent calcium (larger) and zinc (smaller) ions. Most mobile loops are indicated by crosses, open circles and star. The star also labels one of the two binding regions of TIMP2 C-tail [4].

the MMP2 catalytic moiety are in opposition of phase (gray-black color in Figure 4a). On the other hand, the structural moiety and the hemopexin domain regions II and III display motions both in-phase and out-of-phase (Figure 4a, top left yellow rectangle).

Coupling among the multi-domain MMP2 highlights a non-trivial long-range interaction between regions far apart both in sequence and in space. This interplay comprises the S1′ loop in the catalytic domain, suggesting that correlated movements regulate the access to the S1′ pocket, which otherwise is compressed or twisted. Besides, the portion of hemopexin blade III (light blue ribbon in Figure 1) which is correlated to fibronectin modules 1 and 3 is also contiguous – both in sequence and in space – to the string of blade III residues that contact the C-terminus of tissue

inhibitor TIMP2 in the complex with proMMP2 [4] (star in Figure 1). This suggests that an event which occurs relatively far away from the catalytic domain – such as, for example, binding of TIMP2 C-terminus to blade III of MMP2 hemopexin domain – may directly affect MMP2 activation or substrate recognition. Moreover, a number of exosites – secondary specificity sites involved in metalloproteinase regulatory events – are found in hemopexin and fibronectin domains of MMP2 [3], which suggests that the observed inter-domain correlations may have several functional roles.

Cross-correlations within catalytic domain from both simulations

No internal correlations are observed in the map relevant to the isolated catalytic domain system (1QIB simulation), indicating that the presence of hemopexin and fibronectin domains is necessary in order to induce specific motions within the catalytic domain. The cross-correlation matrix of the catalytic domain in multi-domain MMP2 (1CK7 simulation) shown in Figure 4b, is extracted from the multi-domain map of Figure 4a in order to highlight these emerging correlations. In Figure 4b, residue range 1–110 corresponds to the structural moiety, sm, and residue range 111–161 to the catalytic one, cm. We observe several cross-correlated areas internal to each of the two moieties of MMP2 catalytic domain, while coupling between the two moieties is provided by cross-correlation between one portion of helix H1 (residue range: 38–43) and both the S1′ loop and contiguous helix H3 (residue range: 137–158). This region is framed by a blue rectangle in Figure 4b. All interplays internal to MMP2 catalytic domain occur with in-phase motions (orange-red color in Figure 4b). In the catalytic moiety, all structural elements seem to be coupled to each other to some degree (top right yellow rectangle). Within the structural moiety, instead, we identify three contiguous cross-correlated regions. The one framed by the blue oval in Figure 4b corresponds to one end of the longest α-helix (H1) mentioned above, the following loop and contiguous β-sheet (residue range 40–60), which are cross-correlated to part of the loop connecting 8 to 12. The left green oval in Figure 4b involves the long β-sheet near the catalytic zinc and its preceding calcium loop (residue range 80–90 in Figure 4b), which are correlated to the initial part of the loop mentioned above. The large green oval in Figure 4b frames three separated small spots, approximately at the crossing between the previously mentioned range of residues (80–90) and the residues comprised in range 52–74. It is composed by several contiguous structures (the loop already mentioned for the largest spot, followed by one long β-sheet and the consecutive long loop).

The flexibility of the multi-domain MMP2 protein indicates that there is room for long-range interactions within MMP2. Moreover, this flexibility is necessary in order to induce specific motion in the catalytic domain, suggesting that binding and processing of substrate may involve several inter/intra domain motions.

A model for MMP2/collagen interaction

The occurrence of inter/intra-domain motions and the identification of two cavities during the 1CK7 molecular dynamics simulation led us to build a model for the complex of multi-domain MMP2 and a collagen fiber. The triple-helix collagen fiber fits into the largest cavity (C1), with a portion of one fibril, unwound and bent, which can be accommodated in the smaller hydrophobic crevice (C2), as shown in Figure 5. Diameter of the collagen fiber is about 12 Å, whereas its length is 70 Å. The bent fibril accounts for 25 Å – approximately the distance from the one end connected to collagen to the other end near the catalytic zinc ion. A large portion of the collagen fiber body contacts fibronectin modules 1 and 3 (bottom surface of the modules), and a smaller one the hemopexin domain (upper surface of blade II). Thus, it is conceivable that interdomain motions – the observed interplay between hemopexin and fibronectin domains – could be responsible for the unwinding of the triple-helix collagen fiber. Once partially unfolded, one collagen helix may bind to the catalytic crevice for proteolysis.

Conclusions

The unwinding mechanism for collagen fibers involves both fibronectin and hemopexin domains, which are also important for inducing specific internal motions in the catalytic domain. Our study suggests that propeptide removal from MMP2 induces correlated motions between hemopexin and fibronectin. The activity of multidomain MMP2 implicates several steps in which the catalytic domain operates in combination with at least one of the other domains [10, 11]. A two-step mechanism for collagen cleavage

by MMP2 has been recently proposed, the first one involving both catalytic and hemopexin domains, the second one promoted by the fibronectin domain only [11]. On the other hand, it has been also suggested that hemopexin domain is not involved in collagen processing by MMP2 (see [3] for literature). Thus, according to a recent model (the 'molecular tectonics' [3]), independent modular movements in fibronectin domain facilitate penetration of the gelatinase in collagen fiber, with no involvement of hemopexin domain. Although our model cannot account for a multi-step dynamical process, our simulation data are compatible with the two-step mechanism [11]. However, our cross-correlations map does not support the molecular tectonics hypothesis [3], which suggests 'interdigited' movements for fibronectin modules, because the fibronectin modules 1 and 3 are cross-correlated in phase independently of fluctuations of fibronectin module 2.

We emphasize that the dynamically identified crevice, which surrounds the catalytic zinc (C2), is lined by both hydrophobic and hydrogen bond acceptor/donor atoms, is large enough to accommodate small inhibitors, and it might be exploited for selective inhibitor design. A recent experiment [7], based on comparison of X-ray absorption spectra of the inactive, active and inhibited MMP2, proposes that the small synthetic inhibitor SB-3CT directly binds the catalytic zinc by re-establishing the proenzyme structural motif. In the model complex shown in Reference [7], the inhibitor SB-3CT (designed to be highly selective in gelatinases) occupies an area that evolves into the cavity C2 observed during the molecular dynamics simulations.

References

1. Whittaker, M., Floyd, C.D., Brown, P. and Gearing, J.H., Chem. Rev., 99 (1999) 2735.
2. Coussen, L.M., Fingleton, B. and Matrisian, L.M., Science, 295 (2002) 2387.
3. Overall, C.M., Mol. Biotech., 22 (2002) 51.
4. Morgunova, E., Tuttila, A., Bergmann, U. and Tryggvason, K., Proc. Natl. Acad. Sci. USA, 99 (2002) 7414.
5. Dhanaraj, V., Williams, M.G., Ye, Q.Z., Molina, F., Johnson, L.L., Ortwine, D.F., Pavlosky, A., Rubin, R.J., Skeean, R.W., White, A.D., Humblet. C., Hupe, D.J. and Blundell, T.L., Croatica Chem. Acta, 72 (1999) 575.
6. Hou T., Zhang W. and Xu, X., J. Comput.-Aided Mol. Des., 16 (2002) 27.
7. Kleifeld, O., Kotra, L.P., Gervasi, D.C., Brown, S., Bernardo, M.M., Fridman, R., Mobashery, S. and Sagit, I., J. Biol. Chem., 276 (2001) 17125.
8. Rosenblum, G., Meroueh, S.O., Kleifeld, O., Brown S., Singson, S.P., Fridman R., Mobashery S. and Sagi, I., J. Biol. Chem., 278 (2003) 27009.
9. Morgunova, E., Tuttila, A., Isupov, M., Linqvist, Y., Schneider, G. and Tryggvason, K., Science, 284 (1999) 1667.
10. Sanchez-Lopez, R., Alexander, C.M., Behrendtsen, O., Breathnach, R. and Werb, Z., J. Biol. Chem., 268 (1993) 7238.
11. Patterson, M.L., Atkinsons, S.J., Knauper, V. and Murphy, G., FEBS Lett., 503 (2001) 158.
12. Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N. and Bourne, P.E., Nucleic Acids Res., 28 (2000) 235.
13. Bella, J., Brodsky, B. and Berman, H.M., Structure, 3 (1995) 893.
14. Bella, J., Eaton, M., Brodsky, B. and Berman, H.M., Science, 266 (1994) 75.
15. Steward, J.J.P., Rossi, I., Hu, W.P., Lynch, G.C., Liu, Y.P. and Trulhar, D.G., *MOPAC–version 5.07mn*, University of Minnesota, Minneapolis, MN, 1997.
16. Allen, M.P. and Tildesley, D.J., Computer Simulation of Liquids. Clarendon Press, Oxford, UK, 1987.
17. Melchionna, S. and Ciccotti, G., J. Chem. Phys., 106 (1997) 195.
18. Smith, W. and Forester, T.R., J. Mol. Graph., 14 (1996) 136.
19. Melchionna, S., Luise, A., Venturosi, M. and Cozzini, S., In: Voli M. [Ed], Science and Supercomputing at CINECA - 1997 report. Supercomputing Group CINECA, Bologna, Italy, 1998.
20. van Gunsteren, W.F. and Berendsen, H.J.C., *GROMOS Manual*, University of Groningen, Groningen, The Netherlands, 1987.
21. Berendsen, H.J.C., Grigera, J.R. and Straatsma, T.P., J. Phys. Chem., 91 (1987) 6269.
22. Ryckaert, J.P., Ciccotti, G. and Berendsen, H.J.C., J. Comput. Phys., 23 (1977) 327.
23. Essmann, U., Perera, L., Berkowitz, M.L., Darden, T., Lee, H. and Pedersen, L.G., J. Chem. Phys., 103 (1995) 8577.
24. Kabsch, W. and Sander, C., Biopolymers, 22 (1983) 2577.
25. Kneller, G., Mol. Sim., 7 (1991) 113.
26. McCammon, J.A. and Harvey, S.C., Dynamics of Proteins and Nucleic Acids, Cambridge University Press, London, 1987.
27. Chillemi, G., Fiorani, P., Benedetti, P. and Desideri, A., Nucleic Acids Res., 31 (2003) 1525.
28. Lee, B. and Richards, F.M., J. Mol. Biol., 55 (1971) 379.
29. Laskowski, R.A., J. Mol. Graph., 13 (1995) 323.
30. Richards, F.M., J. Mol. Biol., 82 (1974) 1.
31. Richards, F.M., Methods Enzymol., 115 (1985) 440.
32. Voronoi, G.F.Z., Reine Angew. Math., 134 (1908) 198.
33. Ashcroft, N.W. and Mermin, N.D., Solid State Physics, Saunders College, Philadelphia, PA, 1976.