Peer-reviewed version available at *Int. J. Mol. Sci.* **2019**, *20*, 4194; doi:10.3390/ijms20174194

Research Article

Conformation and Domain Movement Analysis of Human Matrix Metalloproteinase-2: Role of Associated Zn^{2+} and Ca^{2+} Ions.

Leah Voit-Ostricki[1], Sándor Lovas[2], and Charles R. Watts[1,3],*

[1] Department of Neurosurgery, Mayo Clinic Health System-Franciscan Healthcare in La Crosse, Wisconsin, 54601 USA; voitost.leah@gmail.com
[2] Department of Biomedical Sciences, Creighton University, Omaha, Nebraska, USA, 68178; SandorLovas@creighton.edu
[3] Department of Neurologic Surgery, Mayo Clinic, Rochester, Minnesota, USA, 55905; charles.watts@parknicollet.com

* Correspondence: charles.watts@parknicollet.com; Tel.: +1-952-993-3200, Fax: +1-952-993-7407

Abstract: Matrix Metaloproteinase-2 (MMP-2) is an extracellular Zn^{2+} protease specific to type I and IV collagens. Its expression is associated with several inflammatory, degenerative, and malignant diseases. Conformational properties, domain movements, and interactions between MMP-2 and its associated metal ions were characterized using a 1.0 μs molecular dynamics simulation. Dihedral principle component analysis revealed 10 families of conformations with the greatest degree of variability occurring in the link region connecting the catalytic and hemopexin domains. Dynamics cross correlation analysis indicated domain movements corresponding to opening and closing of the hemopexin domain in relation to the fibronectin and catalytic domains facilitated by the link region. Interaction energies were calculated using the MMPBSA-interaction entropy analysis method and revealed strong binding energies for the catalytic Zn^{2+} ion 1, Ca^{2+} ion 1, and Ca^{2+} ion 3 with significant conformational stability at the binding sites of Zn^{2+} ion 1 and Ca^{2+} ion 1. Ca^{2+} ion 2 diffuses freely away from its crystallographically defined binding site. Zn^{2+} ion 2 plays a minor role in conformational stability of the catalytic domain while Ca^{2+} ion 3 is strongly attracted to the highly electronegative sidechains of the Asp residues around the central β-sheet core of the hemopexin domain.

Keywords: molecular dynamics; matrix metalloproteinase; domain movement; zinc binding protein; calcium binding protein.

---

1. Introduction

Matrix Metaloproteinase-2 (MMP-2) is a 550 amino acid residue extracellular Zn^{2+} protease that degrades type I and IV collagens [1,2]. It is related to a family of 24 known endopeptidases with an active site Zn^{2+} ion. On the basis of evolutionary relationships and the structure of their domains, the family is divided into 4 subfamilies [3-5]. MMP-2 expression is associated with normal physiology as well as several inflammatory, degenerative, and malignant diseases [6-12]. As shown in Figure 1; MMP-2 has three domains, catalytic (Cat), fibronectin (Fib), and hemopexin (Hpx) and five crystallographically assigned divalent cations (two Zn^{2+} and three Ca^{2+}) [13]. The Cat domain (Tyr^{110} through Tyr^{445}) is a conserved matrixin fold consisting of five β-sheets and three α-helices. The Fib domain (Glu^{217} through Gln^{393}) is inserted within the catalytic domain between the β5-sheet and α2-helix and contains three type II fibronectin fingers consisting of two antiparallel β-sheets connected by a short α-helix forming a three

Peer-reviewed version available at Int. J. Mol. Sci. 2019, 20, 4194; doi:10.3390/ijms20174194

prong treble hook-like arrangement. This domain and its arrangement may play a crucial role in substrate binding and presentation to the catalytic site. The Hpx domain (Leu^{461}-Cys^{660}) is a four bladed propeller fold that is partially oriented toward the catalytic domain. This domain binds an endogenous inhibitor TIMP-2, however, its role in enzymatic function is unknown. The Hpx and Cat domains are connected by a 16 amino acid proline rich Lnk region (Gly^{446} through Thr^{460}) which is unresolved in the 1CK7 X-ray crystal structure and has been shown to be highly flexible in other molecular dynamics investigations [14,15].

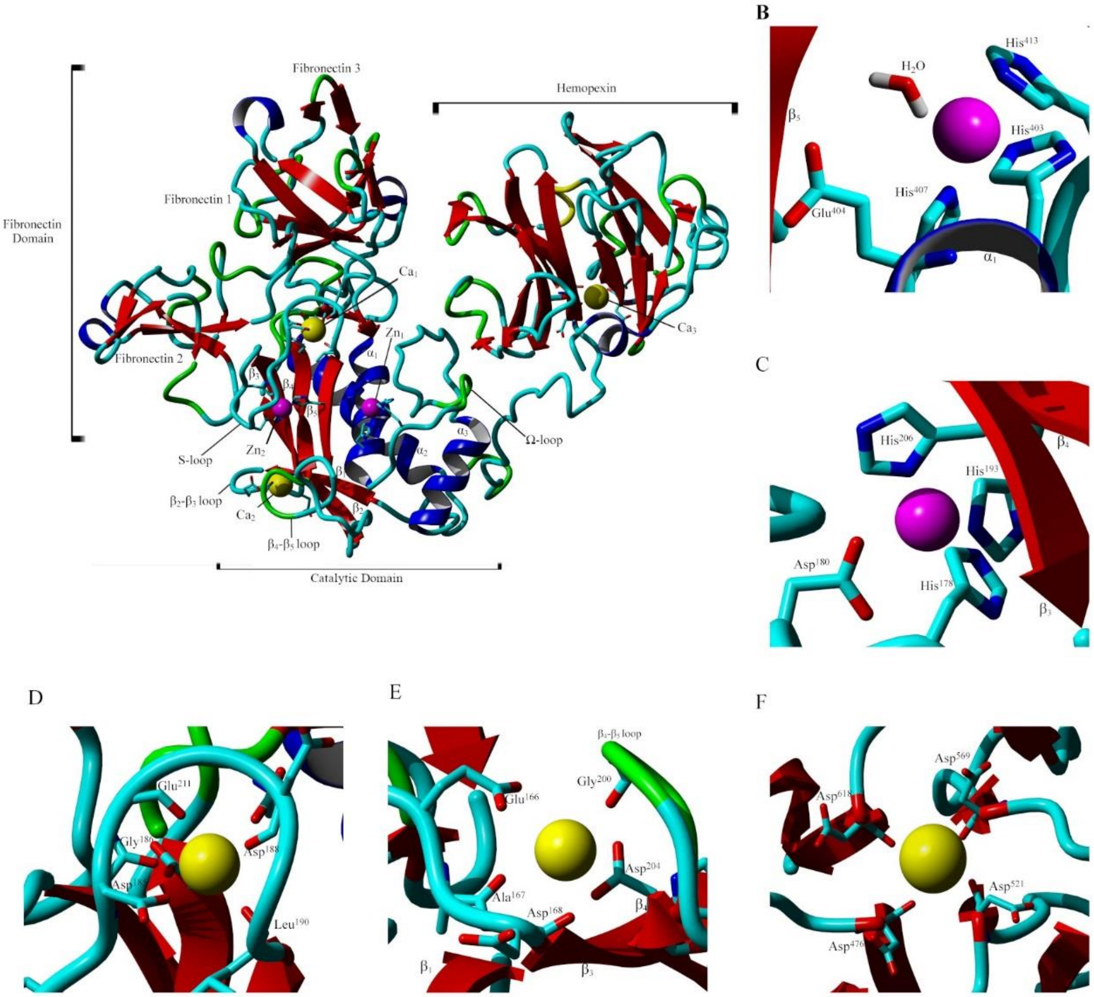


Figure 1. The ribbon diagram of the X-ray crystal structure of 1CK7. The pro-peptide (Pro^{31}-Gln^{109}) region is removed and the unresolved link (Asp^{450}-Thr^{460}) connecting the Cat and Hpx domains built with YASARA [16]. Domains, subdomains, and secondary structural features of the catalytic domain are labeled accordingly (blue, α-helix; red, β-sheet; green, β-turn/bend; and aqua, coil) (**A**). The associated ions are shown as van der Waals radii with Zn^{2+} pink and Ca^{2+} yellow. The catalytic Zn^{2+} ion 1 is bound to His^{403}, Glu^{404}, His^{407}, His^{413}, and catalytic water (**B**). Zn^{2+} ion 2 is bound to His^{178}, Asp^{180}, His^{193}, and His^{206} (**C**). Ca^{2+} ion 1 is bound to Asp^{185}, Gly^{186}, Asp^{188}, Leu^{198}, Asp^{208}, and Glu^{211} (**D**). Ca^{2+} ion 2 is bound to Glu^{166}, Ala^{167}, Asp^{168}, Gly^{200}, Gly^{202}, and Asp^{204} (**E**). Ca $^{2+}$ ion 3 is bound to Asp ^{476}, Asp^{521}, Asp^{569}, and Asp^{618} (**F**).

Table 1. Distances between ions and the interacting MMP-2 residue atoms and associated binding geometry identified from the 1CK7 X-ray crystal structure. Binding geometries were determined for those peptide atoms within 0.35 nm of the divalent cation.[a,b,c]

		Distance / nm				
Protein Atom	**Geometry**	**Zn^{2+} ion 1**	**Zn^{2+} ion 2**	**Ca^{2+} ion 1**	**Ca^{2+} ion 2**	**Ca^{2+} ion 3**
His^{403}:Nε2	tetrahedral	0.23				
His^{407}:Nε2	tetrahedral	0.22				
His^{413}:Nε2	tetrahedral	0.25				
Water:O	tetrahedral	0.25				
Glu^{404}:Oε1		0.83				
Glu^{404}:Oε2		0.76				
His^{178}:Nε2	trigonal planar		0.21			
Asp^{180}:Oδ1			0.35			
Asp^{180}:Oδ2	trigonal planar		0.23			
His^{206}:Nδ1	trigonal planar		0.21			
His^{193}:Nε2	trigonal planar		0.21			
Asp^{185}:C=O	octahedral			0.29		
Asp^{185}:Oδ1				0.38		
Asp^{185}:Oδ2				0.61		
Gly^{186}:C=O	octahedral			0.24		
Asp^{188}:C=O	octahedral			0.26		
Leu^{198}:C=O	octahedral			0.25		
Asp^{208}:Oδ1				0.44		
Asp^{208}:Oδ2	octahedral			0.25		
Glu^{211}:Oε1				0.49		
Glu^{211}:Oε2	octahedral			0.27		
Glu^{166}:Oε1					0.42	
Glu^{166}:Oε2	pentagonal pyramidal				0.32	
Ala^{167}:C=0	pentagonal pyramidal				0.29	
Asp^{168}:C=O	pentagonal pyramidal				0.29	
Gly^{200}:C=O	pentagonal pyramidal				0.27	
Gly^{202}:C=O	pentagonal pyramidal				0.34	
Asp^{204}:Oδ1	pentagonal pyramidal				0.35	
Asp^{204}:Oδ2					0.40	
Asp^{476}:C=O	seesaw					0.25
Asp^{521}:C=O	seesaw					0.28
Asp^{569}:C=O	seesaw					0.27
Asp^{618}:C=O	seesaw					0.27

[a] Zn^{2+} ion 1 is the catalytic ion.

[b] Glu^{404} is critical to catalytic activity.

[c] Location and orientation of the catalytic water was derived from the sidechain of Cys^{102} and the X-ray crystal structure of MMP-13 (PDB ID: 1XUD).

Expanded views of the bound divalent cations to MMP-2 residues are shown in Figures 1B through 1F with the X-ray crystal structure 1CK7 coordination geometry and interatomic metal cation to MMP-2 residue distances given in Table 1 [13,17]. The catalytic Zn^{2+} ion 1 is bound by the conserved MMP extended zinc binding motif [5]:

$$\mathrm{HExxHxxGxxH/D} \tag{1}$$

consisting of the sidechains of His^{403} and His^{407} from the α2-helix, and His^{413} from the Ω-loop (Figure 1B). The structural Zn^{2+} ion 2 is bound in a tetrahedral arrangement involving the sidechains of His^{193} and His^{206} from the β5- and β4-sheets respectively and Asp^{180} and His^{178} of the long S-loop of the Cat domain which is a conserved motif in the MMP family (Figure 1C). Two of the Ca^{2+} ions are bound near the Cat domain with Ca^{2+} ion 1 bound by the sidechains of Asp^{208} and Glu^{211} of the interim loop connecting the Cat and Fib domains and the carbonyl oxygens of Asp^{185}, Gly^{186}, Asp^{188}, and Leu^{190} of the loop connecting

Peer-reviewed version available at *Int. J. Mol. Sci.* **2019**, *20*, 4194; doi:10.3390/ijms20174194

the β3- and β5-sheets (Figure 1D). Ca^{2+} ion 2 is bound by the sidechains of Glu^{166} and carbonyl oxygens of Ala^{167} and Asp^{168} arising from the loop connecting the β1- and β3-sheets, the carbonyl oxygen of Gly^{200} of the loop connecting β5- and β4-sheets, and the sidechain of Asp^{204} arising from the β4-sheet, Figure 1E. The third structural Ca^{2+} ion 3, is bound by the carbonyl oxygens of Asp^{476}, Asp^{521}, Asp^{569}, and Asp^{618} at the edge of the central cavity of the Hpx domain (Figure 1F).

In the present study we examine the domain movements within MMP-2 and evaluate the structural stability of the bound (2 Zn^{2+} and 3 Ca^{2+}) divalent ions using 1.0 µs NPT MD simulations. Protein-metal cation distances and MMPBSA-interaction entropy binding energies (ΔG) were calculated, metal cation hydration evaluated and the sampled conformational space analyzed with dihedral Principle Component Analysis (dPCA) and Dynamic Cross-Correlation Matrix (DCCM) analysis.

2. Results and Discussion

2.1 System Equilibration and Conformational Stability.

In order to evaluate convergence of the trajectory, the configurational entropy of MMP-2 together with the associated divalent ions was calculated as a function of time (Figure S1) [18-20]. After a sharp rise in the configurational entropy over the first 100 ns, the value plateaus prior to 200 ns. Based on these results, we used the 200 ns to 1000 ns portion of the trajectory for our analysis with a sampling frequency of 0.1 ns. The radius of gyration (R_g) and inter-domain center-of mass (COM) distances: Cat-Hpx and Fib-Hpx, were analyzed with k-means clustering and the associated means and standard deviations of each population calculated (Figures S2 and S3) [21,22]. Five different distributions of protein conformation were identified with R_g: 2.65±0.02 nm, 2.70±0.06 nm, 2.80.±0.05 nm, 2.83±0.04 nm, and 3.09±0.13 nm. The COM distances mirror the R_g results, identifying five distributions of Cat-Hpx COM distances: 3.55±0.10 nm, 3.70±0.17 nm, 3.79±0.06 nm, 3.86±0.13 nm, and 4.20±0.23 nm. Five distributions of Fib-Hpx COM distances: 3.70±0.06 nm, 3.84±0.18 nm, 4.32±0.22 nm, 4.48±0.18 nm, and 5.52±0.48 nm, were also identified. The R_g and COM distance data are consistent with inter-domain motions between Cat/Fib and Hpx and the presence of inter-domain motions and the sampling of more extended conformations of MMP-2 in solution compared to the more compact X-ray crystal structure (PDB ID: 1CK7) which has R_g: 2.77 nm, Cat-Hpx COM distance: 3.81 nm, and Fib-Hpx COM distance: 2.00 nm [13-15].

Table 2. Average RMSD of the Cα-trace of MMP-2 as a whole and divided into its individual domains: Cat, Fib, and Hpx with the Lnk region

	All	Cat w/ Fib	Fib	Cat w/o Fib	Hpx	Lnk
RMSD / nm	3.20±0.14	3.34±0.02	3.69±0.03	0.50±0.03	0.34±0.04	0.54±0.08

The average Cα-trace RMSD data for MMP-2 as a whole and divided into its individual domains are given in Table 2. The Cα-trace RMSF with the >50% sampled secondary structure assigned by the DSSP method are shown in Figure 2 [23]. Those regions with defined rigid secondary structure (β-sheets and α-helices) have lower RMSF values compared to more flexible β-turn/bend and coil regions. The greatest degree of conformational variability occurs within the Fib domain while the most stable regions of the protein are within the Cat and Hpx domains. The stability of the Cat and Hpx domains is most likely secondary to the presence of the three long α-helices of the Cat domain and the prominent β-sheets and ordered arrangement of the Hpx domain. Although ordered secondary structure is present within the Fib domain, it consists of three separate subdomains with significant portion of the fold consisting of flexible β-turn/bend and coils. Flexibility within the treble hook arrangement of the Fib domain may play an important role in collagen binding and unraveling [24-27].

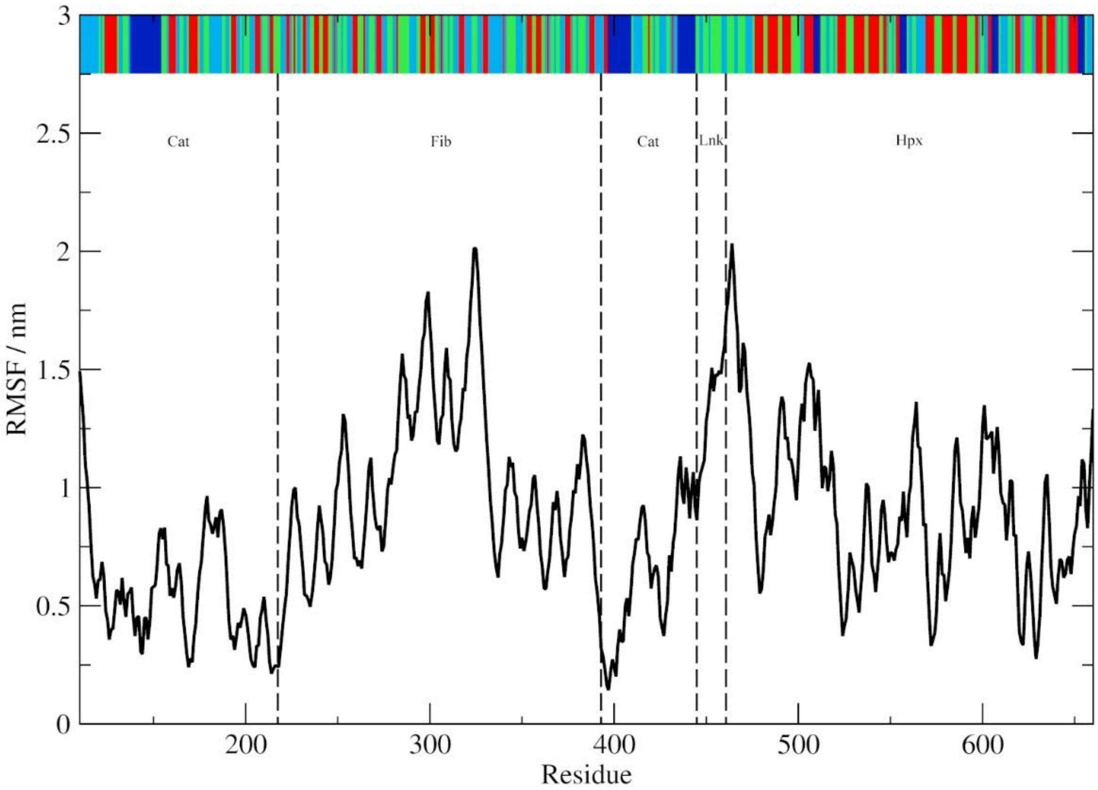
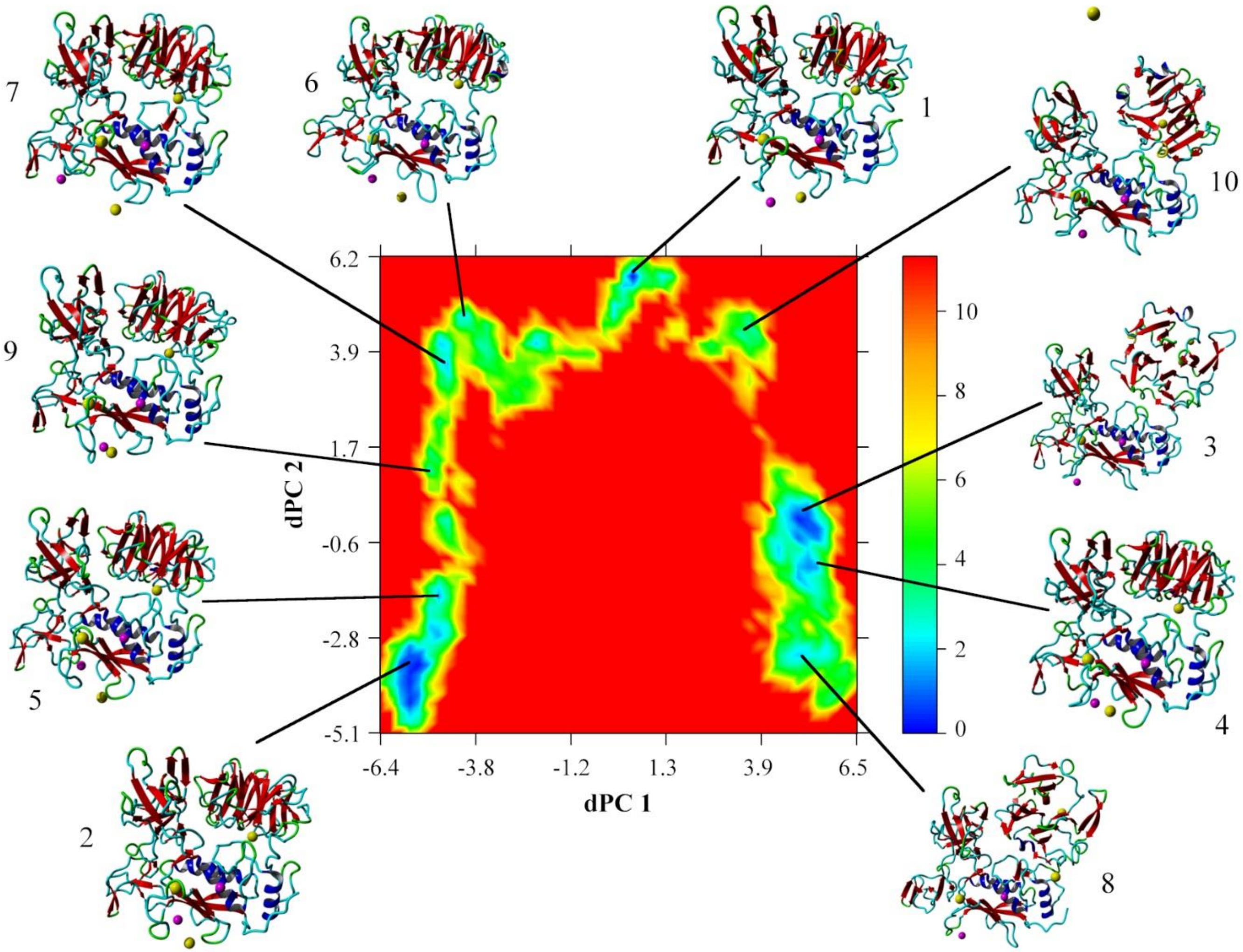


Figure 2. The Cα-trace RMSF. The DSSP assigned secondary structure (sampled >50%) is shown at the top of the graph (blue, α-helix; red, β-sheet; green, β-turn/bend; and aqua, coil). The Cat, Fib and Hpx domains and Lnk regions are demarcated with dotted lines.

2.2 Conformational Analysis.

The free energy landscape created by the first two dihedral principle components (dPC) is shown in Figure 3 with the corresponding lowest energy centroid conformations as determined by k-means clustering. The most conformationally stable region is the Cat domain in which the active site Zn^{2+} ion 1 is bound by His residues of the α1-helix and the Ω-loop (Figures 3 and 4). The structure around Ca^{2+} ion 1 also is stable with the cation bound within a pocket created by the distal portion of the S-loop and the interim loop connecting the Cat and Fib domains. This stability is also confirmed by the low RMSD and RMSF values for the Cat domain and those of Zn^{2+} ion 1 and Ca^{2+} ion1 with both ions staying closely associated with the Cat domain of the protein and approximate to their crystallographically defined positions during simulation (Tables 2 and 3 and Figure 3 and 4). The other associated metal ions of the Cat domain do not share this degree of stability (Table 3). Zn^{2+} ion 2 remains associated with the S-loop but loses contact with the crystallographically demonstrated interactions with the His residues on the β4- and β5-sheets. This may not be unexpected since conformational flexibility within the S-loop region, as was reported previously [28] and may allow for changes in the binding pocket conformation necessary for substrate recognition. Ca^{2+} ion 2 does not remain in close contact with the β1-β3-loop and diffuses out of the binding pocket (Figures 3 and 4). The Hpx associated Ca^{2+} ion 3 remains close to its crystallographically observed binding site (Figures 3 and 4) with a relatively low RMSF (Table 3) but to a lower degree that either Zn^{2+} ion 1 or Ca^{2+} ion 1.

Peer-reviewed version available at Int. J. Mol. Sci. 2019, 20, 4194; doi:10.3390/ijms20174194



137
138 **Figure 3.** Free energy landscape (kJ mol^{-1}) as a function of dPC1 and dPC2; the lowest energy conformations of each family as identified by k-means clustering are shown.
139 Secondary structural motifs and ions are shown (blue, α-helix; red, β-sheet; green, β-turn/bend; aqua, coil; pink, Zn^{2+}; and yellow, Ca^{2+}).

Peer-reviewed version available at Int. J. Mol. Sci. 2019, 20, 4194; doi:10.3390/ijms20174194

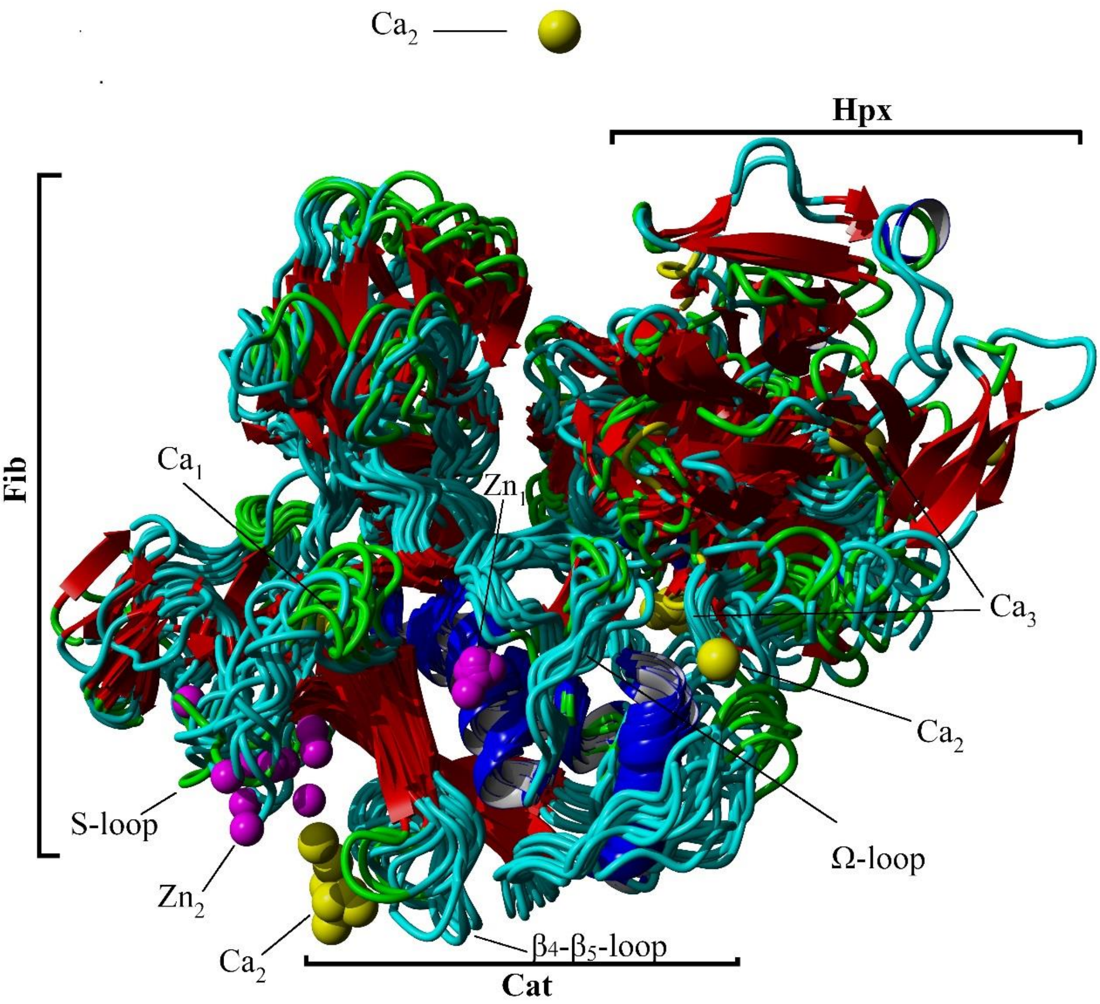


Figure 4. Overlay of 10 cluster centroid structures from the cluster analysis of dPC1 and dPC2. Secondary structural motifs and ions are shown (blue, α-helix; red, β-sheet; green, β-turn/bend; aqua, coil; pink, Zn^{2+}; and yellow, Ca^{2+}).

Table 3. Average RMSF of the associated divalent cations to the Cα-trace of MMP-2

	Zn^{2+} ion 1*	**Zn^{2+} ion 2**	**Ca^{2+} ion 1**	**Ca^{2+} ion 2**	**Ca^{2+} ion 3**
RMSF / nm	0.593	0.882	0.571	2.906	1.062

The majority of the conformational fluctuations within MMP-2 are within the Fib domain (Figures 3 and 4 and Table 2). The three type II Fib subdomains are highly flexible due in part to the large amount of β-turn/bend and coil structure within the subdomains. This degree of flexibility may be important for interactions between MMP-2 and its collagen substrates [24-27]. There is also clearly an inter-domain interaction that occurs between Hpx and the Cat and Fib domains mediated by the Lnk region. The Lnk region acts as a complex hinge allowing the COM distance between the Cat/Fib and Hpx domains to open and close while changing the orientation of the Hpx domain from an edge view to an end on view (Figure 3).

The catalytic Zn^{2+} ion 1 is hydrated with a single water molecule within 0.5 nm (Table 4). This is consistent with its catalytic function and prior computational studies [29,30]. Ca^{2+} ion 1 is the most solvent sequestered of the bound divalent cations. The remaining divalent cations (Zn^{2+} ion 2, Ca^{2+} ion 2 and Ca^{2+} ion 3) have increased solvent accessibility (Table 4).

Peer-reviewed version available at Int. J. Mol. Sci. 2019, 20, 4194; doi:10.3390/ijms20174194

Table 4. Probability (ρ) of a water molecule within 0.5 nm of the associated divalent cations.

	Zn2+ ion 1	Zn2+ ion 2	Ca2+ ion 1	Ca2+ ion 2	Ca2+ ion 3
ϱ(g(r))	0.056	0.071	0.035	0.063	0.067

2.3 Domain Movement Analysis.

DCCM (Figure 5) demonstrated that motions within Cat, Fib and Hpx domains and Lnk region are for the most part highly correlated with a few exceptions. The first five principle components (PC) account for total of 93.7% of the total domain motions (Figure 6). The plots of the Cα-trace RMSF of the projection of the trajectory onto the first five eigenvectors are shown (Figure S4). For PC1, the majority of the contributions arises from the first two subdomains of Fib, the Hpx domains and Lnk region. PC2, has significant motions in the third subdomain of Fib and Hpx. There are also anti-correlated motions with the Cat domain involving the α1-helix and β1- through β5-sheets that contain the active site. Motions along PC3 through PC5 represent minor fluctuations within the domains and global conformation. The distal portion of the Cat domain which contains the active site has correlated motions with the first and third subdomains of the Fib domain and the first and second blades of the Hpx domain. There are also anti-correlated motions between the active site on the Cat domain and the second subdomain of the Fib domain and the third and fourth blades of the Hpx domain.

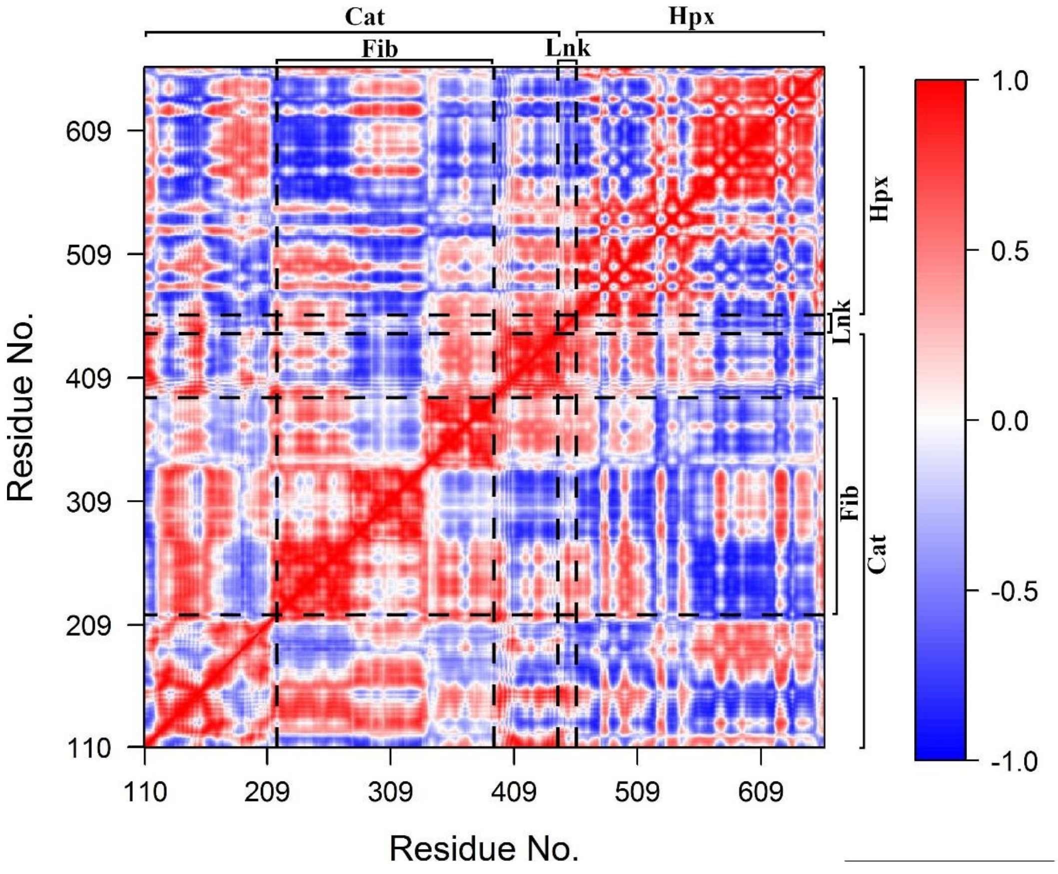
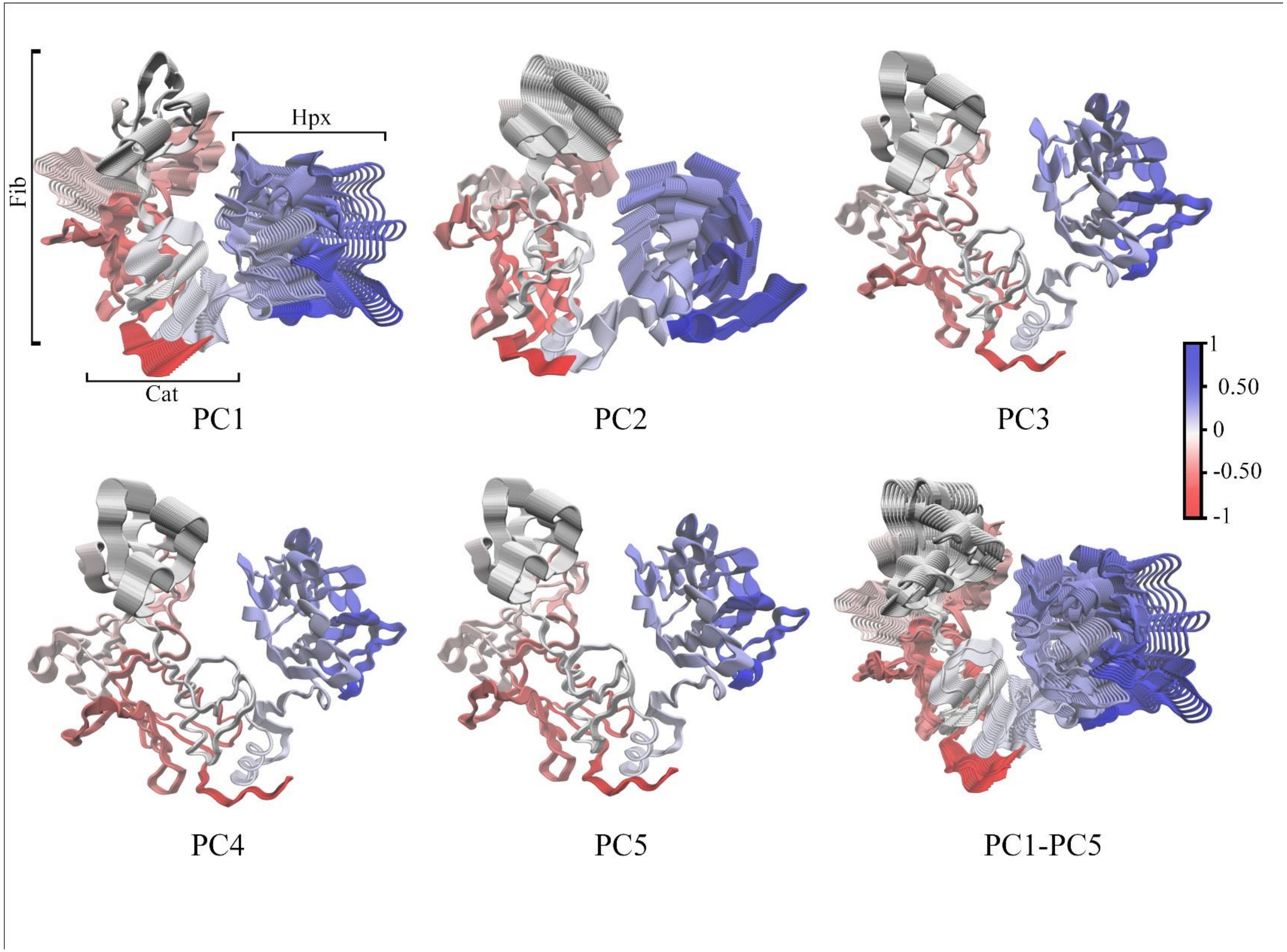


Figure 5. Dynamic Cross Correlation Matrix. Values range from -1 (complete anti-correlation) to +1 (complete correlation). The Cat, Fib, and Hpx domains and Lnk regions are demarcated with dashed lines.

Peer-reviewed version available at Int. J. Mol. Sci. 2019, 20, 4194; doi:10.3390/ijms20174194



177
178 **Figure 6.** Cα-trace of PC1 through PC5 and an overlay of all principle components PC1-PC5. Correlated domain movements are indicated in blue and anti-correlated are in
179 red.

2.4 Protein-Metal Ion Interaction Energies.

MMP-2 has a high affinity for the bound divalent cations (Table 5) with major contribution from the electrostatic interactions. The solvation term ΔG_{polar} is unfavorable particularly for the catalytic Zn^{2+} ion 1, the structural Ca^{2+} ion 1, and Ca^{2+} ion 3 and the $\Delta G_{non\text{-}polar}$ is only weakly favorable. For the weakly bound Zn^{2+} ion 2 and the freely diffusing Ca^{2+} ion 2, the ΔG_{polar} is significantly smaller. The entropic contributions to MMP-2 divalent cation interaction energies are very small but positive indicating decreased system entropy with binding of the metal ions to the peptide. The entropic term is lowest for those ions that for those metal ions (Ca^{2+} ion 2 and Ca^{2+} ion 3) that either diffuse freely away from their crystallographically determined binding sites or do not form close associations and stable binding geometries with the electronegative backbone and sidechain atoms of MMP-2 [31].

Table 5. Binding energies between MMP-2 and the associated divalent cations as determined by the MMPBSA and interaction entropy methods with its associated components.

	Mean ± standard deviation / kJ mol^{-1}						
Ion(s)	**ΔE_{vdw}**	**ΔE_{elec}**	**ΔG_{polar}**	**$\Delta G_{non\text{-}polar}$**	**$-T\Delta S$**	**$\Delta E_{binding}$**	**$\Delta G_{binding}$**
All ions	95.04 ± 2.92	-9264.45 ± 194.34	3142.58 ± 81.6	-3.95 ± 0.13	6.77	-6020.73 ± 116.75	-6013.96 ± 116.75
Zn^{2+} ion 1	25.41 ± 0.77	-2427.66 ± 72.19	1291.3 ± 42.75	-0.88 ± 0.04	0.63	-1114.16 ± 27.12	-1113.51 ± 27.12
Zn^{2+} ion 2	13.79 ± 0.96	-799.35 ± 63.45	188.73 ± 35.14	-0.47 ± 0.04	0.52	-598.61 ± 29.09	-598.09 ± 29.09
Ca^{2+} ion 1	38.15 ± 1.90	-1796.42 ± 89.21	605.84 ± 30.55	-1.14 ± 0.07	1.03	-1154.73 ± 56.65	-1153.69 ± 56.65
Ca^{2+} ion 2	6.52 ± 0.58	-367.21 ± 36.96	73.71 ± 9.94	-0.35 ± 0.05	0.18	-289.24 ± 27.33	-298.07 ± 27.33
Ca^{2+} ion 3	14.94 ± 0.57	-1480.59 ± 42.61	320.38 ± 11.31	-1.32 ± 0.05	0.23	-1147.96 ± 30.78	-1147.73 ± 30.78

Residue contribution to the binding energy with their associated interatomic distances and geometries are given in Table 6. The catalytic Zn^{2+} ion1 maintains interactions with His^{403}, His^{407}, and His^{413} similar to what is observed in the x-ray crystal structure. The Glu^{404} sidechain Oε atoms are in closer proximity than what is observed in the crystal structure. The binding geometry for the bound His residues is trigonal pyramidal however, if a coordinated hydration water is considered, this geometry would be tetrahedral. Other important interactions are also noted between Zn^{2+} ion 1 and Asp and Glu residues within the Cat domain. These residues contribute significantly to MMP-2 to Zn^{2+} ion 1 interaction energy despite being outside what is considered to be the normal coordination sphere of the ion. Zn^{2+} ion 2 loses contact with His^{178}, His^{193}, and His^{206} shifting to a more linear coordination geometry that is depended on a strong interaction with the Oδ atoms of Asp^{180}.

Ca^{2+} ion 1 has strong interactions with the adjacent Oδ atoms of Asp and Oε atoms of Glu of adjacent residues. There is a shift from the divalent cation to backbone carbonyl oxygen interactions that are observed in the crystal structure to interactions dominated by the acidic sidechain groups. The coordination geometry changes from pentagonal pyramidal to a seesaw geometry. The strong interaction and coordination with these sidechains is expected and has been previously observed for other systems [32]. In general the favorability of interaction is Glu>Asp with the difference attributed to increased flexibility of the Glu residue secondary to presence of the extra methyl group. The binding of this ion is also similar to that of the catalytic Zn^{2+} ion 1 in that adjacent but non-coordinated electronegative Asp and Glu residues make significant contributions to its binding energy. Ca^{2+} ion 2 freely diffuses out of its binding site and although it has strong interactions with several electrostatic sidechains, these interactions are mostly transient. Ca^{2+} ion 3 is more stable in its RMSF compared to Ca^{2+} ion 2 (Table 3) however, it is still much more variable that the other associated ions. The interaction energies and geometry are again dominated by the electronegative sidechains of Asp residues surrounding the central core of the Hpx domain.

Table 6. Distances of ions to the interacting protein residue atoms identified from the 1CK7 X-ray crystal. Statistically significant protein residue atoms as identified by outlier analysis with the associated per residue interaction energies and binding geometry. 1CK7 identified interactions are marked with a dagger (†), statistically significant interactions are marked with an asterisk (*), and binding geometries were determined for those atoms within 0.35 nm. Distance are given in nm and energies in kJ mol^{-1}.[a,b]

Protein Atom	$\Delta E_{binding}$	Geometry	Zn^{2+} Ion 1	Zn^{2+} ion 2	Ca^{2+} ion 1	Ca^{2+} ion 2	Ca^{2+} ion 3
†*Glu^{404}:Oε1	-154.8223		0.46 ± 0.05				
†*Glu^{404}:Oε2			0.48 ± 0.06				
*Glu^{412}:Oε1	-88.9662		0.96 ± 0.13				
*Glu^{412}:Oε2			0.96 ± 0.13				
*Asp^{188}:Oδ1	-73.8901		1.46 ± 0.18				
*Asp^{188}:Oδ2			1.47 ± 0.18				
†*His^{403}:Nε2	-72.1511	trigonal pyramidal	0.21 ± 0.01				
*Asp^{437}:Oδ1	-67.5468		1.29 ± 0.11				
*Asp^{437}:Oδ2			1.29 ± 0.11				
*Asp^{185}:Oδ1	-66.7783		1.65 ± 0.20				
*Asp^{185}:Oδ2			1.68 ± 0.16				
†*His^{407}:Nε2	-66.3450	trigonal pyramidal	0.21 ± 0.01				
*Asp^{208}:Oδ1	-65.6438		1.60 ± 0.12				
*Asp^{208}:Oδ2			1.49 ±0.09				
*Glu^{211}:Oε1	-64.0824		1.57 ± 0.13				
*Glu^{211}:Oε2			1.58 ± 0.11				
†*His^{413}:Nε2	-61.5998	trigonal pyramidal	0.21 ± 0.01				
*Asp^{416}:Oδ1	-61.3612		1.52 ± 0.05				
*Asp^{416}:Oδ2			1.52 ± 0.05				
†*Asp^{180}:Oδ1	-92.3424	linear		0.19 ± 0.01			
†*Asp^{180}:Oδ2		linear		0.19 ± 0.01			
*Glu^{177}:Oε1	-42.684			0.71 ± 0.27			
*Glu^{177}:Oε2				0.71 ± 0.27			
†His^{178}:Nε2	-1.1803			1.13 ± 0.17			
†His^{206}:Nδ1	-1.0497			1.00 ± 0.42			
†His^{193}:Nε2	-0.2562			1.34 ± 0.51			
†*Asp^{185}:C=O	-167.499				0.70 ± 0.09		
†*Asp^{185}:Oδ1		seesaw			0.29 ± 0.16		
†*Asp^{185}:Oδ2		seesaw			0.28 ± 0.16		
†*Glu^{211}:Oε1	-147.6924	seesaw			0.33 ± 0.11		
†*Glu^{211}:Oε2		seesaw			0.35 ± 0.10		
†*Asp^{208}:Oδ1	-142.7074	seesaw			0.34 ± 0.09		
†*Asp^{208}:Oδ2		seesaw			0.24 ± 0.02		
*Asp^{210}:Oδ1	-112.8532	seesaw			0.33 ± 0.14		
*Asp^{210}:Oδ2		seesaw			0.35 ± 0.14		
*Asp^{209}:Oδ1	-61.7053				0.77 ± 0.16		
*Asp^{209}:Oδ2					0.76 ± 0.14		
†*Asp^{188}:C=O	-57.9066				0.98 ± 0.14		
*Asp^{188}:Oδ1					1.27 ± 0.19		
*Asp^{188}:Oδ2					1.27 ± 0.19		
*Glu^{404}:Oε1	-50.3849				1.37 ± 0.14		
*Glu^{404}:Oε2					1.38 ± 0.15		
*Asp^{134}:Oδ1	-49.6054				1.19 ± 0.13		
*Asp^{134}:Oδ2					1.19 ± 0.14		
*Asp^{392}:Oδ1	-49.3403				1.37 ± 0.17		
*Asp^{392}:Oδ2					1.36 ± 0.17		
*Asp^{180}:Oδ1	-46.489				1.50 ± 0.25		
*Asp^{180}:Oδ2					1.51 ± 0.25		
†Gly^{186}:C=O	-7.1579				0.75 ± 0.17		
†Leu^{198}:C=O	-0.4421				2.68 ± 0.17		

[a] Zn^{2+} ion 1 is the catalytic ion.

[b] Glu^{404} is critical to catalytic activity.

Peer-reviewed version available at Int. J. Mol. Sci. 2019, 20, 4194; doi:10.3390/ijms20174194

Table 6 Continued. Distances of ions to the interacting protein residue atoms identified from the 1CK7 X-ray crystal. Statistically significant protein residue atoms as identified by outlier analysis with the associated per residue interaction energies and binding geometry. 1CK7 identified interactions are marked with a dagger (†), statistically significant interactions are marked with an asterisk (*), and binding geometries were determined for those atoms within 0.35 nm. Distance are given in nm and energies in kJ mol^{-1}.[a,b]

Protein Atom	$\Delta E_{binding}$	Geometry	Zn^{2+} ion 1	Zn^{2+} ion 2	Ca^{2+} ion 1	Ca^{2+} ion 2	Ca^{2+} ion 3
†*Glu^{166}:Oε1	-59.5996					2.48 ± 2.14	
†*Glu^{166}:Oε2						2.48 ± 2.14	
†Asp^{204}:Oδ1	-31.9972					2.61 ± 1.65	
†Asp^{204}:Oδ2						2.61 ± 1.66	
†Asp^{168}:C=O	-31.2060					2.51 ± 1.60	
Asp^{168}:Oδ1						2.57 ± 1.59	
Asp^{168}:Oδ2						2.54 ± 1.60	
†Ala^{167}:C=O	-1.3321					2.52 ± 1.75	
†Gly^{200}:C=O	0.3799					2.57 ± 1.70	
†Gly^{202}:C=O	-1.0605					2.55 ± 1.56	
*Asp^{521}:Oδ2							0.98 ± 0.16
†*Asp^{569}:C=O	-79.0260						0.71 ± 0.18
*Asp^{569}:Oδ1							0.92 ± 0.15
*Asp^{569}:Oδ2							0.91 ± 0.15
*Asp^{490}:Oδ1	-69.0106						1.19 ± 0.11
*Asp^{490}:Oδ2							1.19 ± 0.11
*Asp^{615}:Oδ1	-68.1150						1.29 ± 0.14
*Asp^{615}:Oδ2							1.30 ± 0.14
*Asp^{153}:Oδ1	-63.3125						2.52 ± 0.96
*Asp^{153}:Oδ2							2.51 ± 0.95
*Asp^{472}:Oδ1	-54.3377						1.54 ± 0.15
*Asp^{472}:Oδ2							1.54 ± 0.15

[a] Zn^{2+} ion 1 is the catalytic ion.

[b] Glu^{404} is critical to catalytic activity.

3. Materials and Methods

3.1 Matrix Metalloprotease-2 Starting Conformation.

Initial coordinates were obtained from the X-ray l structure of the Glu^{404} to Ala^{404} mutant of the human MMP-2 (PDB ID: 1CK7) [13]. Crystallographically resolved waters and sulfate ions were removed while the protein bound Zn^{2+}, Ca^{2+}, Na^{+}, and Cl^{-} ions where retained. Residues 31-109 were removed as in the biologically active form of the enzyme. The crystallography non-resolved loop from residues Asp^{450}-Thr^{460} was build using the homology modelling script of YASARA [16]. The coordinates for the sidechain of Glu^{404} and the Zn^{2+} coordinated water molecule that replaces Cys^{102} at the enzyme active site were derived from the MMP-13 crystal structure (PDB ID: 1XUD) by a least squares fitting of the backbone and Zn^{2+} atoms of both x-ray structures [33]. Sidechain protonation states of the Zn^{2+} associated His residues were assigned based on the 1CK7 crystal structure as follows: HND1 for His^{178}, His^{403}, His^{193}, His^{407}, His^{413}, and HNE2 for His^{206}. The remaining Histidine residues (His^{163}, His^{276}, and His^{628}) were assigned automatically with HNE2 atoms using the *pdb2gmx* module of GROMACS version 5.1.2 [34,35]. The protonation state and charges of all other residues within the protein were set to correspond to a pH of 7.0.

4.2 Molecular Dynamics.

Simulations were performed using the CHARMM36m force field with modified TIP3Pm water model and the CM model of divalent metal cation parameters of Li *et al.* as implemented in GROMACS version 5.1.2 [Zn^{2+} (σ = 0.226466454151 nm, ε = 0.01381916624 kJ•mol^{-1}) and Ca^{2+} (σ = 0.293818397243 nm, ε = 0.44320568080 kJ•mol^{-1})] [36-41]. The CM model attempts to balance

Peer-reviewed version available at Int. J. Mol. Sci. 2019, 20, 4194; doi:10.3390/ijms20174194

hydration free energy by optimizing the ion-oxygen distance in the first solvation shell. The metal ions are represented with a standard Lennard-Jones and coulomb potential energy function [42]:

$$\boldsymbol{E_{ion}} = \sum_{j}^{N} 4\boldsymbol{\varepsilon_{ij}}\left[\left(\frac{\boldsymbol{\sigma_{ij}}}{\boldsymbol{r_{ij}}}\right)^{12} - \left(\frac{\boldsymbol{\sigma_{ij}}}{\boldsymbol{r_{ij}}}\right)^{6}\right] + \sum_{j}^{N} \frac{\boldsymbol{q_i q_j}}{4\boldsymbol{\pi\varepsilon_0 r_{ij}}} \tag{2}$$

where,

$$\boldsymbol{\sigma_{ij}} = \frac{(\boldsymbol{\sigma_i} + \boldsymbol{\sigma_j})}{2} \tag{3}$$

and

$$\boldsymbol{\varepsilon_{ij}} = \sqrt[2]{(\boldsymbol{\varepsilon_i} \bullet \boldsymbol{\varepsilon_j})} \tag{4}$$

and *i* and *j* are atom indices. The system was solvated in a truncated dodecahedron with 45105 TIP3Pm water molecules. The minimal distance of the protein to the edge of the dodecahedron was 1.4 nm. The system was neutralized with 141 and 132, Na^+ and Cl^- ions, respectively, so that the final concentration of the NaCl was set to 150 mM; the initially retained protein bound Zn^{2+}, Ca^{2+}, Na^+, and Cl^- ions are not included. . The system was subjected to 5000 steps of steepest descent energy minimization allowing all bond distances and angles to relax. This was followed with 10 ns of NVT simulation at 310 K so that the position of the protein heavy atoms and retained Zn^{2+}, Ca^{2+}, Na^+, and Cl^- ions and catalytic Zn^{2+} associated water were constrained to their energy -minimized coordinates with force constant of 1000 $kJ\bullet mol^{-1}$. The solvent and non-restrained Na^+ and Cl^- ions were then subjected to 10 ns of NPT simulation at 310 K and 101.325 kPa using Berendsen temperature and pressure scaling with a relaxation constant of 0.1 ps and 4.5×10^{-5} bar^{-1} isothermal compressibility [43]. The heavy atom restraints were then removed and the entire system was subjected to 10 ns of NPT at 310 K and 101.325 kPa using the stochastic velocity-rescaling method of Bussi and Berendsen pressure coupling with a relaxation constant of 1.0 ps and 4.5×10^{-5} bar^{-1} isothermal compressibility [44]. The production run consisted of 1.0 μs NPT simulation at 310 K and 101.325 kPa. The integration step was 2 fs, the LINCS algorithm was used to constrain all bonds to their correct length, with a warning angle of 30° [45,46]. The long-range electrostatic interactions were calculated using the PME method with 1.2 nm cutoff distance and 0.15 nm Fourier spacing [47]. For the calculations of Van der Waals interactions the short-range and long-range cutoff, respectively, was 1.0 and 1.2 nm. The protein and solvent with ions were separately coupled to a 1 bar Parrinello-Rahman barostat and the temperature was maintained by the velocity-rescaling method [44,48].

4.3 Biophysical Properties

The Cα-trace root mean square deviation (RMSD) and per-residue root mean square fluctuation (RMSF) from the average sampled peptide conformation were calculated with the *g_rmsd* and *g_rmsf* utilities of GROMACS, respectively [35]. The >50% sampled per-residue DSSP assigned [23] secondary structure (α-helix, β-sheet, β-bend/turn, and coil) were determined using the *do_dssp* utility of GROMACS and an in house *perl* script to calculate sampling statistics. The hydration of the Zn^{2+} and Ca^{2+} was determined by calculating the radial distribution function for the oxygen atom of the surrounding water molecules using the *g_rdf* utility in GROMACS [35]. The integral of the radial distribution function is equal to the probability of finding water molecules within the defined radius (0.5 nm for the first hydration shell) [49,50].

doi:10.20944/preprints201907.0085.v1

Peer-reviewed version available at Int. J. Mol. Sci. 2019, 20, 4194; doi:10.3390/ijms20174194

4.4 Conformational Analysis.

The time-dependent ϕ/ψ dihedral angles from residues 111 to 659 of the protein were extracted from the trajectories using the *g_rama* utility of GROMACS. An in-house python script was used to transform the data for the input to the dPCA program provided by Dr. Yuguang Mu [51,52]. Lowest energy conformations were identified by projecting the trajectories of the first two principal components (dPC1 and dPC2) onto a two-dimensional free energy (ΔG) landscape:

$$\Delta G = -R \cdot T \cdot \ln\frac{\rho_{x,y}}{\rho_{max}} \tag{5}$$

where R is the universal gas constant, T is the temperature, x and y are the first two dihedral principal components from the trajectory. The free energy (ΔG) landscape was calculated by dividing the dPC1-dPC2 subspace into grids creating a 2D histogram of the sampled phase space and calculating the probability $\rho_{x,y}$ where ρ_{max} corresponds to the grid with the maximum probability of occurrence. Results were visualized using the *scatterplot3D, akima*, and *latticeExtra* packages in the R software environment with conformations and secondary structural elements rendered using YASARA [16,21,53-55]. Families of low energy conformations were identified using k-means clustering as implemented in the *cluster* package in R and the identified lowest energy conformations extracted for further analysis [21,22]. The optimal number of clusters was determined by visual inspection, sum of squared error (SSE), average silhouette width (S_{AVG}), silhouette coefficient (SC) and distribution plots [56-59].

4.5 Dynamic Cross-Correlation Matrix.

Correlated motions within and between domains were evaluated by calculating dynamic cross correlation matrices (DCCM) from the principal components of the Cα-trace covariance matrix using the *GeoStaS* method as implemented in the *Bio3D* package written in R [60-62]. The results are displayed as a color coded matrix of Pearson correlation coefficients with a value of -1 indicated completely anti-correlated motions and a value of +1 indicating completely correlated motions [63,64].

4.6 Interaction Energy.

The free energy of binding between the metal cations and protein was calculated using the *g_mmpbsa* program [65]. The polar component of the solvation energy was calculated using the Poisson-Boltzmann equation and non-polar component calculated from the solvent-accessible surface area approximation [66,67]. Dielectric constants for the solute and water were 4 and 80, respectively. The entropic contribution to the binding energy was determined using the interaction entropy method of Zhang and coworkers [31,68]. The trajectory was sampled every 0.1 ns for the equilibrium phase (200 ns to 1000 ns). A bootstrap analysis (n = 5000) was performed to obtain standard errors and the residue contributions to the binding energy were also calculated.

The residue contributions to binding were deconvoluted. To determine the most significant residue interactions between MMP-2 and the divalent cations, an outlier analysis was performed to identify statistically significant interactions. The distribution of interaction energies were not Gaussian (normal). In the setting of non-normal distributions, the method of Tukey's fences can be used to identify those observations that are outside of the expected fluctuations within the data [69]. Tukey's fences defines minimum and maximum values of the interaction energy:

$$[MinimumValue, MaximumValue] \tag{6}$$

such that measurements less than or equal to the minimum value or greater than or equal to the maximum value are considered statistical outliers. The respective minimum and maximum values of the fences, are defines as:

$$[Q_1 - k(Q_3 - Q_1), Q_3 + k(Q_3 - Q_1)] \tag{7}$$

where Q_1, Q_3 are the interquartile values, and k is the constant that defines the outlier range (k = 1.5 is an outlier, k = 3.0 is an extreme outlier) [69].

5. Conclusions

Here, we report a microsecond scale molecular dynamics analysis of the full biologically active protein (Cat, Fib, and Hpx domains with Lnk region) with its crystallographically associated (structural) divalent metal ions (Zn^{2+} and Ca^{2+}). So far, most of the simulations were done on the truncated structure of MMP-2 protein [28-30]. Diaz et al. had previously investigated the role of the structural metal ions using a truncated version of the protein with the AMBER force field, a non-bonded model of Ca^{2+} and a bonded tetrahedral model of Zn^{2+} using 10 ns simulations [29]. Their results are in agreement with ours: the Zn^{2+} in 1 (catalytic ion) and Ca^{2+} ion 1 demonstrate the lowest RMSFs and are tightly bound within their crystallographically defined pockets. Zn^{2+} ion 2 was more flexible and associated with the S-loop which has demonstrated increased flexibility in both prior simulations and our own however, the use of short simulation time (10 ns) and a bonded potential representing the interaction between the Zn^{2+} ion and the associated His and Asp/Glu residues may have artificially stabilized the protein-Zn^{2+} ion 2 interaction. This is an important point since prior investigators have identified variations in the stoichiometry of the MMP-2-Zn^{2+} interaction that are strongly dependent on the purification procedure used [70]. The current non-bonded model of Li *et al.* with the CHARMM36m force field appears to be a reasonable model of protein metal cation interactions and is in agreement with prior studies [71,72].

Dihedral principle component analysis demonstrates that only minor structural fluctuations occur within the Cat domain. There is increased fluctuations within the Fib domain however, this is expected given the large number of β-turn/loop structures connecting the individual subdomains. The three lowest energy populations of conformations from differ predominantly in the orientation of the Hpx domain in relation to the Cat and Fib domains. This is confirmed by the DCCM analysis where the difference in orientation of the Hpx to the Cat and Fib domains comprises the first two principle components. These inter domain movements are facilitated by the flexible linker region Gly^{446} through Asp^{476} and may play and important role in collagen substrate binding, unravelling and subsequent catalysis [24-27].

Supplementary Materials: Supplementary materials can be found at www.mdpi.com/xxx/s1.

Author Contributions: conceptualization, C.R.W and S.L.; methodology, C.R.W and S.L.; software, C.R.W., L.V.-O., and S.L.; validation, C.R.W. and L.V.-O.; formal analysis, C.R.W. and L.V.-O.; investigation, L.V.-O.; resources, C.R.W.; data curation, L.V.-O.; writing—original draft preparation, C.R.W.; writing—review and editing, C.R.W, L.V.-O., and S.L.; visualization, L.V.-O.; supervision, C.R.W. and S.L. ; project administration, C.R.W.; funding acquisition, C.R.W.

Funding: Funding for this investigation was provided in part through a grant from Mayo Clinic Health System-Franciscan Healthcare Foundation Inc. and Mayo Foundation for Medical Education and Research.

Acknowledgments: The molecular dynamics simulations and subsequent analysis were completed using the High Performance Cluster, Resource Computing Services, Mayo Clinic, Rochester, Minnesota.

Conflicts of Interest: Dr. Charles R. Watts is a consultant for Medtronic Spine and Biologics. The remaining authors have disclosed that they do not have any conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results”.

Abbreviations

dPCA	Dihedral Principle Component Analysis
DCCM	Domain Cross-Correlation Matrix
MD	Molecular Dynamics

Peer-reviewed version available at Int. J. Mol. Sci. 2019, 20, 4194; doi:10.3390/ijms20174194

MMP-2	Matrix Metalloproteinase-2
PDB	Protein Databank
RMSD	Root Mean Square Deviation
RMSF	Root Mean Square Fluctuation
SASA	Solvent Exposed surface Area

References

1. Basbaum, C.B.; Werb, Z. Focalized proteolysis: spatial and temporal regulation of extracellular matrix degradation at the cell surface. *Curr Opin Cell Biol.* **1996**, *8(5)*, 731-738, DOI: 10.1016/S0955-0674(96)80116-5.
2. Birkedal-Hansen, H.; Moore, W.G.; Bodden, M.K.; Windsor, L.J.; Birkedal-Hansen, B.; DeCarlo, A.; Engler, J.A. Matrix metalloproteinases: a review. *Crit Rev Oral Biol Med.* **1993**, *4(2)*, 197-250, DOI: 10.1177/10454411930040020401.
3. Maskos, K. Crystal structure of MMPs in complex with physiological and pharmacological inhibitors. *Biochimie.* **2005**, *87(3-4)*, 249-263, DOI: 10.1016/j.biochi.2004.11.019.
4. Ra, H.-J.; Parks, W.C. Control of Matrix Metalloproteinase Catalytic Activity. *Matrix Biol.* **2007**, *26(8)*, 587-596, DOI: 10.1016/j.matbio.2007.07.001.
5. Verma, R.P.; Hansch C. Matrix metalloproteinases (MMPs): chemical-biological functions and (Q)SARs. *Bioorg. Med. Chem.* **2007**, *15 (6)*, 2223–68. doi:10.1016/j.bmc.2007.01.011. PMID 17275314.
6. van Meurs, J.; van Lent, P.; Holthuysen, A.; Lambrou, D.; Bayne, E.; Singer, I.; van den Berg, W. Active matrix metalloproteinases are present in cartilage during immune complex mediated arthritis: a pivotal role for stromelysin-1 in cartilage destruction. *J Immunol.* **1999**; *163(10)*, 5633-5639, http://www.jimmunol.org/content/163/10/5633.
7. Nagase, H.; Woessner, J.F. Jr. Matrix metalloproteinases. *J Biol Chem.* **1999**; *274(31)*, 21491-21494, DOI: 10.1074/jbc.274.31.21491.
8. Steffensen, B.; Häkkinen, L.; Larjava, H. Proteolytic events of wound-healing-coordinated interactions among matrix metalloproteinases (MMPs), integrins, and extracellular matrix molecules. *Crit Rev Oral Biol Med.* **2001**, *12(5)*, 373-398, DOI: 10.1177/10454411010120050201.
9. Egeblad. M; Werb, Z. New functions for the matrix metalloproteinases in cancer progression. *Nat Rev Cancer.* **2002**; *2(3)*, 163-176, DOI: 10.1038/nrc745.
10. Fingleton, B. Matrix metalloproteinases: roles in cancer and metastasis. *Front Biosci.* **2006**, *11*, 479-491, DOI: 10.2741/1811.
11. Butler, G.S.; Overall, C.M. Updated biological roles for matrix metalloproteinases and new "intracellular" substrates revealed by degradomics. *Biochemistry.* **2009**, *48(46)*, 10830-10845, DOI: 10.1021/bi901656f.
12. Rodríguez, D.; Morrison, C.J.; Overall, C.M. Matrix metalloproteinases: what do they not do? New substrates and biological roles identified by murine models and proteomics. *Biochim Biophys Acta.* **2010**, *1803(1)*, 39-54, DOI: 10.1016/j.bbamcr.2009.09.015.
13. Morgunov, E.; Tuuttila, A.; Bergmann, U.; Isupov, M.; Lindqvist, Y.; Schneider, G.; Tryggvason, K.; Structure of Human Pro-Matrix Metalloproteinase-2: Activation Mechanism Revealed. *Science.* **1999**, *284(5420)*, 1667-1670, DOI: 10.1126/science.284.5420.1667.
14. Díaz, N.; Suárez, D.; Valdés, H. From the X-ray Compact Structure to the Elongated form of the Full-Length MMP-2 Enzyme in Solution: A Molecular Dynamics Study. *J Am Chem Soc.* **2008**, *130(43)*, 14070-14071, DOI: 10.1021/ja806090v.
15. Díaz. N.; Suárez, D. Alternative Interdomain Configurations of the Full-Length MMP-2 Enzyme Explored by Molecular Dynamics Simulations. *J Phys Chem B.* **2012**, *116(9)*, 2677-2686, DOI: 10.1021/jp211088d.
16. Krieger, E.; Vriend, G. YASARA View - molecular graphics for all devices - from smartphones to workstations. *Bioinformatics.* **2011**, *30(20)*, 2981-2982., DOI: 10.1093/bioinformatics/btu426.
17. Petrucci, R.H.; Herring, F.G.; Madura, J.D.; Bissonnette, C. *General Chemistry: Principles and Modern Applications,* 11th ed.; Pearson: Don Mills, Ontario, Canada, 2017; p 411-465, ISBN-13: 978-0132931281.
18. Amadei, A.; Ceruso, M.A.; Di Nola, A. On the convergence of the conformational coordinates basis set obtained by the essential dynamics analysis of proteins' molecular dynamics simulations. *Proteins.* **1999**, *36(4)*, 419-424, DOI: 10.1002/(SICI)1097-0134(19990901)36:4<419::AID-PROT5>3.0.CO;2-U.
19. Hayward, S.; de Groot, B.L. Normal modes and essential dynamics. In: *Methods in Molecular Biology: Molecular modeling of proteins,* Kukol, A., Ed.; Humana Press, New York City, New York, USA, 2008; Volume 443, pp. 89-106, ISBN 978-1-59745-177-2.

Peer-reviewed version available at *Int. J. Mol. Sci.* **2019**, *20*, 4194; doi:10.3390/ijms20174194

20. Andricioaei, I.; Karplus, M. On the calculation of entropy from covariance matrices of the atomic fluctuations. *J Chem Phys.* **2001**, *115(14)*: 6289–6292, DOI: 10.1063/1.1401821.
21. R Core Team. R: A language and environment for statistical computing. Available online: http://www.R-project.org/ (accessed on 2 Jan 2019).
22. Maechler, M.; Rousseeuw, P.; Struyf, A.; Hubert, M.; Hornik, K.; Studer, M.; Roudier, P.; Gonzalez, J.; Kozlowski, K. cluster: methods for cluster analysis. Available online: https://cran.r-project.org/web/packages/cluster/ (accessed on 2 Jan 2019).
23. Kabsch, W.; Sander, C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers.* **1983**, *22(12)*, 2577-2637, DOI: 10.1002/bip.360221211.
24. Steffensen, B.; Wallon, U.M.; Overall, C.M. Extracellular matrix binding properties of recombinant fibronectin type II-like modules of human 72-kDa gelatinase/type IV collagenase. High affinity binding to native type I collagen but not native type IV collagen. *J Biol Chem.* **1995**, *270(19)*, 11555–11566, DOI: 10.1074/jbc.270.19.11555.
25. Gehrmann, M.L.; Douglas, J.T.; Bányai, L.; Hedvig, T.; Patthy, L.; Llinás, M.; Modular autonomy, ligand specificity, and functional cooperativity of the three in-tandem fibronectin type II repeats from human matrix metalloproteinase 2. *J Biol Chem.* **2004**, *279(45)*, 46921–46929, DOI: 10.1074/jbc.M408859200.
26. Xu, X.; Wang, Y.; Lauer-Fields, J.L.; Fields, G.B.; Steffensen, B. Contributions of the MMP-2 collagen binding domain to gelatin cleavage. Substrate binding via the collagen binding domain is required for hydrolysis of gelatin but not short peptides. *Matrix Biol.* **2004**; *23(3)*, 171−181, DOI: 10.1016/j.matbio.2004.05.002.
27. Xu, X.; Mikhailova, M.; Llangovan, U.; Chen, Z.; Yu, A.; Pal, S.; Hinck, A.P.; Steffensen, B. Nuclear magnetic resonance mapping and functional confirmation of the collagen binding sites of matrix metalloproteinase-2. *Biochem.* **2009**, *48(25)*, 5822−5831, DOI: 10.1021/bi900513h.
28. Díaz N, Suárez D, Molecular dynamics simulations of the active matrix metalloproteinase-2: Positioning of the N-terminal fragment and binding of a small peptide substrate. *Proteins.* **2008**, 72(1), 50-61, DOI: 10.1002/prot.21894.
29. Díaz, N.; Suárez, D. Molecular Dynamics Simulations of Matrix Metalloproteinase 2: Role of the Structural Metal Ions. *Biochem.* **2007**, *46(31)*, 8943-8952, DOI: 10.1021/bi700541p.
30. Díaz, N.; Suárez, D. Peptide Hydrolysis Catalyzed by Matrix Metalloproteinase 2: A computational Study. *J Phys Chem B.* **2008**, *112(28)*, 8412-8424, DOI: 10.1021/jp803509h.
31. Duan, L.; Liu, X.; Zhang, J.Z.H. Interaction Entropy: A New Paradigm for Highly Efficient and Reliable Computation of Protein-Ligand Binding Free Energy, *J Am Chem Soc.* **2016**, *138(17)*, 5722-5728, DOI: 10.1021/jacs.6b02682.
32. Cates, M.S.; Teodoro, M.L.; Phillips, G.N. Jr., Molecular Mechanisms of Calcium and Magnesium Binding to Parvalbumin. *Biophys J.* **2002**, *82(3)*, 1133-1146, DOI: 10.1016/S0006-3495(02)75472-6.
33. Engel, C.K.; Pirard, B.; Schimanski, S.; Kirsch, R.; Habermann, J.; Klingler, O.; Schlotte, V.; Weithmann, K.U.; Wendt, K.U. Structural basis for the highly selective inhibition of MMP-13. *Chem.Biol.* **2005**, *12(2)*, 181-189, DOI: 10.1016/j.chembiol.2004.11.014
34. Abraham, M.J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J.C.; Hess, B.; Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX.* **2015**, *1-2*, 19-25, DOI: 10.1016/j.softx.2015.06.001
35. Abraham, M.J.; van der Spoel, D.; Lindahl, E.; Hess, B. and the GROMACS development team, *GROMACS User Manual* ver. 5.1.2 2016, http://www.gromacs.org, (accessed Jan 2, 2019).
36. Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; de Groot, B.L.; Grubmüller, H.; MacKerell, A.D. Jr. CHARMM36m: an improved force field for folded and intrinsically disordered proteins, *Nat Methods.* **2017**, *14(1)*, 71-73, DOI: 10.1038/nmeth.4067.
37. Best, R.B.; Zhu, X.; Shim, J.; Lopes, P.E.; Mittal, J.; Feig, M.; Mackerell, A.D. Jr. Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone ϕ, ψ and Side-Chain χ^1 and χ^2 Dihedral Angles, *J Chem Theory Comput.* **2012**, *8(9)*, 3257-3273. DOI: 10.1021/ct300400x
38. MacKerell, A.D. Jr.; Feig, M.; Brooks, C.L. III. Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations, *J Comp Chem.* **2004**, *25(11)*, 1400-1415, DOI: 10.1002/jcc.20065
39. MacKerell, A.D. Jr.; Bashford, D.; Bellott, M.; Dunbrack, R.L.; Evanseck, J.D.; Field, M.J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F.T.; Mattos, C.; Michnick, S.; Ngo, T.;

Nguyen, D.T.; Prodhom, B.; Reiher, W.E.; Roux, B.; Schlenkrich, M.; Smith, J.C.; Stote, R.; Straub, J.; Watanabe, M.; Wiórkiewicz-Kuczera, J.; Yin, D.; Karplus, M. All-atom empirical potential for molecular modeling and dynamics Studies of proteins, *J. Phys. Chem. B* **1998**, *102(18)*, 3586-3616, DOI: 10.1021/jp973084f.

40. Bjelkmar, P.; Larsson, P.; Cuendet, M.A.; Bess, B.; Lindahl, E. Implementation of the CHARMM force field in GROMACS: Analysis of protein stability effects from correction maps, virtual interaction sites, and water models, *J Chem Theory Comput.* **2010**, *6(2)*, 459-466, DOI: 10.1021/ct900549r.
41. Li, P. Roberts, B.P.; Chakravorty, D.K.; Merz, K.M. Jr. Rational Design of Particle Mesh Ewald Compatible Lennard-Jones Parameters for +2 Metal Cations in Explicit Solvent. *J. Chem Theory Comput.* **2013**, *9(6)*, 2733-2748, DOI: 10.1021/ct400146w.
42. Stote, R.H.; Karplus, M. Zinc Binding in Proteins and Solution: A Simple but Accurate Nonbonded Representation. *Proteins.* **1995**, *23(1)*, 12-31, DOI: 10.1002/prot.340230104.
43. Berendsen, H.J.C.; Postma, J.P.M.; van Gunsteren, W.F.; DiNola, A.; Haak, J.R. Molecular dynamics with coupling to an external bath. *J Chem Phys* **1984,** *81(8)*, 3684–3690, DOI: 10.1063/1.448118.
44. Bussi, G.; Donadio, D.; Parrinello, M. Canonical Sampling through Velocity Rescaling. *J. Chem. Phys.* **2007**, *126(1)*, 014101, DOI: 10.1063/1.2408420.
45. Hess, B.; Bekker, H.; Berendsen, H.J.C.; Fraaije, J.G.E.M. LINCS: A Linear Constraint Solver for molecular simulations. *J Comp Chem.* **1997**, *18(12)*, 1463-1472, DOI: 10.1002/(SICI)1096-987X(199709)18:12<1463::AID-JCC4>3.0.CO;2-H.
46. Hess, B. P-LINCS: A Parallel Linear Constraint Solver for Molecular Simulation. *J Chem Theory Comput.* **2008**, *4(1)*, 116-22, DOI: 10.1021/ct700200b.
47. Essmann, U.; Perera, L.; Berkowitz, M.L.; Darden, T.; Lee, H.; Pedersen, L.G. A smooth particle mesh Ewald method. *J Chem Phys.* **1995**, *103(19)*, 8577-8592, DOI: 10.1063/1.470117.
48. Parrinello, M.; Rahman, A. Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method. *J Appl Ph.* **1981**, *52(12)*, 7182-7190, DOI: 10.1063/1.328693.
49. White, A.D.; Keefe, A.J.; Ella-Menye, J.R.; Nowinski, A.K.; Shao, Q.; Pfaendtner, J.; Jiang, S. Free energy of solvated salt bridges: A simulation and experimental study. *J. Phys. Chem. B* **2013**, *117*, 7254–7259 DOI: 10.1021/jp4024469
50. Nguyen, B.L.; Pettitt, B.M. Effects of Acids, Bases, and Heteroatoms on Proximal Radial Distribution Functions for Proteins. *J. Chem. Theory Comput.* **2015**, *11*, 1399–1409, DOI: 10.1021/ct501116v
51. Mu, Y.; Nguyen, P.H.; Stock, G. Energy landscape of a small peptide revealed by dihedral angle principal component analysis. *Proteins.* **2005**, *58(1)*, 45-52, DOI: 10.1002/prot.20310.
52. Altis, A.; Nguyen, P.H.; Hegger, R.; Stock, G. Dihedral Angle Principal Component Analysis of Molecular Dynamics Simulations. *J Chem Phys.* **2007**, *126(24)*, 216-225, DOI: 10.1063/1.274633.
53. Ligges, U.; Mächler, M. Scatterplot3d - an R package for visualizing multivariate data. *J Stat Software* **2003**, *8(11)*, 1-20, Available online, https://cran.rproject.org/web/packages/scatterplot3d/ (accessed on 2 Jan 2019).
54. Akima. H.; Gebhardt, A.; Petzold, T.; Maechler, M. akima: interpolation of irregularly and regularly spaced data. Available online, https://CRAN.R-project.org/package=akima (accessed on 2 Jan 2019).
55. Sarkar. D.; Andrews, F. latticeExtra: extra graphical utilities based on lattice. Available online, https://CRAN.R-project.org/package=latticeExtra (accessed on 2 Jan 2019).
56. Tan, P.-N.; Steinbach, M.; Kumar, V. Chapter 8, Cluster Analysis: Basic Concepts and Algorithms in Introduction to Data Mining. 2nd ed.; Pearson Press, New York, New, York, USA, 2005. pp. 487-568, ISBN-13: 978-0321321367
57. Tan, P.-N.; Steinbach, M.; Kumar, V. Chapter 9, Cluster Analysis: Additional Issues and Algorithms in Introduction to Data Mining. 2nd ed; Pearson Press, New York, New, York, USA, 2005. pp. 569-650, ISBN-13: 978-0321321367
58. Rousseeuw, P.J. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *J Com. Appl Math.* **1987**, *20(1)*, 53-56, DOI: 10.1016/0377-0427(87)90125-7.
59. Thinsungnoena, T.; Kaoungkub, N.; Durongdumronchaib, P.; Kerdprasopb, K.; Kerdprasopb, N. The Clustering Validity with Silhouette and Sum of Squared Errors. *Proc. 3rd Int Conf Ind Appl Eng.* **2015**; 3(1), 44-51, DOI: 10.12792/iciae2015.012.
60. Romanowska, J.; Nowínski, K.S.; Trylska, J. Determining Geometrically Stable Domains in Molecular Conformation Sets. *J Chem Theory Comput.* **2012**, *8(8)*, 2588-2599, DOI: 10.1021/ct300206j.

61. Grant, B.J.; Rodrigues, A.P.; El Sawy, K.M.; McCammon, J.A.; Caves, L.S. Bio3d: an R package for the comparative analysis of protein structures. Available online https://CRAN.R-project.org/package=bio3d, (accessed on 2 Jan 2019).
62. Grant, B.J. Rodrigues, A.P.; El Sawy, K.M.; McCammon, J.A.; Caves, L.S. Bio3d: an R package for the comparative analysis of protein structures. *Bioinformatics*. **2006**, 22(21), 2695-2696, DOI: 10.1093/bioinformatics/btl461.
63. Pearson K. Notes on regression and inheritance in the case of two parents. *Proc Roy Soc Lond.* **1895**, *58(1)*, 240–242, DOI: 10.1098/rspl.1895.0041.
64. Hünenberger, P.H.; Mark, A.E.; van Gunsteren, W.F. Fluctuation and cross-correlation analysis of protein motions observed in nanosecond molecular dynamics simulations. *J Mol Biol.* **1995**, *252(4)*, 492–405, DOI: 10.1006/jmbi.1995.0514.
65. Kumari, R.; Kumar, R.; Lynn A. g_mmpbsa: a GROMACS tool for high-throughput MMPBSA calculations. *J Chem Inf Model*. **2014**, *54(7)*, 1951–1962. DOI: 10.1021/ci500020m
66. Baker, N.A.; Sept, D.; Joseph, S.; Holst, M.J.; McCammon, J.A. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc Natl Acad Sci USA.* **2001**, *98(18)*, 10037−10041, DOI: 10.1073/pnas.181342398.
67. Sitkoff, D.; Sharp, K. A.; Honig, B. Accurate Calculation of Hydration Free-Energies Using Macroscopic Solvent Models. *J Phys Chem*. **1994**, *98(7)*, 1978−1988, DOI: 10.1021/j100058a043.
68. Sun, A.; Yan, Y.N.; Yang, M.; Zhang, J.Z.H. Interaction entropy for protein-protein binding. *J. Chem Phys.* **2017**, *146(12)*, 124124 .DOI: 10.1063/1.4978893.
69. Tukey, J.W. *Exploratory Data Analysis*; Addison-Wesley: Reading, Massachusetts, USA, 1977; pp 688, ISBN 0-201-07616-0. OCLC 3058187.
70. Springman, E.B.; Nagase, H.; Birkedal-Hansen, H.; Van Wart; H.E. Zinc content and function in human fibroblast collagenase. *Biochemistry* 1995; 34: 15713-15720. DOI: 10.1021/bi00048a016
71. Ahlstrand E, Hermansson K, Friedman R, Interaction Energies in Complexes of Zn and Amino Acids: A Comparison of Ab Initio and Force Field Based Calculations. *J Phys Chem A.* 2017; 121: 2643-2654. DOI: 10.1021/acs.jpca6b12969
72. Wambo, T.O.; Chen, L.Y.; McHardy, S.F.; Tsin, A.T. Molecular dynamics study of human carbonic anhydrase II in complex with Zn^{2+} and acetazolamide on the basis of all-atom force field simulations. *Biophys Chem.* 2016; 214-215: 54-60. DOI: 10.1016/j.bpc.2016.05.006