Supplementary information

Identification of substrate-specific allosteric fingerprints in matrix metalloprotease-1 on amyloid-beta peptide aggregates and drug screening with single molecule insights

Sumaer Kamboj[1,#], Chase Harms[1,#], Derek Wright[1], Anthony Nash[2], Judith Klein-Seetharaman[3], and Susanta K. Sarkar[1,*]

[1]Department of Physics, Colorado School of Mines, Golden, CO, USA
[2]Nuffield Department of Clinical Neurosciences, University of Oxford, UK
[3]Department of Chemistry, Colorado School of Mines, Golden, CO, USA
[#]These authors contributed equally

Corresponding author:
[*]ssarkar@mines.edu

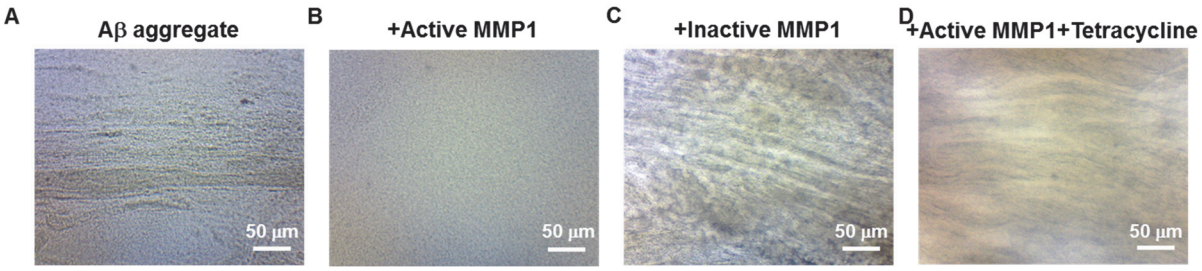

Figure S1. Light microscope images of Aβ aggregates. (**A**) Aggregates on a slide treated with Aβ buffer. (**B**) Aggregates on a slide treated with active MMP1 at 22 °C for 30 min. (**C**) Aggregates on a slide treated with inactive MMP1 at 22 °C for 30 min. (**D**) Aggregates on a slide treated with active MMP1 and 0.1 mg/mL tetracycline at 22 °C for 30 min.

Best-fit parameters to quantify interdomain dynamics at the single molecule level. We fitted a sum of two Gaussians to the histograms in **Figure 1**. The best-fit parameters are in **Table S1A**. We approximated the fit parameters b1 and b2 at the two conformational states, S1 and S2, respectively.

Table S1. Best-fit parameters for histograms and autocorrelations in Figure 1.

A Gaussian fit parameters for experimental histograms

$$y = a_1 \times e^{-\frac{(x-b_1)^2}{c_1^2}} + a_2 \times e^{-\frac{(x-b_2)^2}{c_2^2}}$$

	MMP1 without ligands		MMP1 with tetracycline	
	Active	Inactive	Active	Inactive
a1	14.45±0.04	8.21±0.12	10.07±0.10	25.28±0.11
b1/S1	0.47±0.01	0.45±0.01	0.45±0.01	0.49±0.01
c1	0.01±0.01	0.02±0.01	0.02±0.01	0.01±0.01
a2	6.35±0.02	4.41±0.06	6.33±0.06	4.56±0.03
b2/S2	0.52±0.01	0.51±0.01	0.55±0.01	0.57±0.01
c2	0.06±0.01	0.09±0.01	0.06±0.01	0.08±0.01

B Exponential fit parameters for correlations

$$C_\tau = d \times \exp^{-e \times \tau} + f$$

	MMP1 without ligands		MMP1 with tetracycline	
	Active	Inactive	Active	Inactive
d	0.20±0.01	0.36±0.01	0.15±0.01	0.16±0.01
e	0.03±0.01	0.01±0.01	0.04±0.01	0.08±0.01
f	0.01±0.01	-0.01±0.01	0.02±0.01	-0.01±0.01

C Kinetic rates calculated from histograms and correlations

	MMP1 without ligands		MMP1 with tetracycline	
	Active	Inactive	Active	Inactive
k1 (s^{-1})	0.0160	0.0097	0.0267	0.0488
k2 (s^{-1})	0.0078	0.0039	0.0124	0.0294

We fitted exponential and power law distributions to the autocorrelations in **Figure 1.** Power law distribution does not fit the experimental autocorrelations. The best-fit parameters for the exponential fits are in **Table S1B**. We calculated the kinetic rates of interconversion between S1

and S2 from the fit parameters of histograms and correlations (**Table S1C**). The error bars represent the standard errors of the means.

Calculation of correlations. We used the following equation to calculate autocorrelations:

$$C(\tau)=\frac{1}{N-\tau}\sum_{t=1}^{N-\tau}\left\{I(t)-\frac{1}{N-\tau}\times\sum_{t'=1}^{N-\tau}I(t')\right\}\times\left\{I(t+\tau)-\frac{1}{N-\tau}\times\sum_{t'=1+\tau}^{N}I(t')\right\}$$

where $C(\tau)$ is the correlation at lag number τ, N is the number of points in a time series, and $I(t)$ is the value at t. We illustrate the process using a simple time series (*t,x*):

(0,3), (1,5), (2,1), (3,2), (4,6), (5,4)

$C(\tau)$ at lag number $\tau=0$:

In this time series, *N*=6 and the mean=(3+5+1+2+6+4)/6=21/6=3.5. After subtracting the mean to analyze only the fluctuations, we obtain a new time series:

(0,3-3.5), (1,5-3.5), (2,1-3.5), (3,2-3.5), (4,6-3.5), (5,4-3.5) or

(0,-0.5), (1,1.5), (2,-2.5), (3,-1.5), (4,2.5), (5,0.5)

We do not need to shift the time series to calculate the correlation value for $\tau=0$ and can multiply element by element and calculate $C(\tau)$ at the lag number $\tau=0$ as below:

-0.5, 1.5, -2.5, -1.5, 2.5, 0.5
-0.5, 1.5, -2.5, -1.5, 2.5, 0.5

We multiply element by element, add, and average by $N-\tau=6-0=6$ to get:

0.25+2.25+6.25+2.25+6.25+0.25=17.5, average=17.5/6=2.92.

Therefore, $C(\tau)$ at lag number $\tau=0$ is 2.92.

$C(\tau)$ at lag number $\tau=1$:

For $\tau=1$, we need to shift the time series by 1 time step and multiply element by element:

3, 5, 1, 2, 6, <u>4</u>
5, 1, 2, 6, 4

We disregard the last element, underlined 4, in the first row because it is unpaired. For the first row, the mean is (3+5+1+2+6)/5=17/5=3.4. For the second row, the mean is (5+1+2+6+4)/5=18/5=3.6. After subtracting the means, we get the following time series:

-0.4, 1.6, -2.4, -1.4, 2.6
1.4, -2.6, -1.6, 2.4, 0.4

We multiply element by element, add, and average by $N-\tau=6-1=5$ to get:

-0.56-4.16+3.84-3.36+1.04=-3.2, average=-3.2/5=-0.64.

Therefore, $C(\tau)$ at lag number $\tau=1$ is -0.64.

$C(\tau)$ at lag number $\tau = 2$:

For $\tau = 2$, we need to shift the time series by 2 time steps and multiply element by element::

3, 5, 1, 2, 6, 4
1, 2, 6, 4

We disregard the last two elements, underlined 6 and 4, in the first row because they are unpaired. For the first row, the mean is (3+5+1+2)/4=11/4=2.75. For the second row, the mean is (1+2+6+4)/4=13/4=3.25. After subtracting the means, we get the following time series:

0.25, 2.25, -1.75, -0.75
-2.25, -1.25, 2.75, 0.75

We multiply element by element, add, and average by $N-\tau=6-2=4$ to get:

-0.5625-2.8125-4.8125-0.5625=-8.75, average=-8.75/4=-2.1875.

Therefore, $C(\tau)$ at lag number $\tau = 2$ is -2.1875. The autocorrelations at different lag numbers are:

(0, 2.92), (1, -0.64), (2, -2.1875), …

After normalization by dividing the value at $\tau = 0$, we get:

(0,1), (1,-0.22), (2,-0.75), …

We have followed the procedure above to calculate normalized correlations. For a two-state Poisson process, the decay of autocorrelations provides the sum of two kinetic interconversion rates between the two states.

Table S2. Best-fit parameters for two-state Poisson process simulations in Figure 2.

Input parameters

Active MMP1 (No ligand)
S1 = 0.47, S2 = 0.52
k1 = 0.02 s^{-1}, k2 = 0.01 s^{-1}
k1 + k2 = 0.03 s^{-1}
σ1 = 0.01, σ2 = 0.06

Inactive MMP1 (No ligand)
S1 = 0.45, S2 = 0.51
k1 = 0.01 s^{-1}, k2 = 0.001 s^{-1}
k1 + k2 = 0.01 s^{-1}
σ1 = 0.02, σ2 = 0.09

Recovered parameters
Gaussian fit parameters for simulated histograms

$$y = a_1 \times e^{-\frac{(x-b_1)^2}{c_1^2}} + a_2 \times e^{-\frac{(x-b_2)^2}{c_2^2}}$$

	Without noise		With noise	
	Active	Inactive	Active	Inactive
a1			28.61±0.01	19.34±0.01
b1/S1	0.47	0.45	0.47±0.01	0.45±0.01
c1			0.01±0.01	0.02±0.01
a2			3.30±0.01	1.90±0.02
b2/S2	0.51	0.51	0.52±0.01	0.51±0.01
c2			0.06±0.01	0.09±0.01

Exponential fit parameters for correlations

$$C_\tau = d \times \exp^{-e \times \tau} + f$$

	Without noise		With noise	
	Active	Inactive	Active	Inactive
d	1.08±0.01	1.08±0.01	0.52±0.01	0.36±0.01
e	0.03±0.01	0.02±0.01	0.02±0.01	0.02±0.01
f	-0.08±0.01	-0.09±0.01	-0.05±0.01	-0.03±0.01

Best-fit parameters for two-state Poisson process simulations. We modeled MMP1 dynamics as a two-state Poisson process. To simulate and test, we used experimentally determined values in **Table S1** for active and inactive MMP1 without ligands. We simulated and analyzed 350 smFRET trajectories, each 1000 s long, with the input parameters as noted in **Table S2**. The recovered parameters are on the right side of **Table S2**, indicating the applicability of the two-state description of MMP1 dynamics.

RMSF stabilization of simulations. We defined the input structure at t=0 as the reference structure and calculated the root-mean-square-fluctuation (RMSF) to check the simulations' stabilization. Simulations of MMP1 dynamics stabilize in ~5 ns (**Figure S2**). As such, we simulated 20 ns long dynamics for different conditions.

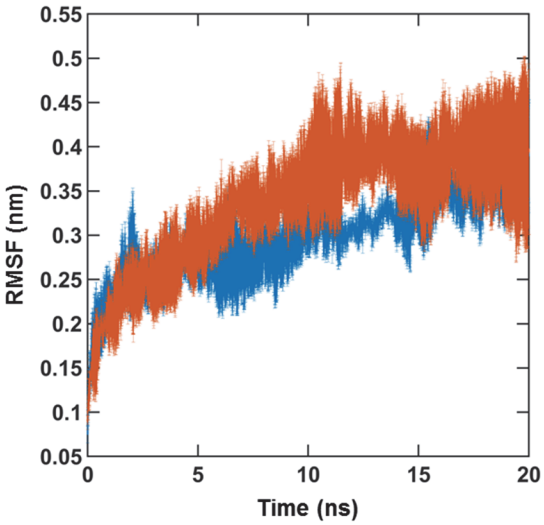

Figure S2. Stabilization of dynamics. RMSF for active (blue) and inactive (orange) MMP1 without ligands at 22 °C.

Calculation of entropy. Entropy can quantify randomness, and there are several measures of entropy, such as Boltzmann and Shannon entropies (1). For patterns, Shannon entropy quantifies randomness in diverse fields. First, we calculated the Gray-Level Co-Occurrence Matrix (GLCM) (1) from the two-dimensional correlation plots. Then, we defined the Shannon entropy (2) from the GLCM. We describe the steps for calculating the Shannon entropy in **Figure S3** using an arbitrary 5×5 matrix, where each element of the matrix has a value between 0 and 3, representing a 2-bit pattern. The dimension of GLCM depends on the bit size. For a 2-bit pattern or image, the possible values are 0, 1, 2, and 3, i.e., the GLCM matrix for a 2-bit pattern is 4×4 irrespective of the pattern dimensions.

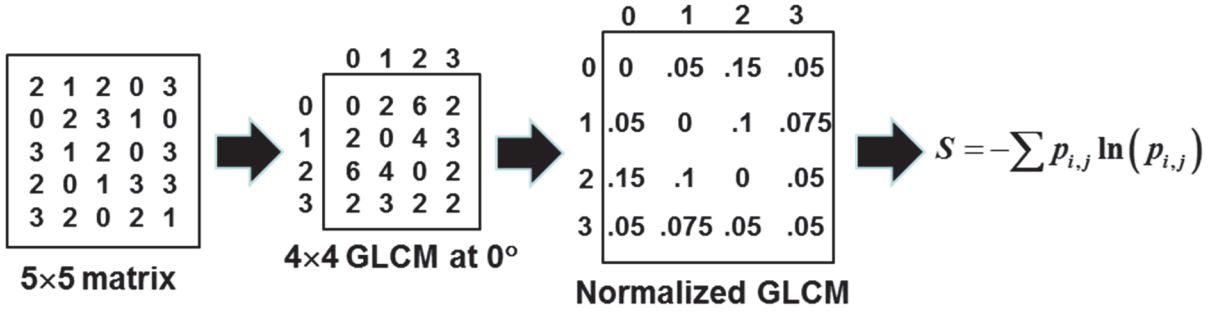

Figure S3. Calculation of Gray-Level Co-Occurrence Matrix for a pattern and corresponding entropy.

Although it is possible to calculate GLCM matrices at 0°, 45°, 90°, and so on (1), we calculated the GLCM at 0° as an example, i.e., we only considered the next neighbor on the left and right. For example, we calculated the (0,0) element of GLCM by finding the value 0s in the original 5×5 matrix and counting the number of times 0s appear on the left or right. The number is 0, and as such, the (0,0) element of GLCM is 0. To calculate the (0,1) element of GLCM, we found the value 0s in the original 5×5 matrix and counted the number of times 1s appear on the left or right.

The number is 2, and as such, the (0,1) element of GLCM is 2. For the (0,2) element of GLCM, we found 2 at the left or right of each 0 in the matrix 6 times. We repeated this process for all the elements of GLCM and obtained the GLCM at 0° (middle panel, **Figure S3**). We calculated the sum of all elements (40 for the GLCM matrix, middle panel, **Figure S3**) and divided each element by the sum to obtain the normalized GLCM (right panel, **Figure S3**). We used $S = -\sum p_{i,j} \ln\left(p_{i,j}\right)$, where $p_{i,j}$ is the (i,j) element of the GLCM to quantify the Shannon entropy. By definition, we considered $\ln\left(p_{i,j}\right) = 0$ for $p_{i,j} = 0$. For MMP1 conformational dynamics, the dimension of correlation matrices is 366×366 because the MMP1 structure used for simulations has 367 residues. We divided the correlation values between 0 and 1 into 10 bins of width 0.1. As such, the dimension of GLCM for MMP1 dynamics is 10×10. If we divided the correlation values into 20 bins, the GLCM matrix would be 20×20.

Table S3. Best-fit parameters for histograms and correlations in Figure 3 from MD simulations. The fit parameters for pose 3 is "NA" because the distributions do not fit to a sum of two Gaussians.

A Gaussian fit parameters for simulated histograms

$$y = a_1 \times e^{-\frac{(x-b_1)^2}{c_1^2}} + a_2 \times e^{-\frac{(x-b_2)^2}{c_2^2}}$$

Catalytic pocket opening						
	Active MMP1			Inactive MMP1		
	Pose 1	Pose 2	Pose 3	Pose 1	Pose 2	Pose 3
a1	1.69±0.04	1.92±0.04	NA	0.43±0.04	0.95±0.07	NA
b1/S1 (nm)	2.52±0.01	2.50±0.01	NA	2.26±0.02	2.67±0.01	NA
c1	0.17±0.01	0.18±0.01	NA	0.18±0.02	0.12±0.01	NA
a2	2.24±0.06	0.80±0.04	NA	1.53±0.03	1.63±0.04	NA
b2/S2(nm)	2.93±0.01	3.14±0.01	NA	2.81±0.01	2.85±0.01	NA
c2	0.10±0.01	0.24±0.01	NA	0.31±0.01	0.28±0.01	NA

Interdomain distance						
	Active MMP1			Inactive MMP1		
	Pose 1	Pose 2	Pose 3	Pose 1	Pose 2	Pose 3
a1	2.76±0.02	2.12±0.04	0.43±0.08	2.03±0.02	1.93±0.04	0.71±0.01
b1/S1 (nm)	4.22±0.01	3.96±0.01	3.67±0.01	4.33±0.01	4.23±0.01	4.30±0.02
c1	0.16±0.01	0.15±0.01	0.10±0.02	0.26±0.01	0.16±0.01	0.55±0.02
a2	0.89±0.03	1.16±0.02	1.53±0.03	0.20±0.03	1.39±0.02	0.95±0.04
b2/S2(nm)	4.49±0.01	4.26±0.01	3.94±0.01	4.74±0.02	4.53±0.01	4.83±0.01
c2	0.14±0.01	0.22±0.01	0.32±0.01	0.15±0.02	0.19±0.01	0.20±0.01

B Linear correlation fit parameters

$$y_i = b_0 + b_1 \times x_i$$

Linear correlation between interdomain distance and catalytic pocket opening						
	Active MMP1			Inactive MMP1		
	Pose 1	Pose 2	Pose 3	Pose 1	Pose 2	Pose 3
b0	0.51±0.06	3.67±0.07	1.76±0.04	5.45±0.05	1.77±0.04	5.78±0.03
b1	0.51±0.02	-0.24±0.02	0.33±0.01	-0.62±0.01	0.24±0.01	-0.63±0.01

MD simulations with zinc and calcium ions in Figure S4. PDB files for active and inactive MMP1 were created by replacing A219 with E219 and Q219, respectively, using Pymol's mutagenesis function. We derived the force field parameters of the metal amino acid complexes (four calcium coordinating sets of amino acids and two zinc coordinating sets of amino acids) following the AmberTools protocol for Developing Non-standard Parameters using the bonded metal model (3). In brief, MCPB.py acts as a wrapper around several tools designed to derive bonded (distance, angular, and dihedral) force constants and equilibrium values along with atomic Restrained Electrostatic Partial (RESP) charges. We checked the amino acids and metal ions in each PDB to ensure consistency with the AMBER naming scheme. Hydrogen atoms were added using the AmberTool reduce function, then mol2 files for non-standard residues were generated using Antechamber, and finally, frcmod files were created using parmchk2. The electronic structure optimization and hessian vibrational frequency calculations were performed using GAMESS US (4) at the B3LYP/6-31G, and level and the Amber ff14SB force field were used to optimize hydrogen atoms. The Seminario method was selected to generate the force field parameters. The coordination between the side chains and the metals was visually checked, and ParmEd was used to check the metal site parameters before finally converting the resulting coordination and topology files to Gromacs compatible files.

Table S4. Best-fit parameters for correlations in Figure 4 from MD simulations.

Exponential fit parameters for correlations

$$C_{\tau} = d \times \exp^{-e \times \tau} + f$$

	MMP1 without ligands	
	Active	Inactive
d	0.40±0.01	0.19±0.01
e	0.01±0.01	0.01±0.01
f	0.55±0.01	0.77±0.01

After parameterization of the side chain-metal coordination complexes and encoded atomic structures using the Amberff14SB force field, each system was centered in a cubic unit cell. We solvated each system with tip3p water molecules and neutralized the overall charge of each system using counter ions.

All-atom Molecular Dynamics simulations were performed using Gromacs (version 2020.5) with the leap-frog algorithm for integrating Newton's equations of motion and a time-step of 0.002 ps. Hydrogen bonds were constrained using the LINCS constraint algorithm with a LINCS iteration of 1 and LINCS order of 4. The neighborhood list was updated every ten steps (20 fs) using the Verlet cut-off scheme, and the unit cell used periodic boundary conditions in all three dimensions. Electrostatic charges were treated using the Particle Mesh Ewald scheme with an order of 4 and modified using a potential shift. The temperature was controlled using the velocity rescale scheme with a time constant of 0.1 ps and fluctuated around a mean of 295 K. Pressure was set to one bar and controlled using isotropic scaling of the unit cell with the Berendsen pressure coupling before switching to the Parrinello-Rahman scheme; both schemes updated every 2 ps. A 1 nm short-range LJ-6-12 cut-off was applied and modified using a potential shift. Finally, velocities and coordinates were collected every 20 ps.

Each system was relaxed using the following steps. First, having solvated and neutralized the total charge, each system was subjected to an energy minimization procedure using the steepest gradient

descent algorithm. The system was left to adjust until all forces were below 1000 kJ/mol. Position restraints of 1000 kJ/mol were then applied to all MMP1 and Aβ heavy atoms. Each system evolved over 50 ps using the NVT ensemble (fixed number of particles, volume, and temperature) with a velocity-rescaling thermostat coupling. The coordinates and velocities from the final frame were preserved, and the system evolved for a further 50 ps using the NPT ensemble (fixed number of particles, pressed and temperature) with the Berendsen pressure coupling whilst maintaining position restraints. Finally, final frame velocities and coordinates were used to perform a production run of 100 ns, having removed all position restraints and switching to the Parrinello-Rahman pressure coupling. For each pose, the series of equilibrium steps and the production simulation were repeated three times using random initial atomic velocities.

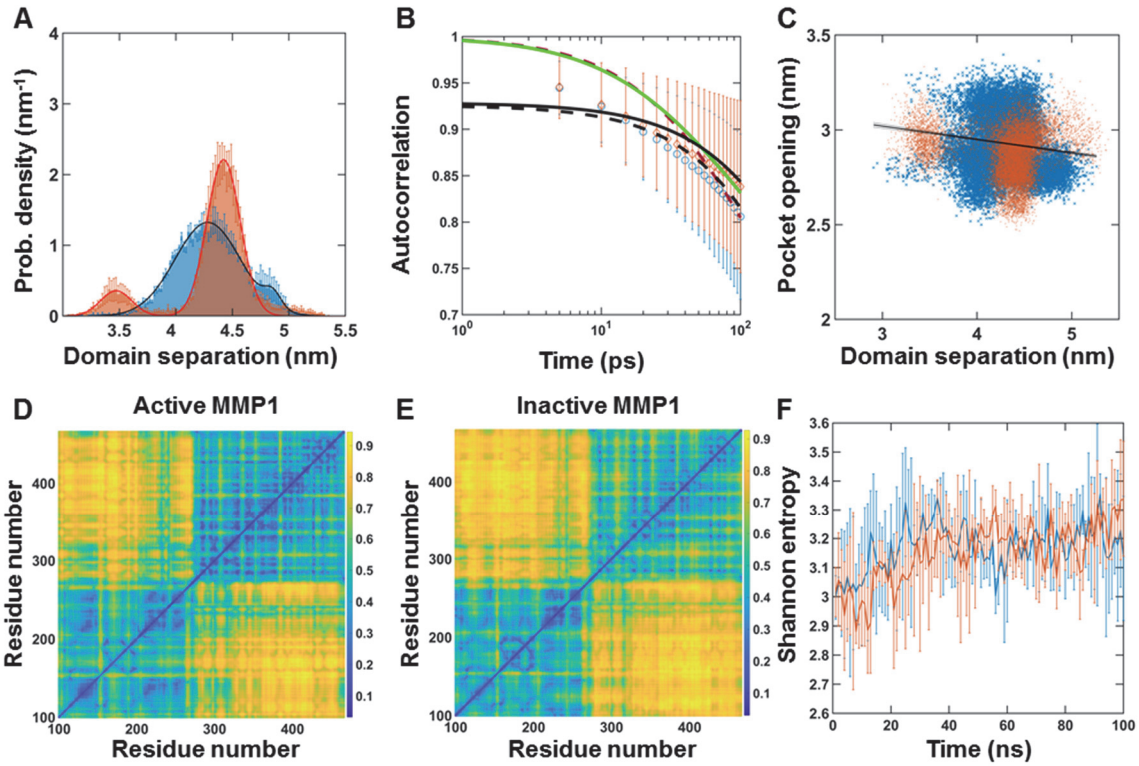

Figure S4. All-atom MD simulation of MMP1 interdomain dynamics of pose 3 with two zinc and four calcium ions at 22 °C. (**A**) Area-normalized histograms of simulated interdomain distance (active MMP1: blue; inactive MMP1: orange) with best fits to a sum of two Gaussians (solid line). (**B**) Autocorrelations of simulated interdomain distance (active: blue; inactive: orange) with fits to exponentials (active MMP1: dashed black line; inactive MMP1: solid black line) and power law fit (active MMP1: dashed red line; inactive MMP1: solid green line). (**C**) Linear correlation plots of catalytic pocket opening and interdomain distance. (**D**) and (**E**) Mean correlations between each residue pair for active and inactive MMP1, respectively. Correlations are normalized between 0 (blue) and 1 (yellow). Yellowish colors indicate higher correlations. (**F**) Shannon entropy calculated from correlation plots for active (S= 3.17±0.08, mean±SEM) and inactive (S= 3.15±0.10, mean±SEM). For best-fit parameters, see **Table S5**. All simulations were repeated three times to calculate error bars.

All electronic structure calculations, parameters, and molecular dynamics topology and coordinates files can be found at https://github.com/acnash/MMP1

Table S5. Best-fit parameters for simulated histograms and autocorrelations in Figure S4.

A Gaussian fit parameters for simulated histograms

$$y = a_1 \times e^{-\frac{(x-b_1)^2}{c_1^2}} + a_2 \times e^{-\frac{(x-b_2)^2}{c_2^2}}$$

MMP1 without ligands		
	Active	Inactive
a1	1.32±0.01	0.36±0.02
b1/S1 (nm)	4.28±0.01	3.47±0.01
c1	0.41±0.01	0.20± 0.01
a2	0.21±0.02	2.21±0.02
b2/S2 (nm)	4.86±0.01	4.42±0.01
c2	0.11±0.01	0.21±0.01

B Exponential fit parameters for correlations

$$C_\tau = d \times \exp^{-e \times \tau} + f$$

MMP1 without ligands		
	Active	Inactive
d	0.41±0.01	0.29±0.01
e	0.002±0.001	0.004±0.001
f	0.52±0.01	0.64±0.01

C Linear correlation fit parameters

$$y_i = b_0 + b_1 \times x_i$$

MMP1 without ligands		
	Active	Inactive
b0	3.23±0.02	2.98±0.01
b1	-0.07±0.01	-0.03±0.01

175

176

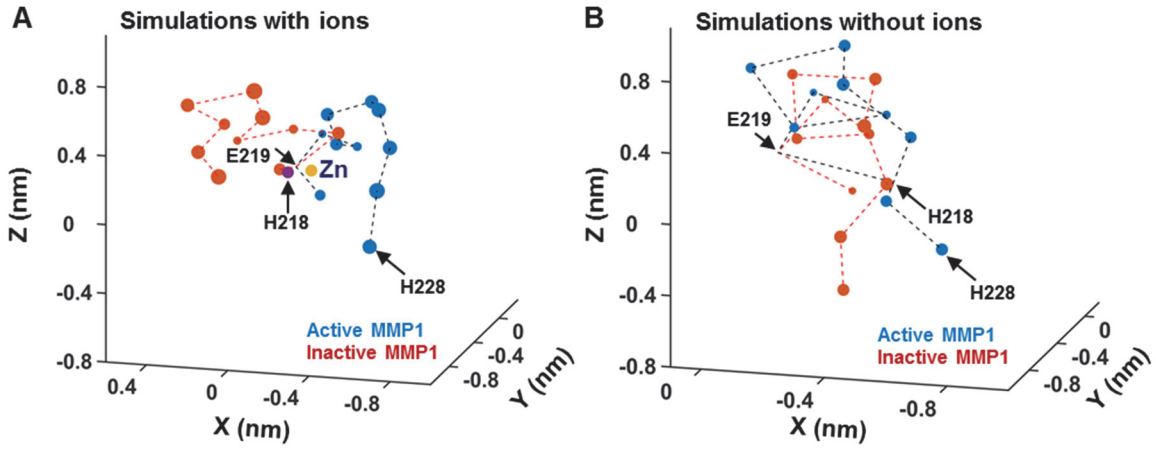

Figure S5. Configuration of the MMP1 catalytic motif at 22 °C with and without metal ions. (**A**) Three-dimensional configurations of the catalytic motif for active MMP1 (blue) and inactive MMP1 (orange) at 22 °C with metal ions. (**B**) Three-dimensional configurations of the catalytic motif residues for active MMP1 (blue) and inactive MMP1 (orange) at 22 °C without considering metal ions during MD simulations.

References

1. Haralick, R. M., K. Shanmugam, and I. H. Dinstein. 1973. Textural features for image classification. IEEE Transactions on systems, man, and cybernetics(6):610-621.
2. Shannon, C. E. 1948. A mathematical theory of communication. The Bell system technical journal 27(3):379-423.
3. Li, P., and K. M. Merz Jr. 2016. MCPB. py: a python based metal center parameter builder. ACS Publications.
4. Barca, G. M., C. Bertoni, L. Carrington, D. Datta, N. De Silva, J. E. Deustua, D. G. Fedorov, J. R. Gour, A. O. Gunina, and E. Guidez. 2020. Recent developments in the general atomic and molecular electronic structure system. The Journal of chemical physics 152(15):154102.