Biophysical Journal, Volume *119*

Supplemental Information

Allosteric Communications between Domains Modulate the Activity of Matrix Metalloprotease-1

Lokender Kumar, Anthony Nash, Chase Harms, Joan Planas-Iglesias, Derek Wright, Judith Klein-Seetharaman, and Susanta K. Sarkar

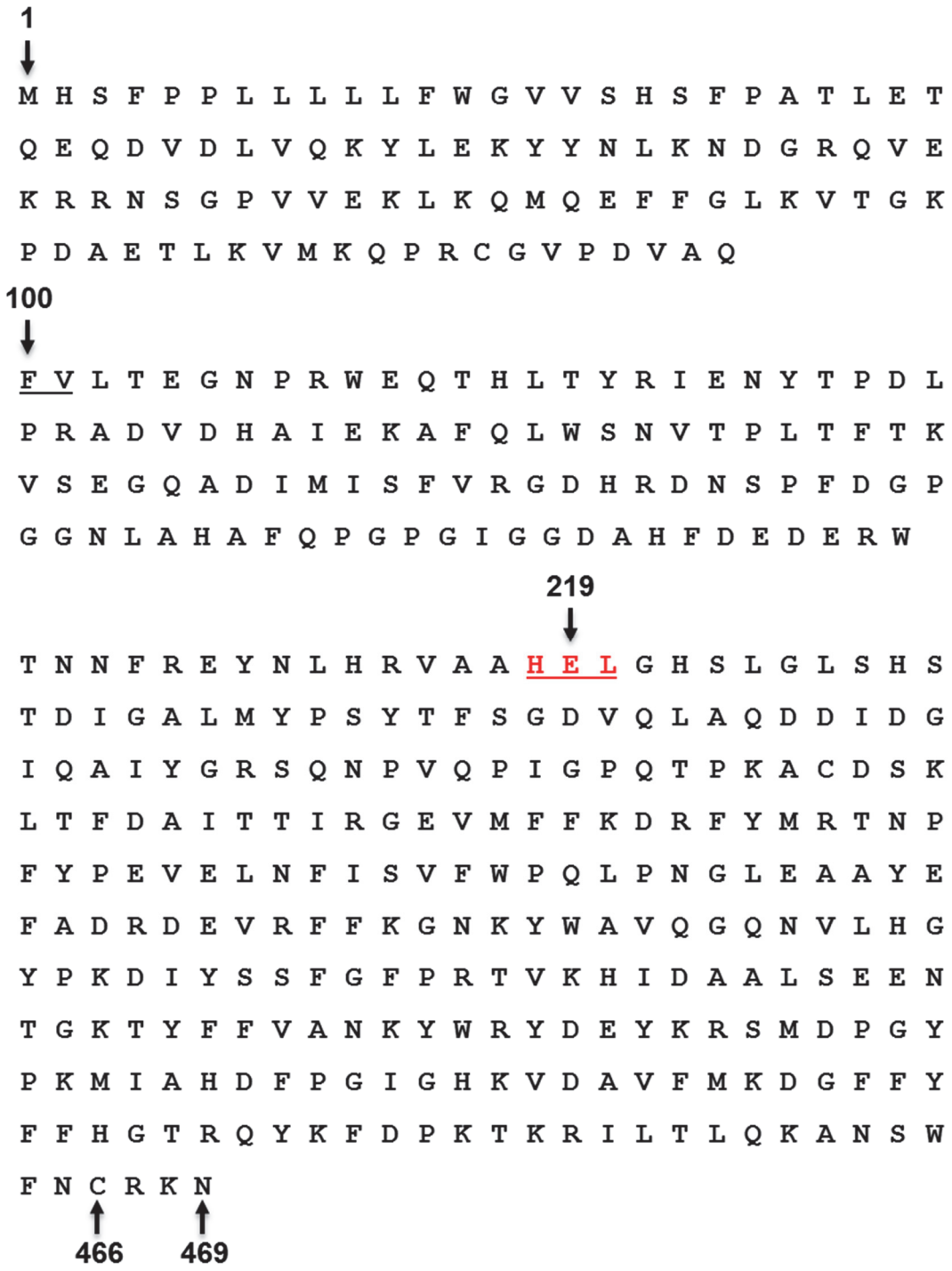

FIGURE S1. MMP1 sequence for experiments and simulations. Full-length MMP1 with the pro domain has 469 residues (M1-N469) (1). Trypsin activates MMP1 by cleaving the F-V bond. Active and inactive MMP1 used in single molecule experiments has the same sequence (V101-N469) except the residue E219 at the catalytic site. For inactive MMP1, there is a single point mutation E219Q. For simulations, we used PDB: 4AUO with the MMP1 sequence between F100 and C466. S142C, S366C, and E219Q mutations were modified in 4AUO to simulate the dynamics of active and inactive MMP1.

AFM and SEM images of MMP-treated fibrils. We made reconstituted fibrils using RatCol® – Rat Tail Type-1 Collagen from Advanced Biomatrix by neutralizing the solution of type-1 collagen monomers and incubating at 37°C according to the manufacturer's protocol.

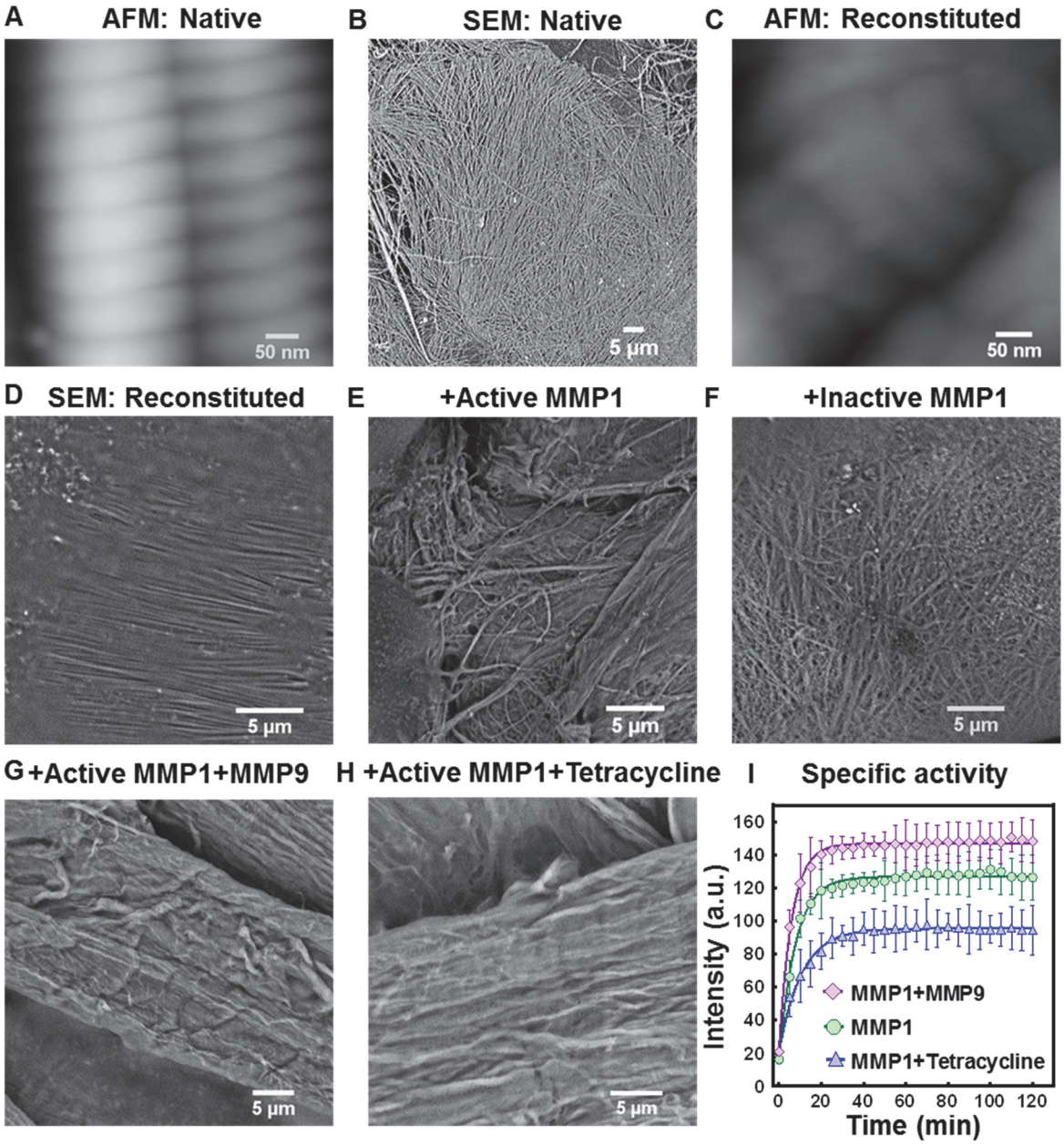

FIGURE S2. Activity of MMP1. (**A-H**) Surface morphology of MMP1-treated type-1 collagen fibrils. (**I**) Fluorescence from degraded peptide substrate, MCA-Lys-Pro-Leu-Gly-Leu-DPA-Ala-Arg-NH_2, as a function of time for MMP1 (green circle), MMP1+MMP9 (magenta diamond), and MMP1+tetracycline (blue triangle). Solid lines are respective best fits to $y = a - b \times exp(-kt)$. The error bars are the standard deviations (SD) of 3 technical repeats. After calibration, the specific activities are ~1000 pmol/min/µg (MMP1), ~1300 pmol/min/µg (MMP1+MMP9), ~760 pmol/min/µg (MMP1+tetracycline).

We used native fibrils extracted from rat tails. We prepared three 200 µL reaction mixtures by adding 100 µL of 1 mg/mL active MMP1 with 1) 100 µL of protein buffer (50 mM Tris, 100 mM NaCl, pH 8.0), 2) 100 µL of 1 mg/mL MMP9, and 3) 100 µL of 100 µg/mL tetracycline. We prepared another 200 µL reaction mixture by adding 100 µL of 1 mg/mL inactive MMP1 with 100 µL of protein buffer. The slides were washed three times with 1 mL sterile distilled water to remove buffer precipitates. After washing, we air-dried the slides at room temperature. We incubated fibrils with these four reaction mixtures at 37°C for 4 h before imaging with AFM (Asylum MFP-3D Atomic Force Microscope) and SEM (Phenom Pro-Scanning Electron Microscope) (**Fig. S2, A-H**).

Specific activity of MMP1 in the presence of MMP9 and tetracycline. We used a synthetic peptide substrate, MCA-Lys-Pro-Leu-Gly-Leu-DPA-Ala-Arg-NH_2 (2), to measure the specific activity of MMP1 following the protocol described in a previous publication (1). MMP9 enhances the activity of MMP1, whereas tetracycline reduces the activity of MMP1 (**Fig. S2, I**).

Best-fit parameters for the experimental histograms and autocorrelations. We fitted two Gaussians to conformational histograms using the following equation:

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

where a's are amplitudes, b's are centers, and c's are widths of Gaussians. Best-fit parameters are in **Table S1,** with the centers highlighted in yellow. The error bars are the standard errors of the mean (SE). The parameters b1 and b2 are states S1 and S2, respectively.

Table S1. Best-fit parameters for Gaussian fits to the histograms in Fig. 2, A-C.

	MMP1 without ligands		MMP1 with MMP9		MMP1 with tetracycline	
	Active	Inactive	Active	Inactive	Active	Inactive
a1	2.95±0.08	6.06±0.02	4.97±0.54	2.59±0.09	6.48±0.04	0.38±0.03
b1/S1	0.44±0.01	0.56±0.01	0.43±0.01	0.52±0.01	0.51±0.01	0.42±0.01
c1	0.08±0.01	0.09±0.01	0.08±0.01	0.10±0.01	0.08±0.01	0.05±0.01
a2	3.50±0.06	0.25±0.05	2.91±0.74	5.75±0.08	0.39±0.04	5.48±0.01
b2/S2	0.55±0.01	0.68±0.01	0.48±0.01	0.54±0.01	0.55±0.01	0.50±0.01
c2	0.09±0.01	0.06±0.01	0.06±0.01	0.05±0.01	0.15±0.01	0.10±0.01

We calculated the autocorrelations of conformational fluctuations by using the following equation:

$$C_\tau = \frac{1}{N-\tau}\sum_{t=0}^{N-\tau}\left\{I(t)-\frac{1}{N-\tau}\sum_{t'=0}^{N-\tau}I(t)\right\}\times\left\{I(t+\tau)-\frac{1}{N-\tau}\sum_{t'=\tau}^{N}I(t)\right\}$$

where C_τ is the autocorrelation at lag number τ, N is the total number of data points in a FRET trajectory, and $I(t)$ is the FRET value at time point t.

We normalized autocorrelations by dividing autocorrelations at each lag by $C_{\tau=0}$. We fitted autocorrelations between $\tau = 1$ and $\tau = 1000$ to both power law and exponential. We used a form of Pareto distribution (3) as power law that satisfies the boundary conditions of our calculated autocorrelations, i.e., $C_{\tau=0} = 1$ at $t = 0$ and $C_{\tau=\infty} = 0$ at $t = \infty$. The power law and exponential functions used are:

$$C_\tau = (a \times \tau + 1)^{-b}$$

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

Generally, the power law does not fit the autocorrelations of trajectories obtained from experiments and two-state Poisson process simulations. We have included them for comparison because power law fits better with the dynamics simulated using all-atom MD (**Fig. 3**). Best-fit parameters for exponential fits to the experimental autocorrelations are in **Table S2**. The parameter e gives the sum k1+k2. The error bars are SEs. The calculated rates from the histograms and autocorrelations are in **Table S3**.

Table S2. Best-fit parameters for exponential fits to the autocorrelations in Fig. 2, D-F.

	MMP1 without ligands		MMP1 with MMP9		MMP1 with tetracycline	
	Active	Inactive	Active	Inactive	Active	Inactive
d	0.57±0.01	0.34±0.01	0.32±0.01	0.13±0.01	0.24±0.01	0.42±0.01
e	0.13±0.01	0.08±0.01	0.03±0.01	0.09±0.01	0.08±0.01	0.05±0.01
f	-0.02±0.01	-0.01±0.01	0.01±0.01	0.01±0.01	-0.01±0.01	-0.01±0.01

Table S3. Calculated kinetic rates (s^{-1}) of interconversion between the states, S1 and S2.

	MMP1 without ligands		MMP1 with MMP9		MMP1 with tetracycline	
	Active	Inactive	Active	Inactive	Active	Inactive
k1	0.0764	0.0022	0.0208	0.0473	0.0081	0.0483
k2	0.0543	0.0759	0.0092	0.0427	0.0719	0.0017

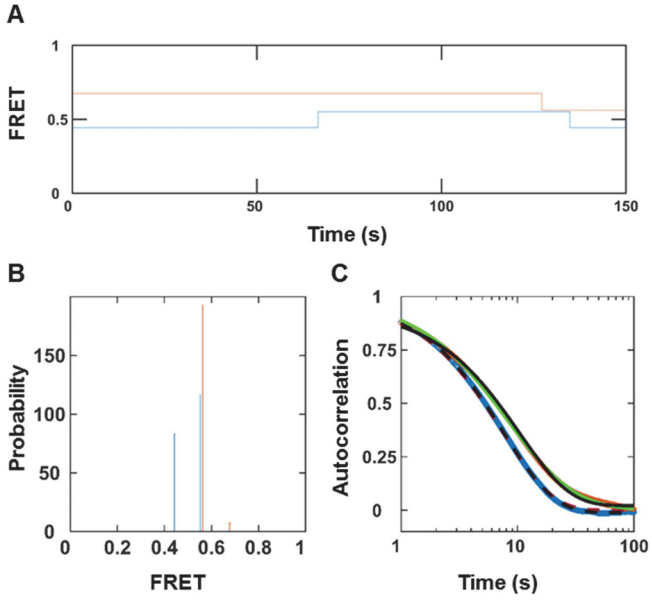

FIGURE S3. MMP1 inter-domain dynamics as a Poisson process without noise. (**A**) An example of simulated two-state FRET trajectory without noise for active (blue) and inactive (orange) MMP1. (**B**) Histograms of the recovered FRET values with bin size=0.005. (**C**) Autocorrelations of simulated trajectories recover the sum, k1+k2, from exponential fits (active: dashed black line; inactive: solid black line). Note that power law does not fit the autocorrelations with noise (**Fig. 3**). However, power law fits the autocorrelations without noise (active: dashed red line; inactive: solid green line). The error bars are the SEs for histograms and autocorrelations and are too small to be seen.

Best-fit parameters for two-state Poisson process simulations. With and without noise, the peak positions (**Table S4**) determined from **Fig. 3, B** and **Fig. S3, B**, respectively, agreed with the input FRET values. The exponential fits to autocorrelations recovered the expected sum of k1 and k2 (**Table S5**, right panel). Note that the power law reasonably fits the two-state Poisson process simulation without noise (**Fig. S3, C**). The error bars are the SEs.

Table S4. Gaussian fit parameters for two-state simulations in Fig. 3 and Fig. S3.

	Without noise		With noise	
	Active	Inactive	Active	Inactive
a1			2.92±0.02	6.07±0.01
b1/S1	0.44	0.56	0.44±0.01	0.56±0.01
c1			0.08±0.01	0.09±0.01
a2			3.54±0.01	0.31±0.0.01
b2/S2	0.55	0.68	0.55±0.01	0.68±0.01
c2			0.09±0.01	0.06±0.01

With noise, the peak positions and autocorrelation decay rates recovered the input parameters from Gaussian and exponential fits, respectively (**Table S4** and **Table S5**). The error bars are the SEs.

Table S5. Fit parameters for two-state simulation autocorrelations in Fig. 3 and Fig. S3.

$$C_\tau = (a \times \tau + 1)^{-b}$$

	Without noise	
	Active	Inactive
a	0.001±0.001	0.04±0.01
b	1.3±5.4	3.16±0.01

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

	Without noise		With noise	
	Active	Inactive	Active	Inactive
d	1.02±0.01	0.10±0.01	0.42±0.01	0.09±0.01
e	0.13±0.01	0.08±0.01	0.13±0.01	0.08±0.01
f	-0.01±0.01	-0.01±0.01	-0.01±0.01	-0.01±0.01

Best-fit parameters for all-atom MD simulations. For all-atom MD simulations, we calculated the distance between S142C and S366C in nm (**Table S6**). The error bars are the SEs.

Table S6. Best-fit parameters for Gaussian fits to MD histograms in Fig. 4.

	Collagen restrained		Collagen unrestrained	
	Active	Inactive	Active	Inactive
a1	1.94±0.01	2.91±0.70	0.08±0.03	2.63±1.43
b1/S1	4.79±0.01	4.68±0.02	4.08±0.02	4.62±0.01
c1	0.20±0.01	0.16±0.01	0.06±0.03	0.19±0.01
a2	0.10±0.01	0.60±0.72	2.23±0.01	0.28±1.10
b2/S2	4.41±0.01	4.83±0.10	4.32±0.01	4.76±0.55
c2	0.16±0.01	0.16±0.04	0.25±0.01	0.23±0.18

Table S7. Best-fit parameters for power law fits to MD autocorrelations in Fig. 4.

	Collagen restrained		Collagen unrestrained	
	Active	Inactive	Active	Inactive
a	0.83±0.03	0.50±0.02	0.41±0.01	0.42±0.01
b	0.03±0.01	0.12±0.01	0.08±0.01	0.11±0.01

Molecular docking of tetracycline with MMP1. We docked tetracycline (ligand) with MMP1 (receptor) using Autodock Vina. We used the crystal structure 4AUO of MMP1 for docking. First, we removed the triple-helical collagen to expose the binding sites of MMP1 and selected a grid box by specifying the number of points and spacing in x, y, and z directions. The number of points in x, y, and z dimensions in MMP1 was 36, 52, and 54. Also, the spacings (Angstrom) in x, y, and z directions were 16.8, 120.1, and 23.5. The docking poses revealed two clusters of docking: one at the active site and the other between the catalytic and hemopexin domains (**Fig. S4, A**).

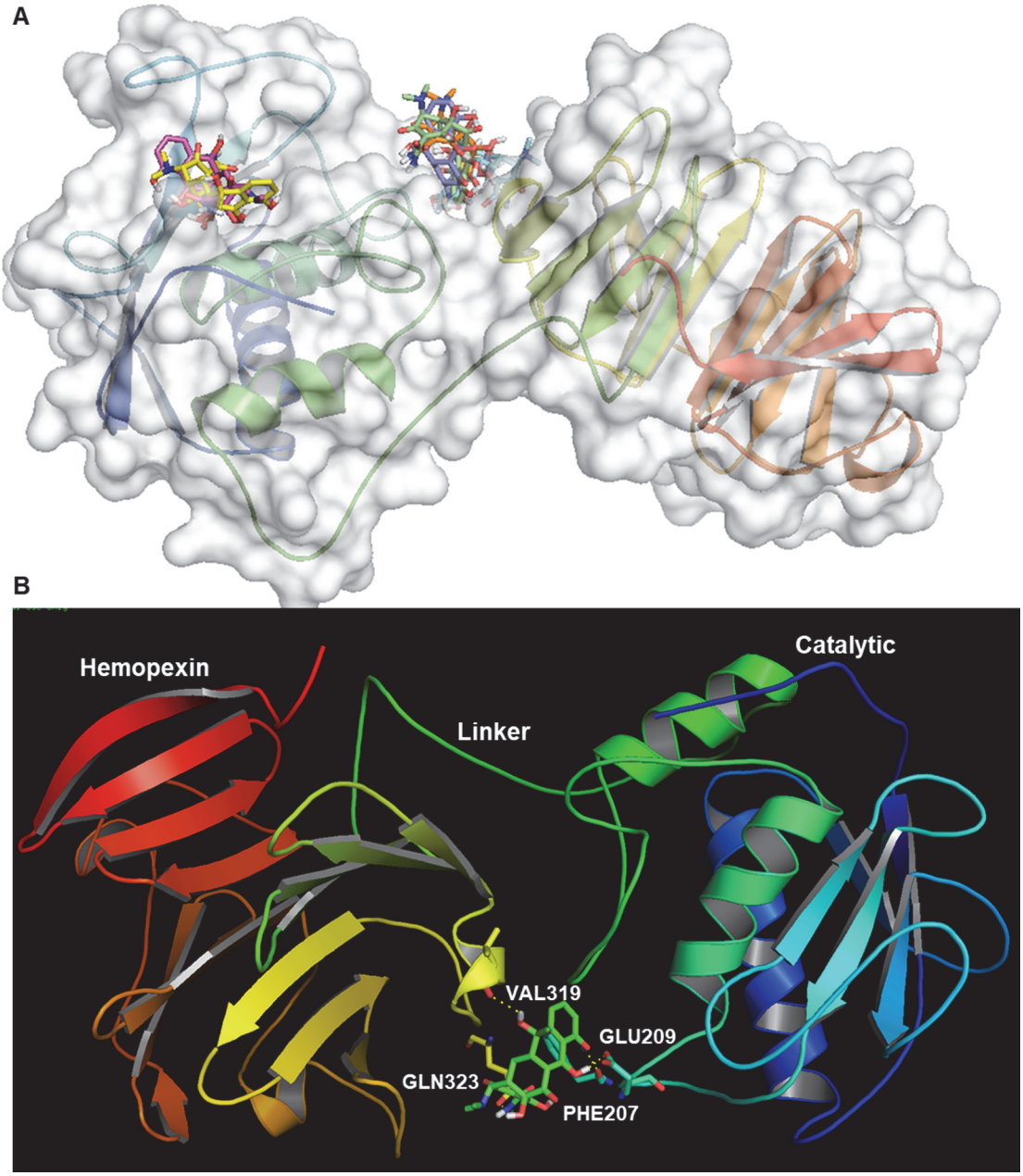

FIGURE S4. Molecular docking of tetracycline with MMP1. (**A**) Two clusters of docking poses near the catalytic site and linker region. (**B**) The docking pose with the highest docking score of -7.5 kcal/mol. Tetracycline forms two hydrogen bonds with both the catalytic and hemopexin domains, which may restrict inter-domain dynamics.

Tetracycline formed two hydrogen bonds with PHE207 and GLU209 in the catalytic domain, and two hydrogen bonds with VAL319 and GLN323 in the hemopexin domain (**Fig. S4, B**). The presence of these hydrogen bonds suggests that tetracycline may restrict inter-domain motion of both active and inactive MMP1. The presence of these hydrogen bonds may explain the similar conformational histograms observed in experiments (**Fig. 2, A-C**).

Root-mean-squared fluctuation of amino acids from all-atom MD simulations. We quantified the conformational changes for active (E219) and inactive (E219Q) from simulations with the collagen backbone unrestrained. For both active and inactive MMP1, we modified 4AUO to include S142C and S366C mutations. Also, 4AUO has Q at 219, which we changed to E to simulate the effect of active MMP1. Note that 4AUO has amino acids between F100 and C466, which means 4AUO has the pro domains (amino acids between N1 and Q99) removed for both active and inactive MMP1. A sampling of simulated trajectories showed differences in the side-chain fluctuations and conformational space explorations. To quantify the difference in variations, we calculated the root-mean-squared fluctuation (RMSF) for active and inactive MMP1.

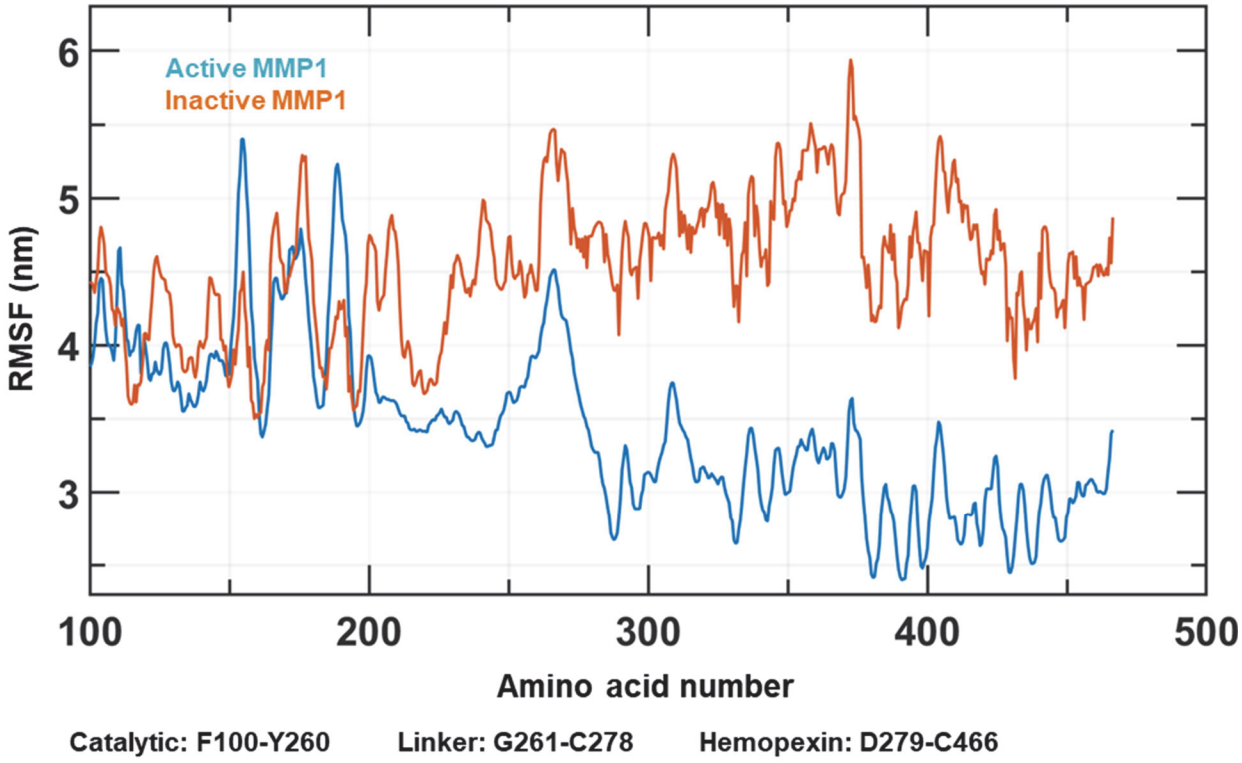

FIGURE S5. Analysis of fluctuations. The RMSF of amino acids for active (blue) and inactive (orange) MMP1 measured with respect to the energy-minimized structure before any simulation integration steps. During the 700 ns-long simulations with time step 2 fs, we saved data every 5 ps with the collagen unrestrained. The catalytic domain, the linker, and the hemopexin domains are defined roughly in the ranges F100-Y260, G261-C278, and D279-C466, respectively. The means ± SDs of RMSF across all the amino acids are 3.6 ± 0.7 nm for active and 4.6±0.4 nm for catalytically inactive MMP1.

We calculated RMSF of MMP1 from the NPT simulation with the collagen backbone unrestrained using the GROMACS gmx toolset. We used the initial energy-minimized conformation before MD integration as a reference structure. In both active and inactive simulations, the catalytic domains showed a similar degree of side-chain fluctuation. However, the linker and hemopexin

domains showed significant differences in the side-chain fluctuations. The changes in the hemopexin domain due to the mutation E219Q in the catalytic domain suggest communications between MMP1 domains. Such allosteric communications are in contrast to a previous NMR-based report on the effect of E219A on MMP12, another of the MMP family, interacting with type-V collagen that suggested no influence outside the catalytic cleft (4). In other words, allosteric communications depend on MMPs and the type of collagen. Further, the higher SD of fluctuations across all amino acids for active MMP1 in comparison to inactive MMP1 is consistent with the wider experimental histograms for active MMP1 (**Fig. 2, A-C**). Surprisingly, allosteric communications appear within 700 ns of simulations shown in **Fig. S5**. However, previous studies have shown that single point mutations can result in protein sampling through different regions of the conformational phase space in short simulation times (5). The difference in RMSF may arise because the size differences between the amino acid substitutions may result in a change in van der Waals interaction (6). The differences could also be the result of a change in hydrogen bond configuration, as reported, for example, on the effects of active site mutations in cyp51A protein leading to resistance against therapeutics (7). Further studies are needed to define the mechanism behind the observed RMSF differences.

Principal Component Analysis. Protein dynamics involve a wide range of length and time scales. For example, our MD simulations calculate Angstrom-scale motion with 2 fs time resolution over durations of a few hundred nanoseconds. In contrast, our smFRET measurements determine nanometer-scale movement with 100 ms time resolution over durations of a few hundred seconds. One approach to reveal the connections in such a wide range of scales is to note that only a few modes out of many protein conformations contain the majority of fluctuations. To find these critical (slow) modes, principal component analysis (PCA) is one of the most efficient methods. PCA involves the calculation of a covariance matrix followed by eigenvalue calculation of the covariance matrix to reduce a multidimensional complex set of variables to a few principal components (8).

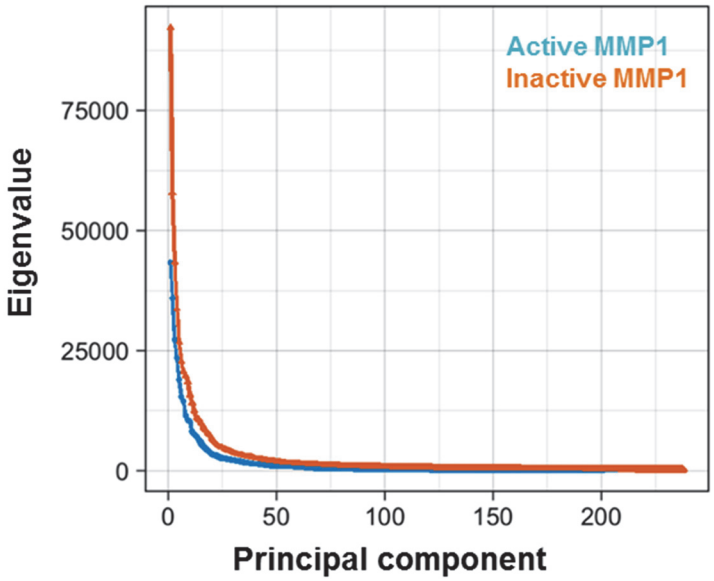

FIGURE S6. Principal component analysis. The distribution of eigenvalues as a function of principal components for active and inactive MMP1. The highest variations in protein dynamics occur in the first twenty principal components.

We performed protein principal component analysis to capture the overall dynamics of active and inactive MMP1 with the collagen unrestrained. We used the alpha-carbons (peptide-backbone carbon adjoining each side chain) in the 700 ns-long simulations. We used the pyPcazip software (9) to compress each trajectory file (0.9 of the original file) and extracted PCA-related metrics using pyPczdump. We generated simulation files (.xtc) without water and removed collagen, ions, and periodic boundary artifacts by treating each simulation with the gmx trjconv GROMACS utility. The majority of fluctuations in MMP1 dynamics falls within the first twenty principal components (**Fig. S6**), which is typical for proteins in general (10).

We analyzed the first four PCAs for active (**Fig. S7**) and inactive MMP1 (**Fig. S8**). Although MMP1 samples conformations around the initial structure, the first four principal components for active MMP1 revealed only one primary populated state (**Fig. S7**). In comparison, the PCA analysis for inactive MMP1 revealed a second populated state (**Fig. S8**), different from the conformation closer to the initial structure. The deviation between the primary populated states compares well with the difference in side-chain fluctuation between active and inactive MMP1.

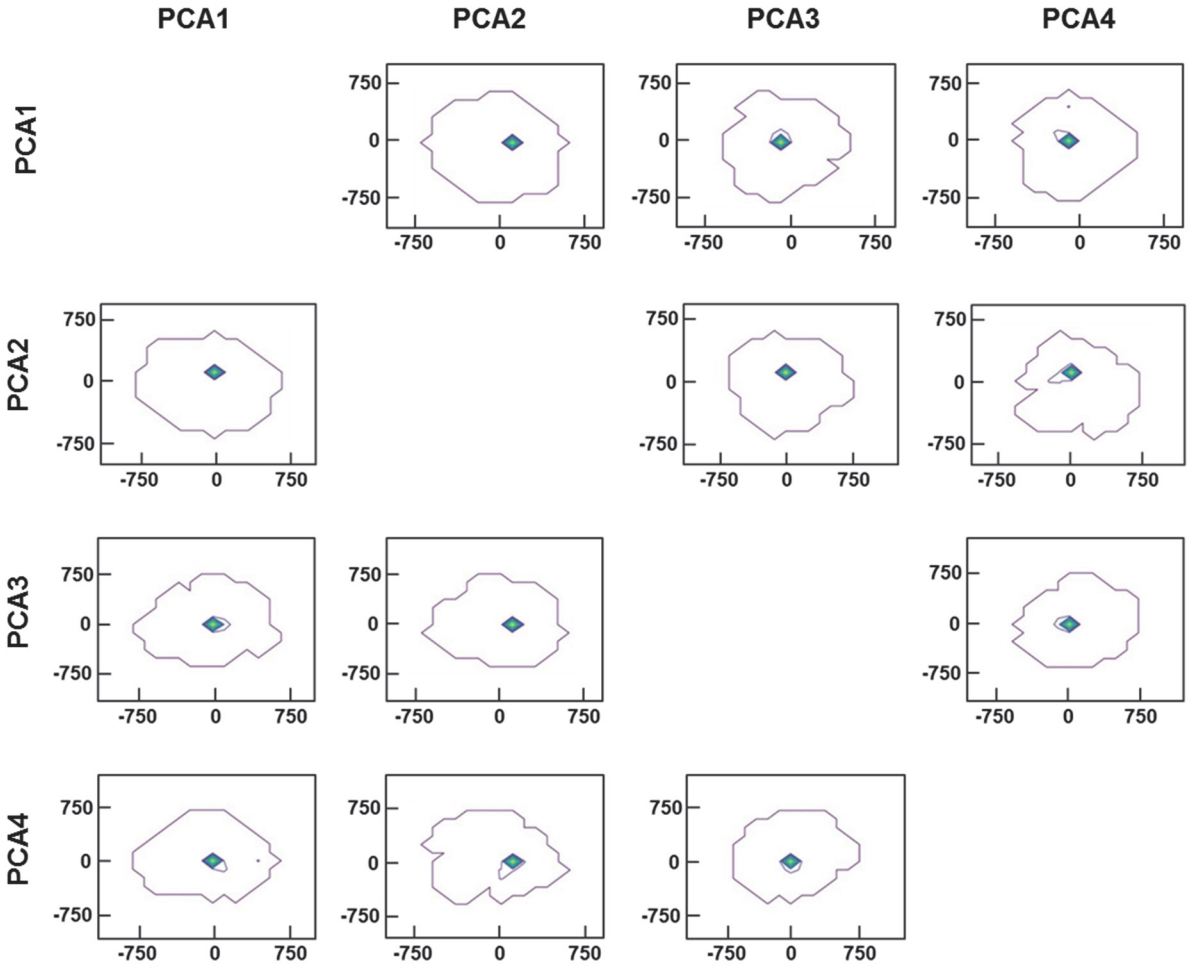

FIGURE S7. Principal component analysis for active MMP1. Conformational sampling plots for active MMP1 projected onto the first four principal components. One primary populated state can be distinguished.

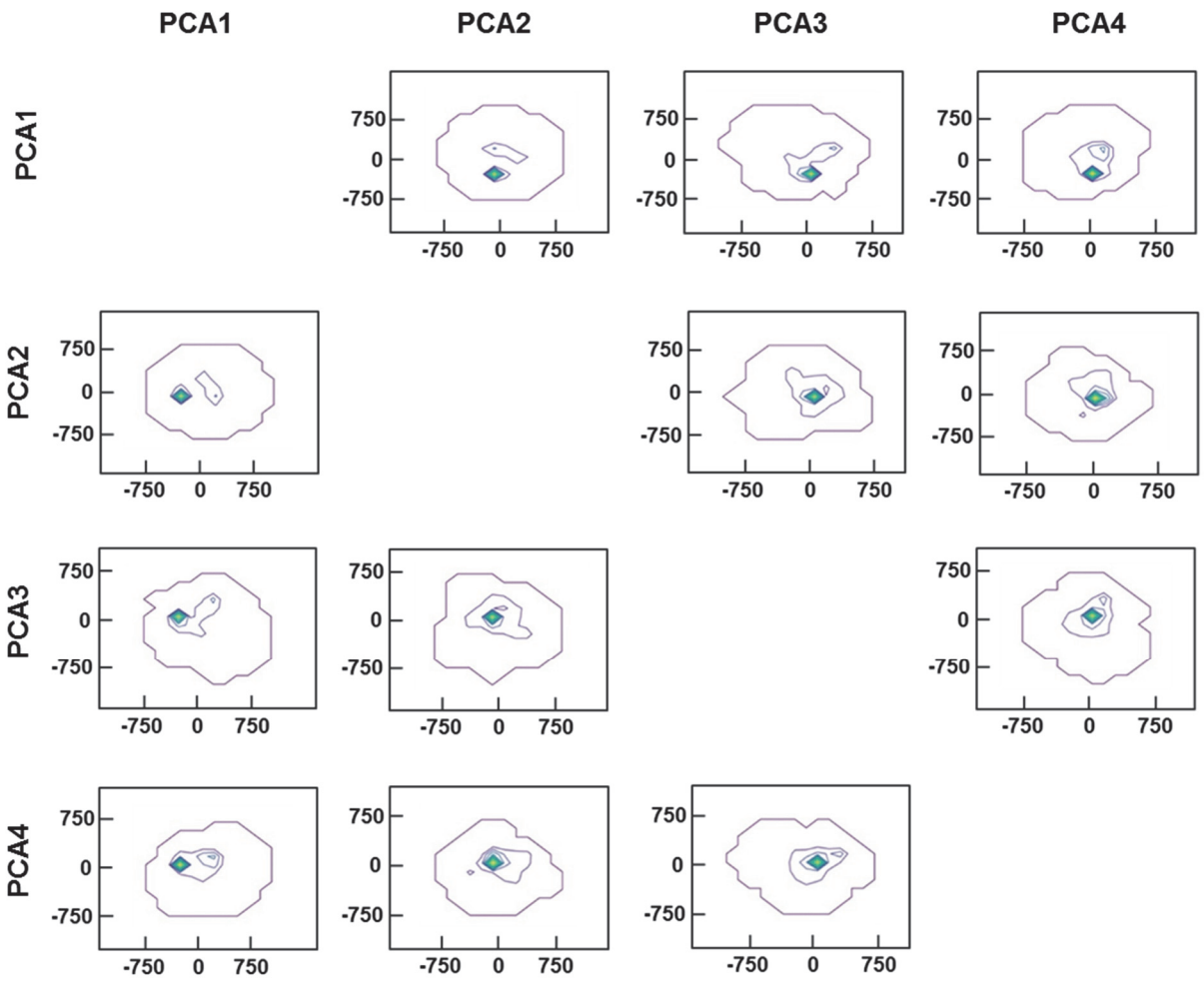

FIGURE S8. Principal component analysis for inactive MMP1. Conformational sampling plots for inactive MMP1 projected onto the first four principal components. The primary populated state accompanies a second less populated state.

References

1. Kumar, L., W. Colomb, J. Czerski, C. R. Cox, and S. K. Sarkar. 2018. Efficient protease based purification of recombinant matrix metalloprotease-1 in E. coli. Protein expression and purification 148:59-67.
2. Neumann, U., H. Kubota, K. Frei, V. Ganu, and D. Leppert. 2004. Characterization of Mca-Lys-Pro-Leu-Gly-Leu-Dpa-Ala-Arg-NH2, a fluorogenic substrate with increased specificity constants for collagenases and tumor necrosis factor converting enzyme. Analytical biochemistry 328(2):166-173.
3. Arnold, B. C. 2014. Pareto distribution. Wiley StatsRef: Statistics Reference Online:1-10.
4. Prior, S. H., T. S. Byrne, D. Tokmina-Roszyk, G. B. Fields, and S. R. Van Doren. 2016. Path to Collagenolysis COLLAGEN V TRIPLE-HELIX MODEL BOUND PRODUCTIVELY AND IN ENCOUNTERS BY MATRIX METALLOPROTEINASE-12. Journal of Biological Chemistry 291(15):7888-7901.
5. Ng, H. W., C. A. Laughton, and S. W. Doughty. 2013. Molecular dynamics simulations of the adenosine A2a receptor: structural stability, sampling, and convergence. Journal of chemical information and modeling 53(5):1168-1178.
6. Loladze, V. V., D. N. Ermolenko, and G. I. Makhatadze. 2002. Thermodynamic consequences of burial of polar and non-polar amino acid residues in the protein interior. Journal of molecular biology 320(2):343-357.
7. Nash, A., and J. Rhodes. 2018. Simulations of CYP51A from Aspergillus fumigatus in a model bilayer provide insights into triazole drug resistance. Medical mycology 56(3):361-373.
8. Maisuradze, G. G., A. Liwo, and H. A. Scheraga. 2009. Principal component analysis for protein folding dynamics. Journal of molecular biology 385(1):312-329.
9. Shkurti, A., R. Goni, P. Andrio, E. Breitmoser, I. Bethune, M. Orozco, and C. A. Laughton. 2016. pyPcazip: A PCA-based toolkit for compression and analysis of molecular simulation data. SoftwareX 5:44-50.
10. David, C. C., and D. J. Jacobs. 2014. Principal component analysis: a method for determining the essential dynamics of proteins. Protein dynamics. Springer, pp. 193-226.