Supporting Information

Theoretical Study of Metal Ion–Induced Modulation of Firefly Bioluminescence Spectra

Jinyu Wang[a], Deping Hu,[b,*] Ya-Jun Liu[a,b,*]

[a] *Key Laboratory of Theoretical and Computational Photochemistry, Ministry of Education, College of Chemistry, Beijing Normal University, Beijing 100875, China*

[b] *Department of Chemistry, Faculty of Arts and Sciences, Center for Advanced Materials Research, Beijing Normal University at Zhuhai 519087, China*

*** Correspondence:** Corresponding Author E-mails: depinghu@bnu.edu.cn; yajun.liu@bnu.edu.cn; ORCID: 0000-0001-7161-1253; 0000-0001-8761-5339

Contents

Computational Methods

Setup of Systems. Because the available crystal structure of *A. vivianii* firefly luciferase lacks the substrate and adopts an open or semi-open conformation, a homology model was constructed. *Photinus pyralis* (*P. pyralis*) firefly luciferase (PDB ID: 4G36, chain B), which shares 79% sequence identity with *A. vivianii*, was selected as the template due to its similar active-site architecture and preserved salt-bridge interactions near the substrate-binding pocket. Missing residues in the template structure were modeled using the ModLoop server prior to homology modeling with Modeller [1, 2]. The resulting *A. vivianii* luciferase models were evaluated based on structural stability and stereochemical quality, and the most reasonable structure was selected and validated using the MolProbity server [3]. Ramachandran plot analysis indicated that 99.6% of residues were located in allowed regions (**Fig S1**), with remaining outliers located far from the active site.

The validated model was aligned with the *P. pyralis* luciferase structure using PyMOL [4], and the substrate D-luciferyl-adenylate (DLSA, **Fig S2**) was transferred into the *A. vivianii* model. DLSA was subsequently converted into **oLu** and AMP to construct the wild-type enzyme–substrate complex (**System I**; **Fig 2a**). Four additional systems (**Systems II–V**) were generated by inserting Ag^{+}, Zn^{2+}, Cd^{2+}, or Hg^{2+} ions between the internal salt bridge (E311–R337) and the external salt bridge (H310–E354), forming metal ion-binding clusters (**Fig 2b**). The coordination environments of these metal ion-binding clusters were parameterized using the MCPB.py module within the AmberTools suite. For all five systems, protonation states of titratable residues (His, Glu, and Asp) were assigned using the H++ server [5]. The Amber ff19SB force field was used for protein residues, while GAFF was applied to **oLu**. Partial charges for **oLu** were derived using the RESP procedure [6] at the HF/6-31G level with Gaussian 09 [7]. Missing force field parameters were generated using parmchk2. Parameters for AMP were taken from the literature. Each system was neutralized with sodium ions and solvated in a rectangular TIP3P water box [8] with a minimum distance of 12 Å between the protein surface and the box boundary.

EEF and IEF Calculations. For EEF calculations, the S_1-state geometries of **oLu** were extracted from the QM region of the optimized QM/MM structures of **Systems I–V**. Each extracted **oLu** molecule was treated as an isolated system in the gas phase and denoted as **Systems *i–v***. Uniform electric fields ranging from −0.08 to +0.08 a.u. were applied independently along the X, Y, and Z axes, and single-point excited-state calculations were performed at the TD-CAM-B3LYP/ma-def2-TZVP level using Gaussian 09.

The IEF generated by the enzyme environment was computed using an in-house Python script. In this approach, the total electric field at a given point on **oLu** was calculated as the vector sum of electrostatic contributions from the surrounding environment, including protein residues, metal ions, AMP, and solvent molecules, according to the formula:

$$\vec{\boldsymbol{E}} = \boldsymbol{k}\sum_{\boldsymbol{i}} \frac{\boldsymbol{q_i}\vec{\boldsymbol{r}}_{\boldsymbol{i}}}{|\vec{\boldsymbol{r}}_{\boldsymbol{i}}|^{\boldsymbol{3}}}$$

where $\vec{\boldsymbol{E}}$ is the total electric field vector at the target point, $\boldsymbol{k}$ is the Coulomb constant, $\boldsymbol{q_i}$ is the partial charge of the *i-th* environmental atom, and $\vec{\boldsymbol{r}}_{\boldsymbol{i}}$ is the displacement vector pointing from the environmental atom *i* to the target point on **oLu**.

Reference

[1] Fiser, A., & Sali, A. (2003). ModLoop: Automated modeling of loops in protein structures. Bioinformatics, *19*(18), 2500–2501. https://doi.org/10.1093/bioinformatics/btg362.

[2] Fiser, A., Do, R. K. G., & Sali, A. (2000). Modeling of loops in protein structures. Protein Science, *9*(9), 1753–1773. https://doi.org/10.1110/ps.9.9.1753.

[3] Davis, I. W., Leaver-Fay, A., Chen, V. B., Block, J. J., Kapral, G. J., Wang, X., et al. (2007). MolProbity: All-atom contacts and structure validation for proteins and nucleic acids. Nucleic Acids Research, *35*(suppl_2), W375–W383. https://doi.org/10.1093/nar/gkm216.

[4] Delano, W. L. (2002). PyMOL: An open-source molecular graphics tool. CCP4 Newsletter on Protein Crystallography, *40*, 82–92.

[5] Anandakrishnan, R., Aguilar, B., & Onufriev, A. V. (2012). H++ 3.0: Automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations. Nucleic Acids Research, *40*(W1), W537–W541. https://doi.org/10.1093/nar/gks375.

[6] Cieplak, P., Cornell, W. D., Bayly, C., & Kollman, P. A. (1995). Application of the multimolecule and multiconformational RESP methodology to biopolymers - charge derivation for DNA, RNA, and proteins. Journal of Computational Chemistry, *16*(11), 1357–1377. https://doi.org/10.1002/jcc.540161106.

[7] Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., et al. (2009). *Gaussian 09*, *Revision D.01*. Gaussian, Inc.

[8] orgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., & Klein, M. L. (1983). Comparison of simple potential functions for simulating liquid water. The Journal of Chemical Physics, *79*(2), 926–935. https://doi.org/10.1063/1.445869.

Table S1 The contribution of HOMO→LUMO transitions to the S_1 state for **Systems I** and **II-V**

System	Metal ion	HOMO-LUMO contribution
I	Free	92.3%
II	Ag^{+}	93.6%
III	Zn^{2+}	92.2%
IV	Cd^{2+}	94.0%
V	Hg^{2+}	93.0%

Table S2 Mulliken charges of **oLu** localized on the benzothiazole and thiazole moieties, together with the net charge-transfer amounts during the $S_1 \rightarrow S_0$ de-excitation in **Systems I** and **II-V**

System	State	Benzothiazole (e)	Thiazole (e)	Net transfer ($S_1 \rightarrow S_0$, e)
I	S_0	-0.407	-0.593	0.085
	S_1	-0.323	-0.677	
II	S_0	0.288	-1.288	0.123
	S_1	0.411	-1.411	
III	S_0	-0.134	-0.866	0.117
	S_1	-0.017	-0.983	
IV	S_0	0.028	-1.028	0.102
	S_1	0.130	-1.130	
V	S_0	-0.095	-0.905	0.131
	S_1	0.036	-1.036	

Table S3 Key bond lengths (Å) and dihedral angles (°) of oLu (see Fig 1) in 5 systems at TD CAM-B3LYP/6-31+G//MM level, D1 represents the dihedral angle C7′–S1′–N3′–S1**

System	R1	R2	R3	R4	R5	R6	R7	D1
I	1.746	1.333	1.415	1.332	1.766	1.269	1.245	-177
II	1.742	1.334	1.421	1.336	1.756	1.269	1.258	172
III	1.747	1.328	1.416	1.336	1.758	1.271	1.263	171
IV	1.746	1.332	1.416	1.334	1.752	1.262	1.258	170
V	1.743	1.333	1.422	1.334	1.757	1.274	1.262	171

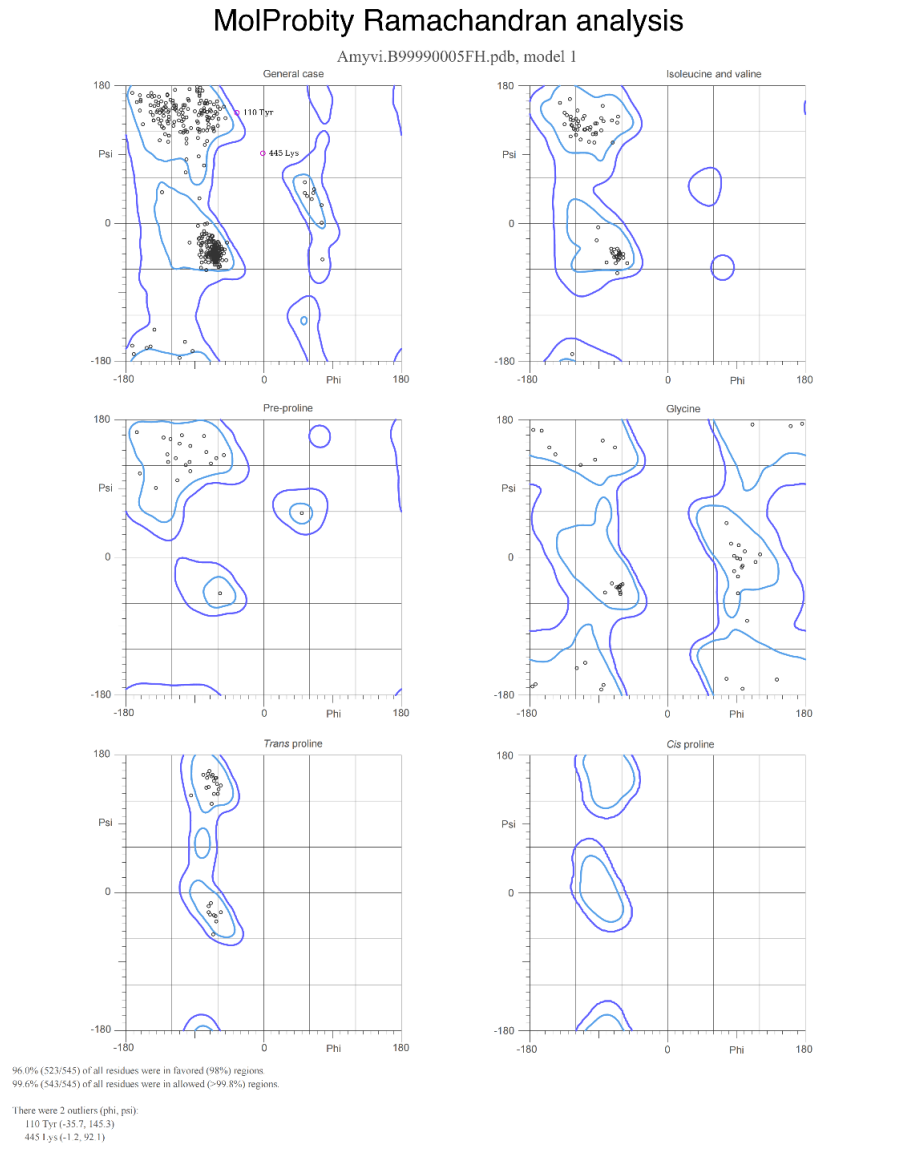

Fig S1 Ramachandran analysis of the most stable conformation of *A. vivianii* luciferase from homology modeling

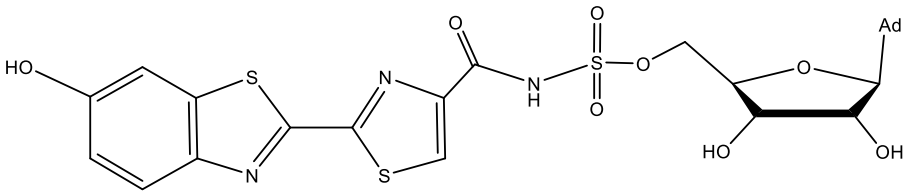

Fig S2 The adenylate analogue 5′-O-[N-dehydroluciferyl)-sulfamoyl]-adenosine (DLSA)

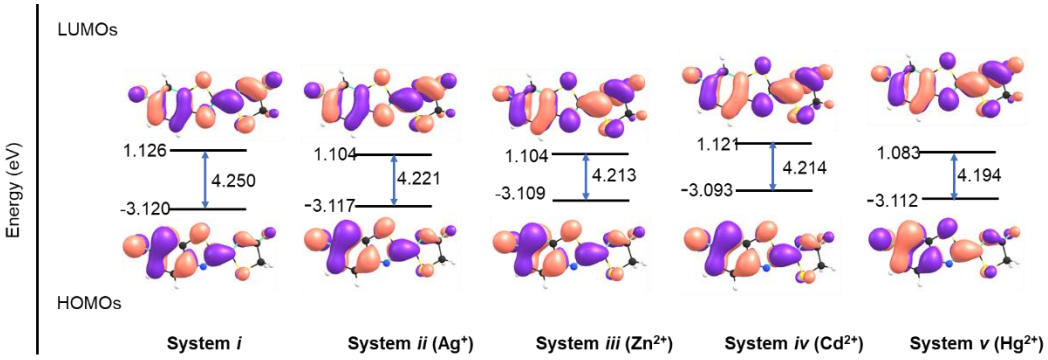

Fig S3 Frontier molecular orbital analysis of the isolated **oLu** in **System *i*** and ***ii-v***

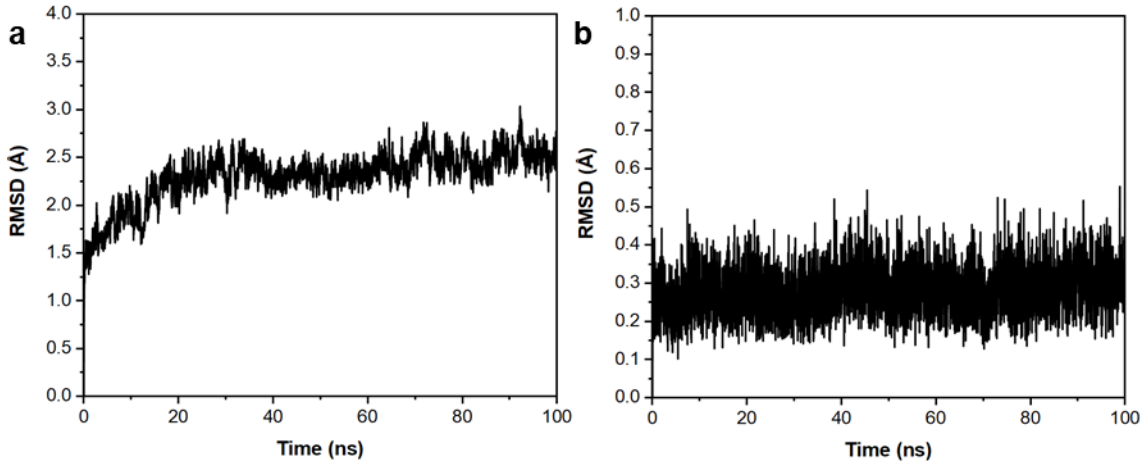

Fig S4 Root mean square deviation (RMSD) of (a) the protein backbone and (b) **oLu** of the *A. vivianii* firefly luciferase complex (**System I**) over the 100 ns molecular dynamics simulation

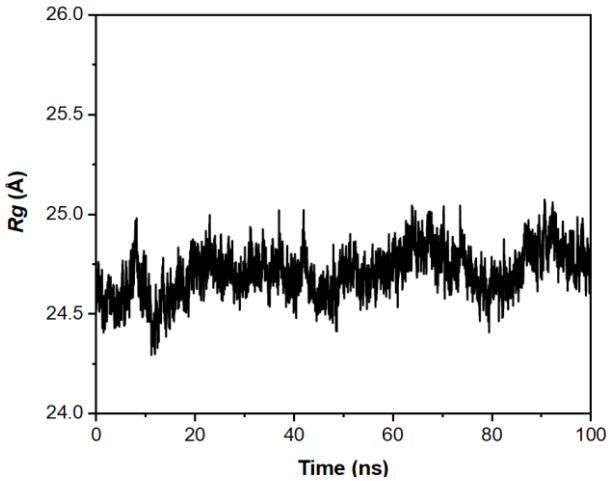

Fig S5 Radius of gyration (Rg) of the protein of the A. vivianii firefly luciferase complex (**System I**) over the 100 ns MD simulation

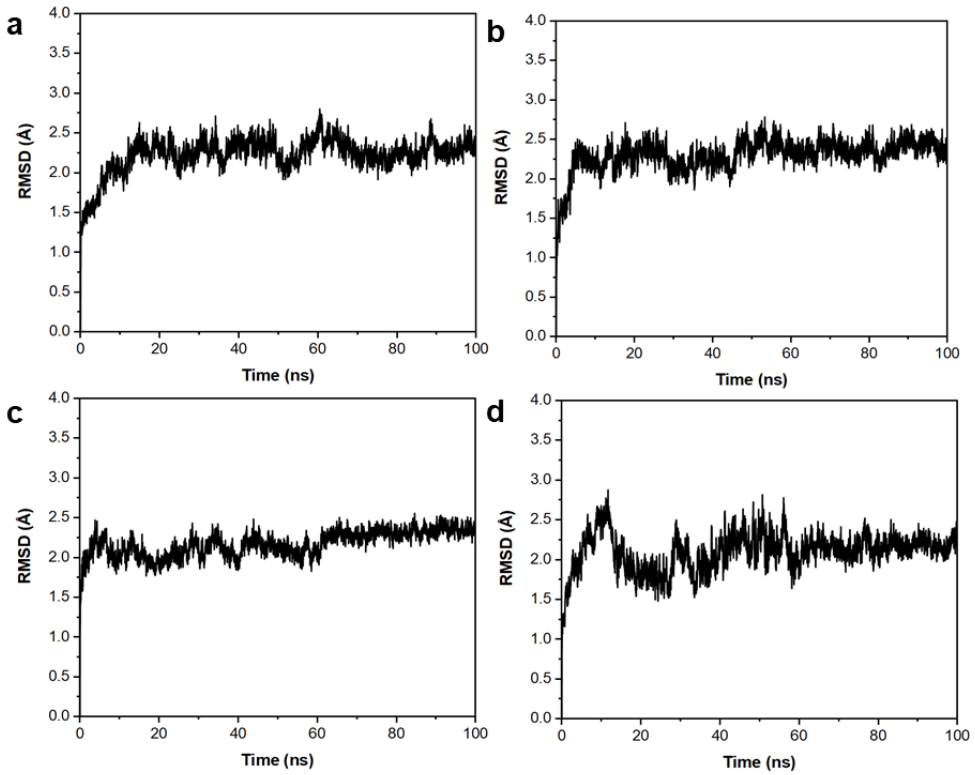

Fig S6 Root mean square deviation (RMSD) of the protein backbone of the *A. vivianii* firefly luciferase complex binding (a) Ag^{+} (**System II**), (b) Zn^{2+} (**System III),** (c) Cd^{2+} (**System IV**) and (d) Hg^{2+} (**System V)** over the 100 ns MD simulation

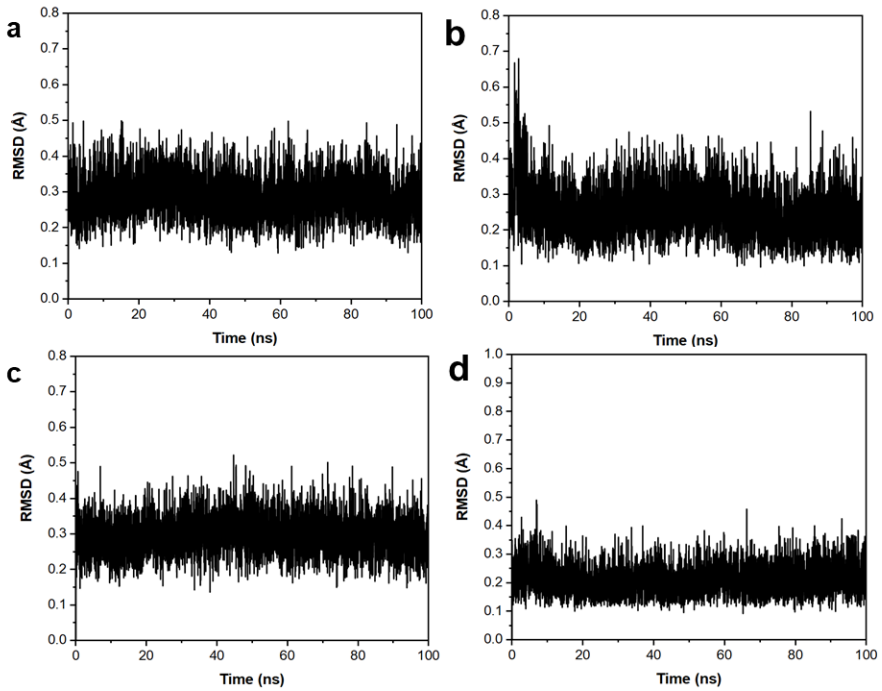

Fig S7. Root mean square deviation (RMSD) of **oLu** of the *A. vivianii* firefly luciferase complex binding (a) Ag^{+} (**System II**), (b) Zn^{2+} (**System III),** (c) Cd^{2+} (**System IV**) and (d) Hg^{2+} (**System V)** over the 100 ns MD simulation

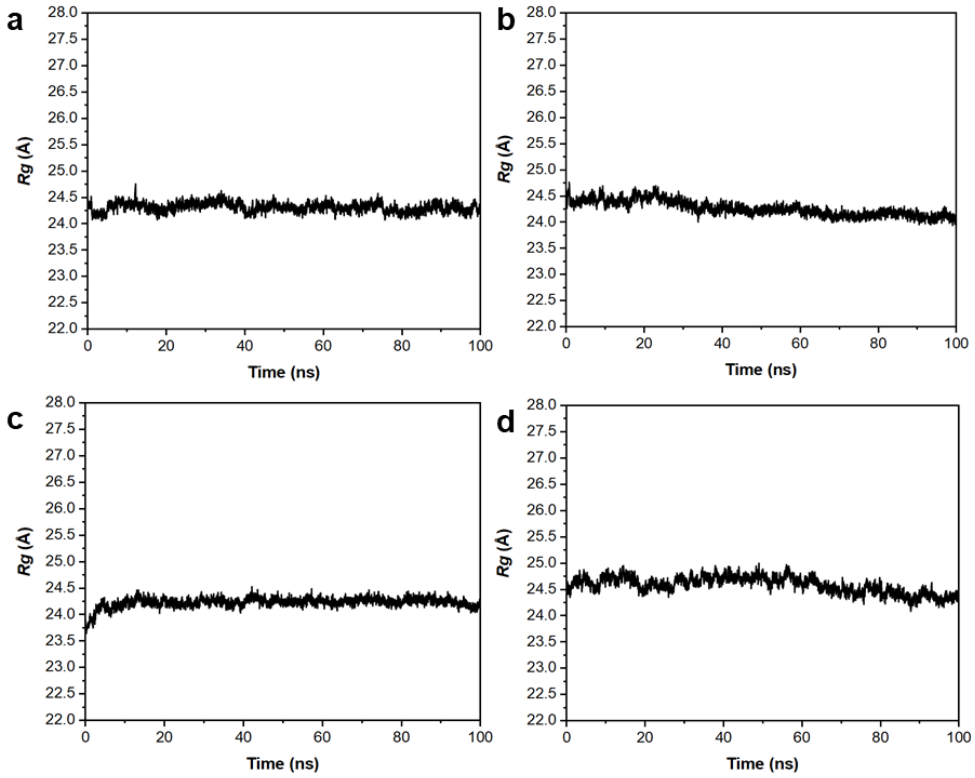

Fig S8 Radius of gyration (Rg) of **oLu** of the *A. vivianii* firefly luciferase complex binding (a) Ag^{+} (**System II**), (b) Zn^{2+} (**System III),** (c) Cd^{2+} (**System IV**) and (d) Hg^{2+} (**System V)** over the 100 ns MD simulation

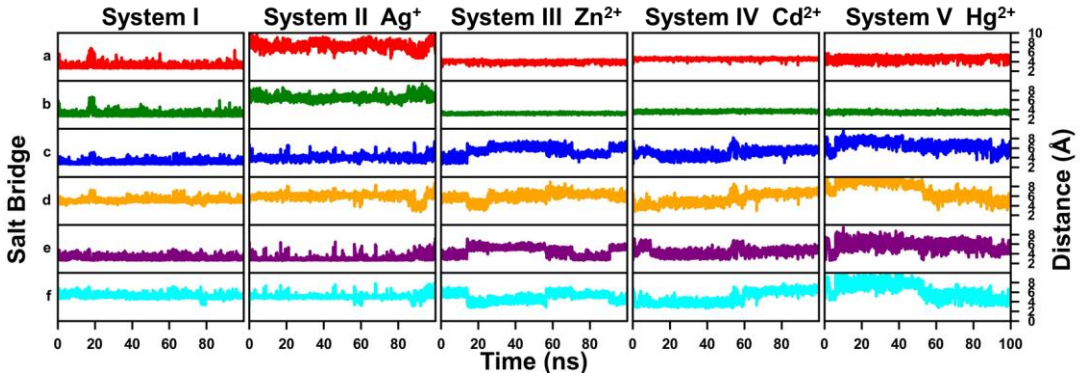

Fig S9 The distances of salt bridge over time for the five systems. Panels a–f correspond to the distances between: (a) His310NE2–Glu354OE1, (b) His310NE2–Glu354OE2, (c) Glu311OE1–Arg337NH1, (d) Glu311OE1–Arg337NH2, (e) Glu311OE2–Arg337NH1, (f) Glu311OE2-Arg337NH2

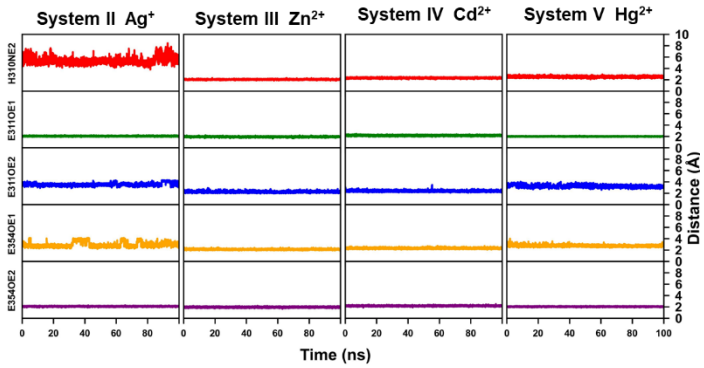

Fig S10 Distances between the metal ion and potentially coordinating atoms of salt-bridge residues (H310NE2, E311OE1, E311OE2, E354OE1, E354OE2) in four metal ion-bound systems during the 100 ns MD simulations

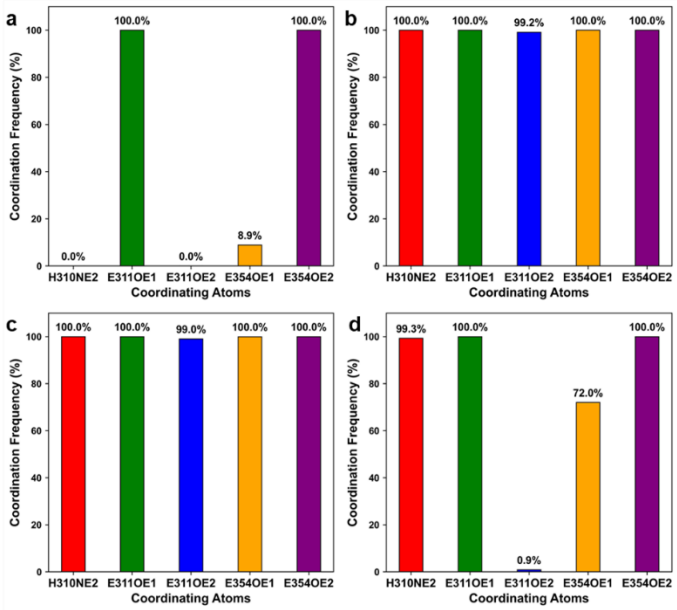

Fig S11 Coordination frequencies of salt-bridge residue atoms with (a) Ag^{+}, (b) Zn^{2+}, (c) Cd^{2+} and (d) Hg^{2+} from equilibrated MD trajectories

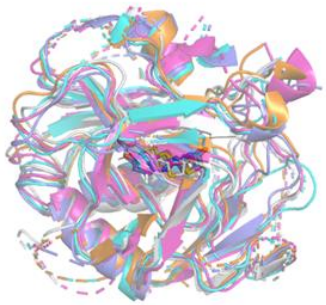

Fig S12 Structural overlay and conformational comparison of **Systems I** (silver), **II** (light blue), **III** (pink), **IV** (dark blue), and **V** (orange)

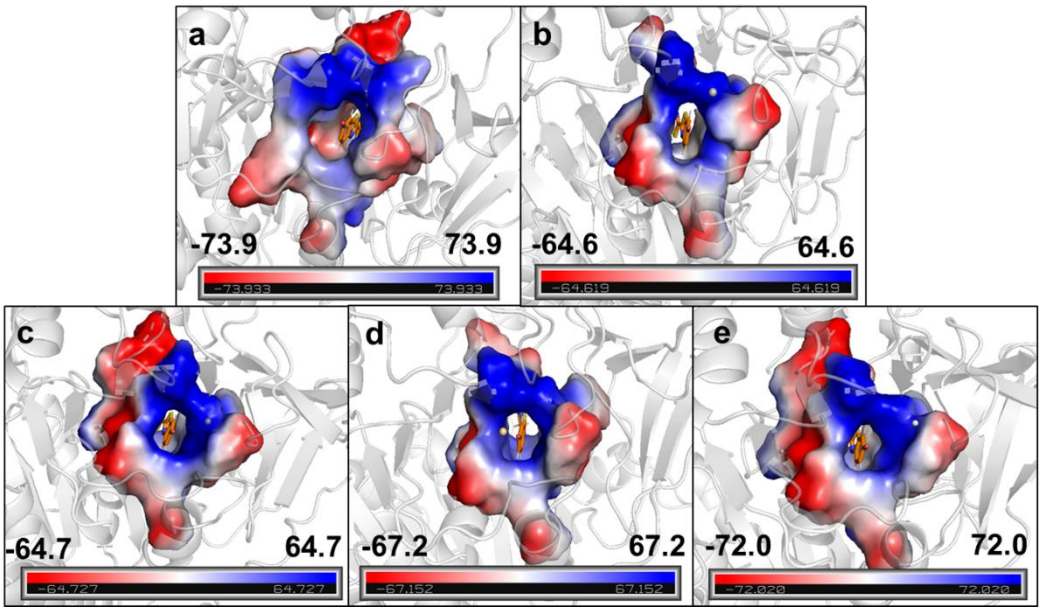

Fig S13 The electrostatic potential of the active-site pocket of **Systems** (a) **I**, (b) **II**, (c) **III**, (d) **IV**, and (e) **V**

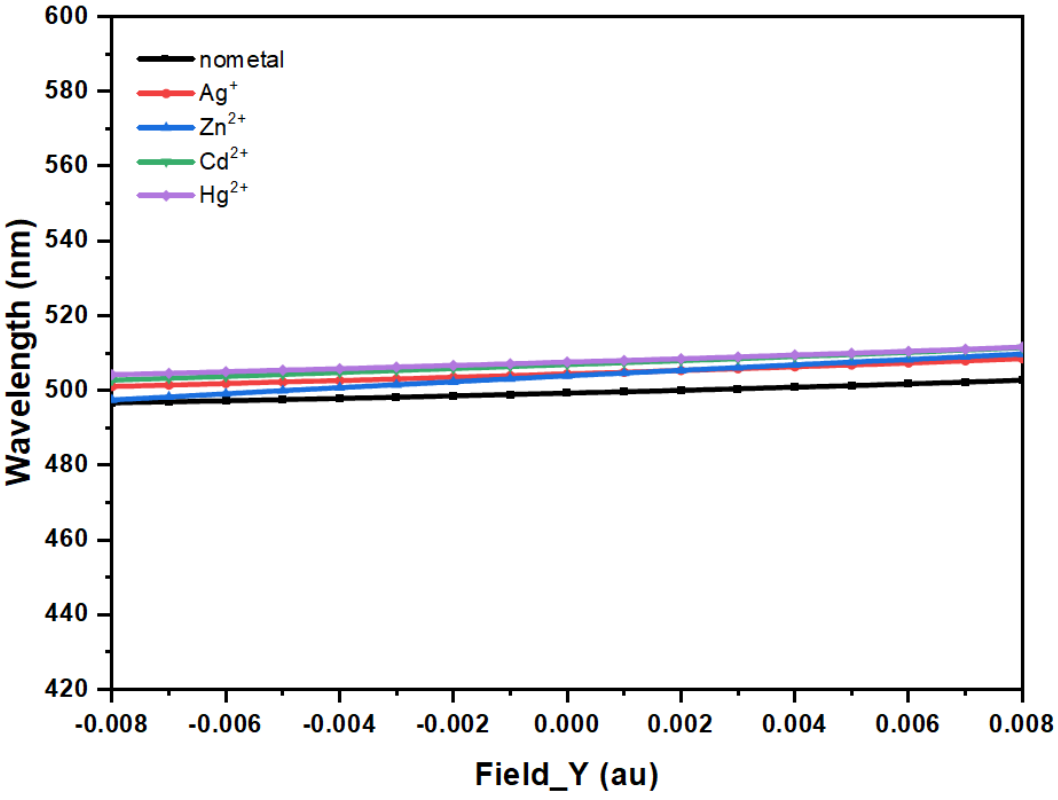

Fig S14 Relationship between the λ_F of **oLu** and the strength of EEF applied along the Y direction

The Cartesian coordinates (Å) of stationary points (only substrate) computed at (TD CAM-B3LYP/6-31+G: Amber) level.**

System I-nometal

C	1.740	1.648	-29.008
N	1.198	0.419	-28.961
O	1.148	2.703	-29.304
S	3.768	0.008	-28.456
C	3.228	1.715	-28.680
N	2.812	-2.834	-28.563
O	0.264	-7.693	-28.541
S	0.219	-2.538	-28.668
C	2.101	-0.523	-28.694
C	1.849	-1.915	-28.630
C	0.870	-4.147	-28.602
C	0.168	-5.344	-28.593
C	0.902	-6.597	-28.562
C	2.280	-4.095	-28.578
C	2.999	-5.319	-28.589
C	2.336	-6.522	-28.583
H	-0.913	-5.389	-28.634
H	2.875	-7.462	-28.601
H	4.084	-5.279	-28.605
H	3.352	2.286	-27.762
H	3.764	2.219	-29.487

System II-Ag^+

C	44.633	50.707	49.381
N	45.095	50.078	48.288
O	44.450	51.946	49.498
S	44.941	48.163	50.117
C	44.393	49.816	50.588
N	46.243	46.625	47.903
O	49.397	44.881	43.774
S	46.320	48.376	45.977
C	45.308	48.780	48.514
C	45.894	47.874	47.589
C	47.115	46.850	45.734
C	47.881	46.457	44.645
C	48.611	45.203	44.715
C	46.962	46.050	46.883
C	47.622	44.797	46.926
C	48.421	44.390	45.885
H	47.990	47.062	43.752
H	48.961	43.451	45.934

H	47.528	44.196	47.824
H	43.339	49.822	50.858
H	44.970	50.181	51.436

System III-Zn^{2+}

C	44.529	49.482	50.564
N	44.780	49.350	49.261
O	44.376	50.592	51.145
S	45.152	46.914	50.233
C	44.497	48.181	51.355
N	46.319	46.604	47.448
O	48.864	47.300	42.649
S	45.660	48.880	46.352
C	45.145	48.106	48.940
C	45.674	47.742	47.678
C	46.689	47.803	45.465
C	47.311	48.035	44.246
C	48.261	47.065	43.742
C	46.924	46.638	46.222
C	47.831	45.678	45.713
C	48.480	45.879	44.520
H	47.148	48.930	43.660
H	49.195	45.154	44.147
H	48.039	44.807	46.322
H	43.460	47.965	51.623
H	45.086	48.242	52.271

System IV-Cd^{2+}

C	46.449	49.524	46.850
N	47.337	48.655	46.359
O	46.429	50.747	46.556
S	46.032	47.282	48.205
C	45.485	48.974	47.885
N	48.355	45.388	47.593
O	53.365	43.383	46.638
S	49.547	46.689	45.666
C	47.277	47.479	46.988
C	48.255	46.468	46.819
C	50.370	45.337	46.387
C	51.653	44.894	46.122
C	52.207	43.807	46.906
C	49.566	44.772	47.399
C	50.111	43.729	48.185
C	51.385	43.267	47.959

H	52.281	45.325	45.353
H	51.813	42.483	48.575
H	49.517	43.351	49.009
H	44.462	48.981	47.505
H	45.519	49.578	48.796

System V-Hg^{2+}

C	44.936	47.526	48.793
N	45.844	47.635	47.817
O	44.033	48.373	49.033
S	46.728	45.646	49.322
C	45.097	46.322	49.704
N	49.063	45.951	47.441
O	53.116	47.797	44.224
S	48.208	47.810	45.818
C	46.833	46.751	47.960
C	48.010	46.719	47.163
C	49.861	47.290	45.687
C	50.850	47.803	44.853
C	52.210	47.319	44.980
C	50.129	46.292	46.642
C	51.456	45.813	46.760
C	52.464	46.309	45.966
H	50.657	48.573	44.116
H	53.482	45.950	46.070
H	51.670	45.080	47.530
H	44.305	45.604	49.474
H	45.018	46.602	50.756