Supporting Information

Metal Coupled Folding of Cys_2His_2 Zinc-Finger

Wenfei Li, Jian Zhang, Jun Wang, and Wei Wang*

National Laboratory of Solid State Microstructure, and Department of Physics,

Nanjing University, Nanjing 210093, China

Email: wangwei@nju.edu.cn

Simulation details

Figure S1-Figure S14

Table S1-Table S3

The complete list of authors for Reference 33 and 45

Simulation Details.

In the simulations, the nuclear magnetic resonance (NMR) structure of the Sp1f2 (PDB code: 1sp2)[1] is solvated in a TIP3P water box. The water density is adjusted to around 1.0 g/cm^3. Nine Cl^- and four Na^+ are added to neutralize the system and mimic the buffer. The Zn(II) is allowed to move freely around the simulation box, namely, no extra constraint is applied to the Zn(II) during the production run. In treating the long-range electrostatic interactions, the Particle Mesh Ewald (PME) summation algorithm is employed. The covalent bonds involving hydrogen atoms are constrained with the SHAKE algorithm and the time step of 0.002 ps is used. The replica-exchange MD (REMD) method is used for conformational sampling[2,3]. 64 replicas are simulated in NVT ensembles, with the temperatures ranging from 289 to 607K. The time intervals between the exchange attempts are 0.8 ps and the atomic coordinates are recorded every 0.2 ps for further analysis. The initial structures of REMD simulation are prepared by high temperature simulation of 1000 K for 0.5 ns. Totally 50 ns are simulated for each replica. The structures of the last 40 ns for each replica are used for analysis. In constructing the free energy landscape at certain temperature, the weighted histogram analysis method is used[4]. Note that at the beginning of the 40ns production run, around 40 percent of the structures are native like, and the Zn(II) is still bound to the peptide. Therefore, the present simulation should be regarded as a conformational sampling among the conformational space rather than a real folding simulation. For comparison, we also conducted a control simulation with the Zn(II) being removed and the cysteines being protonated. The cluster analysis performed for the structures sampled by the peptides with and without zinc binding is based on the rms distance, and the cutoff of the *rmsd* in assigning the clusters is 3.5 Å. The principal component analysis is

conducted in Cartesian-coordinate space. A hydrogen bond is defined as formed when the distance between the donor and acceptor is less than 3.5 Å, and the angle N–H···O is larger than 120.0°.

The reaction coordinates Q, Q_β, N_α, R_g, R_g^{core}, *rmsd* and N_{nl} are used. Q is the fractional native contact. Q_β is the fractional native contacts for the N-terminal β-hairpin region (residues Arg1-Arg16). N_α is defined as the number of helical residues formed among residues Ser17-Lys30. A residue is defined as helical residue whenever three or more consecutive residues satisfy the dihedral constrains of $-95^o<\varphi<-25^o$ and $-77^o<\psi<-17^o$. R_g and R_g^{core} are radius of gyration of the peptide and the hydrophobic core, respectively. *rmsd* is the rms distance for all atoms. N_{nl} represents the number of the native ligands coordinated, namely, Cys5, Cys10, His23 and His27.

(1) Narayan, V. A.; Kriwacki, R. W.; Caradonna, J. P. *J. Biol. Chem.* **1997**, 272, 7801–7809.

(2) Sugita, Y.; Okamoto, Y. *Chem. Phys. Lett.* **1999**, 314, 141–151.

(3) Zhou, R. *J. Mol. Graph Model* **2004**, 22, 451–463.

(4) Ferrenberg, A. M; Swendsen, R. H. *Phys. Rev. Lett.* **1989**, 63, 1195–1198.

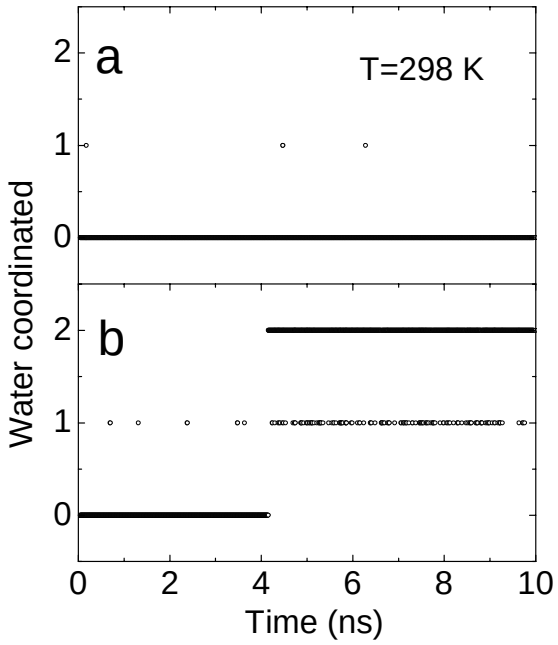

Figure S1. Number of water molecules coordinated to Zn(II) at *T*=298 K with (a) and without (b) the modifications implemented in this work as a function of time. One can see that without modifications to the model, up to two water molecules can come into the ligand shell of Zn(II), which results in a hexacoordination structure. This structure is not consistent with the tetrahedral coordination structure determined experimentally. After modifying the model, the water molecules have almost no chance to come into the ligand shell of Zn(II), which is consistent with the experimental data.

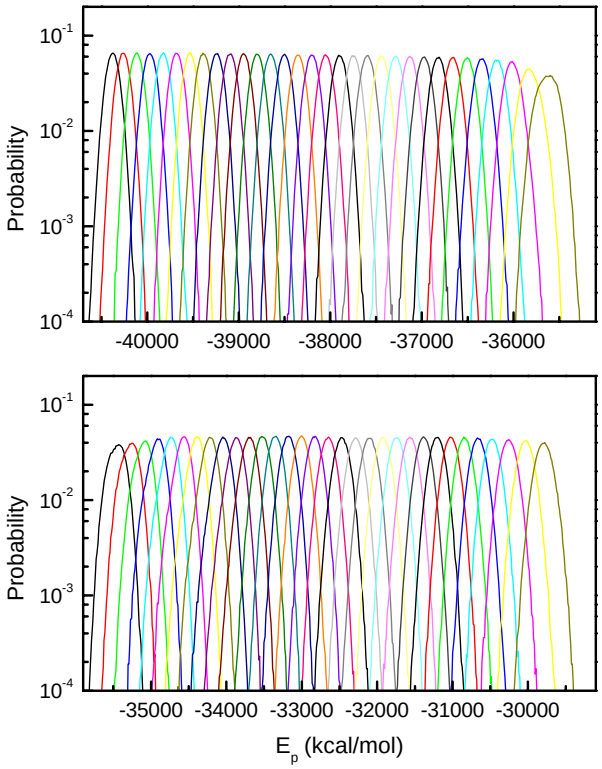

Figure S2. Potential energy distributions at each temperature of the REMD simulation. The large overlap between the potential energy distributions of the neighboring temperatures ensures the high exchange rate (20% – 30%).

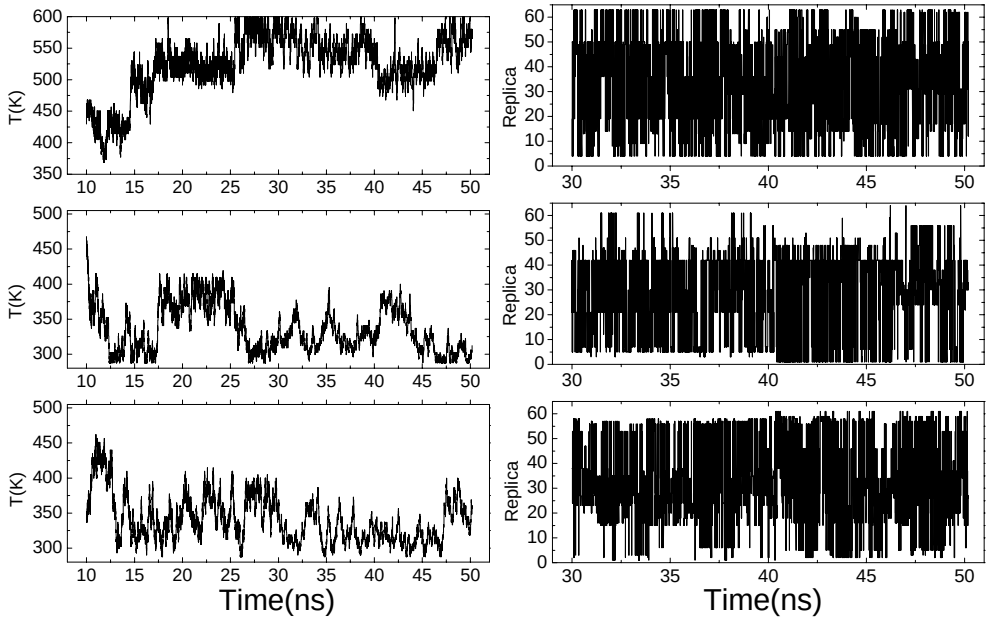

Figure S3. Temperatures as a function of time for three selected replicas (left) and the replica number as a function of time at temperatures of 0.8 T_m (right top), 1.0 T_m (right middle) and 1.2 T_m (right bottom). One can see that during the simulation, the replicas can visit a wide range of temperatures, and each temperature can be assigned into different replicas, elucidating the high sampling performance. For clarity, only the data of the last 20 ns are shown in the right panels.

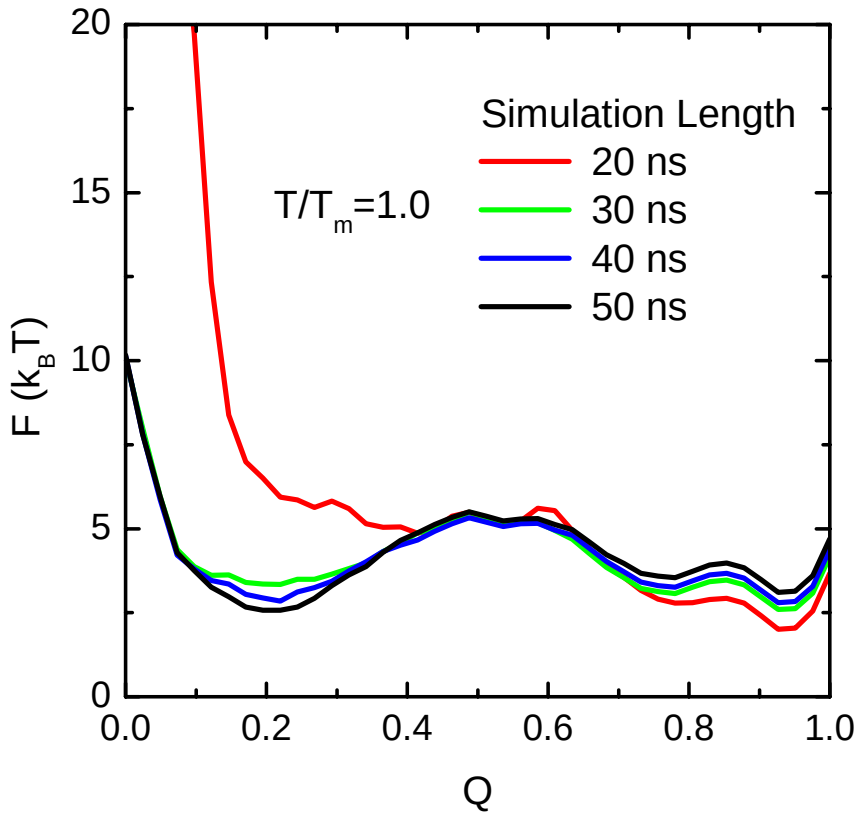

Figure S4. Free energy landscape projected onto reaction coordinate Q at melting temperature T_m with the simulation length of 20 ns, 30 ns, 40 ns and 50 ns, respectively. Good convergence is obtained for the free energy landscape when the simulation length reaches 30 ns. Note that the data of the beginning 10 ns are dropped.

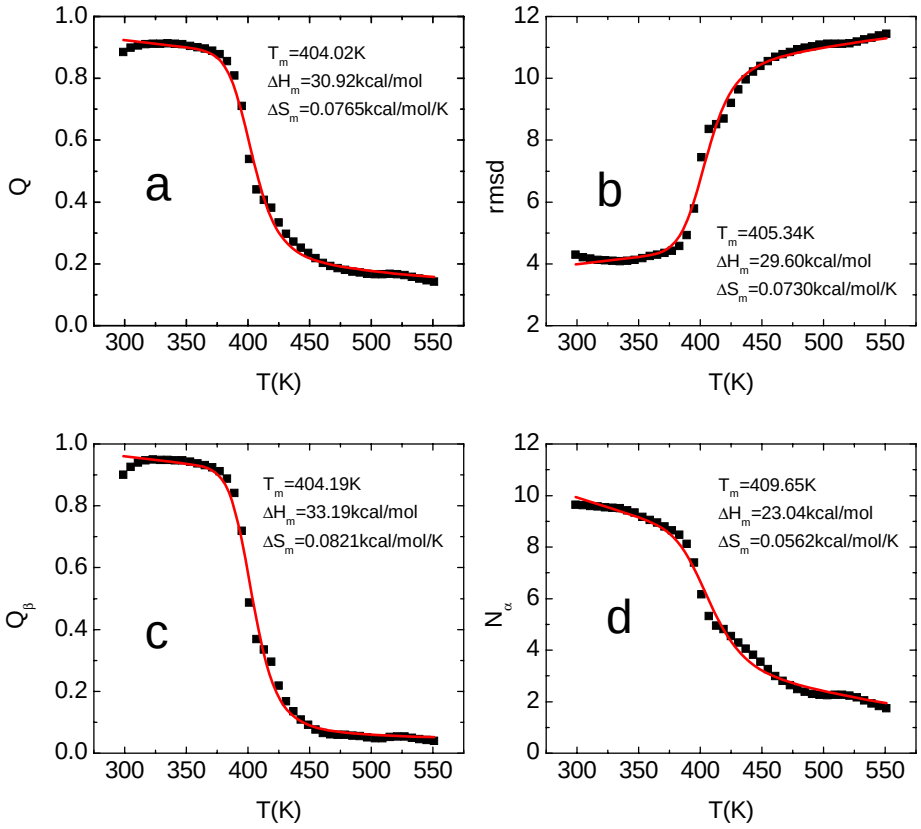

Figure S5. Average Q (a), *rmsd* (b), Q_β (c) and N_α (d) as a function of temperature T (solid squares) and the two-state model fitting (red solid line). The two-state formula $P_f(T) = 1/(1+\exp(\Delta G/RT))$ is used to fit the denaturation curves, where R is the gas constant. P_f represents the probability of folded and is determined by $P_f(T) = (RC(T) - RC_D(T))/(RC_N(T) - RC_D(T))$, where RC represents the reaction coordinate with the RC_N and RC_D being the base lines around the denatured state and native state, respectively. The free energy change ΔG is determined by $\Delta G = \Delta H_m(1 - T/T_m) + \Delta C(T - T_m - T\ln(T/T_m))$ with ΔH_m, T_m and ΔC being the enthalpy change, melting temperature, and heat capacity change, respectively, around transition region[1]. The entropy change ΔS_m is determined by $\Delta G = \Delta H_m - T_m \Delta S_m$ with the $\Delta G = 0$ at T_m. The fitted melting temperature T_m, enthalpy change ΔH_m and entropy change ΔS_m at T_m are shown in the panels. The transition temperature obtained by the fitting varies between 404.02–409.65 K. Up to 400 K, the peptide maintains the native structure very well, which implies the high thermal stability of the zinc-finger. In this

paper, when we perform the analysis at melting temperature, 405 K is used. Here, the transition temperature may be overestimated due to the use of a constant volume in the molecular dynamics simulations and the inaccuracy of the force field in high temperatures[2,3]. The enthalpy change ΔH_m and entropy change ΔS_m vary between 23.04–33.19 kcal/mol and between 0.0562–0.0821 kcal/mol/K, respectively, which are close to the experimental values[4].

(1) Murphy, K. P.; in *Protein structure, Stability and Folding;* Murphy, K. P. Eds; Humana Press Inc., Totowa, N.J., **2001**; pp 1-16.
(2) García, A. E.; Onuchic, J. N. *Proc. Natl. Acad. Sci. USA* **2003**, 100, 3898–13903.
(3) Zhou R. *Proc. Natl. Acad. Sci. USA* **2003**, 100, 13280–13285.
(4) Blasie, C. A.; Berg, J. M. *Biochemistry* **2002**, 41, 15068–15073.

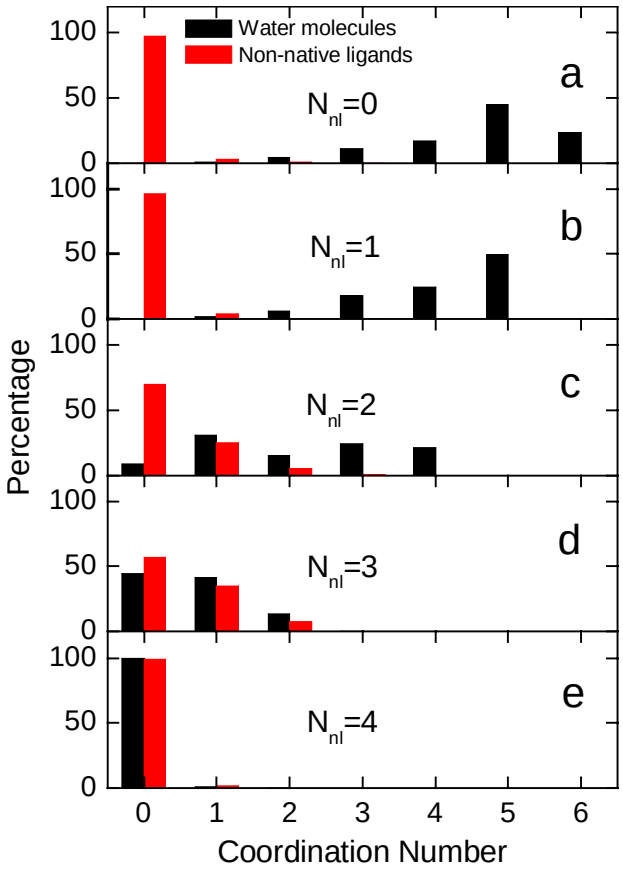

Figure S6. Distribution of the coordinated number of water molecules (black bar) and non-native ligands (red bar) for the conformations with (a) zero, (b) one, (c) two, (d) three and four (e) native ligands coordinated to Zn(II). The non-native ligands refer to all of the potential ligands which can coordinate to Zn(II) except for the four native ligands and water molecules. One can see that when all of the four native ligands coordinate to the Zn(II), no water molecules and non-native ligand atoms come into the ligand shell (e), which is consistent with the experimental data of the native coordination geometry. When the coordination bond Zn(II)-His27 is broken up, the position of the His27 is replaced by the water molecules or non-native ligands. In this case, the tetrahedral geometry may not be reserved since more than four ligands can come into the ligand shell (d). In fact, in Ref.[1], the authors observed a signal of pentacoordination geometry for the same zinc-finger with the His27 being mutated to Ala and the Zn(II) being replaced by Co(II). This figure shows that not only water molecules can occupy the coordination position, other non-native ligand atoms may also

come into the ligand shell with the absence of the His27 coordination. With the further breaking up of the native ligands, more water molecules come into the ligand shell. In particular, when all of the four native ligands are broken up, up to six water molecules can come into the ligand shell, which is consistent with the experimentally detected coordination number of the Zn(II) solvated in the aqueous solvent. One can also observe that the non-native ligands have the highest probability to coordinate with Zn(II) for the conformations with two or three native ligands coordinated.

(1) Nomura, A.; Sugiura, Y. *Inorg. chem.* **2002**, 41, 3693–3698.

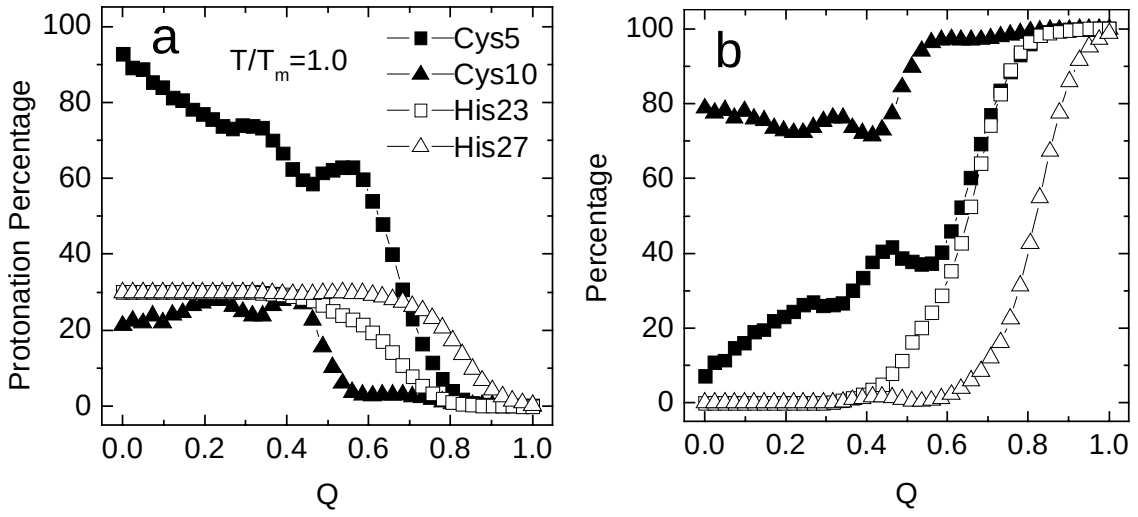

Figure S7. Percentages of ligand protonation (a) and coordination bond formation (b) for the four native ligands as a function of reaction coordinate *Q*. The protonation percentages are calculated according to the distances between the Zn(II) and native ligands. The ligands are treated as deprotonated when the coordination bonds are formed. After the breaking up of the coordination bonds, the Cys5 and Cys10 are treated as fully protonated, and the His23 and His27 are treated as 30% protonated according to their pKa value. By counting the structure number with the coordination bond formed at certain reaction coordinate *Q*, we can determine the averaged protonation state of each native ligand along the folding pathway of the zinc-finger peptide.

One can see from Figure S7a, when the *Q* is small, the Cys5 is mostly protonated. In comparison, the Cys10 is protonated by only around 20%. This is because that at unfolded state, the Zn(II)-Cys5 bond is mostly broken up. Whereas the Zn(II)-Cys10 bond is still formed to large extent as shown in Figure S7b. At unfolded state, both the His23 and His27 are protonated by 30% because the Zn(II)-His23 and Zn(II)-His27 are all broken up. With the increasing of the reaction coordinate *Q*, all the ligands are gradually deprotonated due to the coordination with the Zn(II).

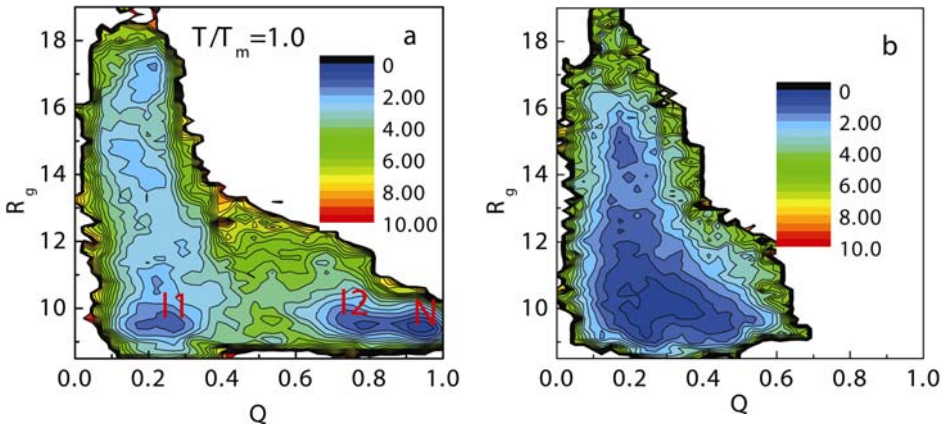

Figure S8 Free energy landscapes projected onto reaction coordinates (R_g, Q) for the peptides with (a) and without (b) zinc binding at T_{m}. The unit of the free energy is $k_B T_m$.

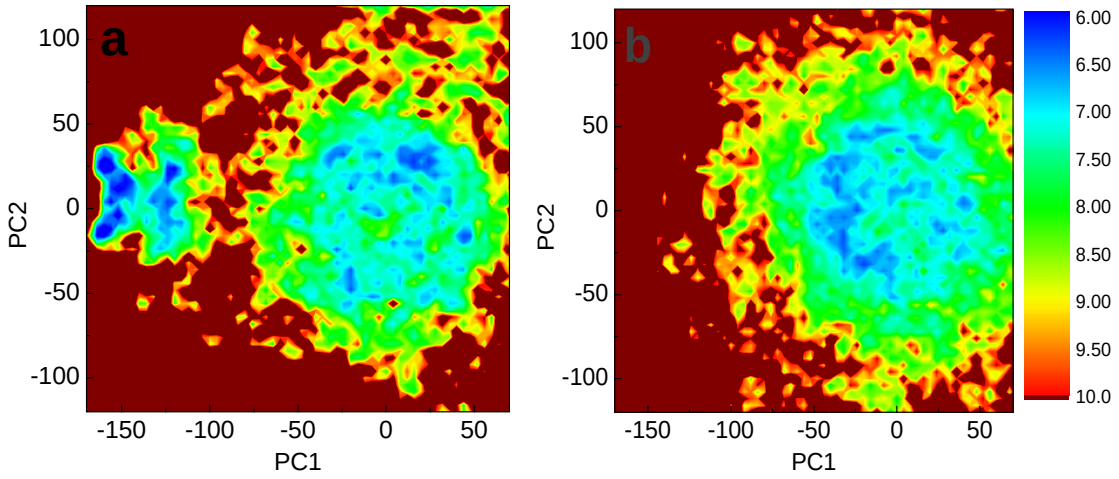

Figure S9 Free energy of the holo-peptide (a) and apo-peptide (b) projected onto the first (*PC1*) and second (*PC2*) principal components of the holo-peptide at T_m. In this figure, the free energies are represented by $-k_B T_m lnP(PC1,PC2)$ with $P(PC1,PC2)$ being the distribution probability calculated by the structures sampled at T_m. The unit of the free energy is $k_B T_m$.

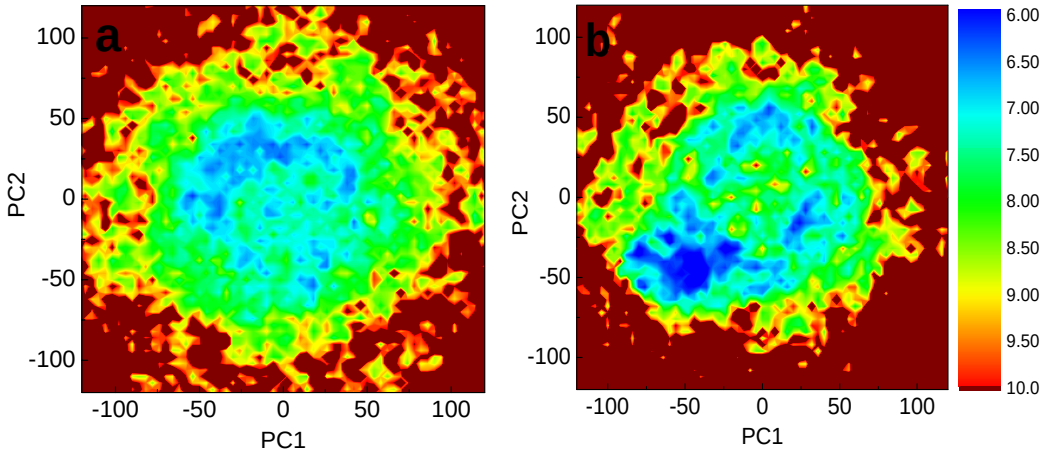

Figure S10 Free energy of the apo-peptide (a) and holo-peptide (b) projected onto the first (*PC1*) and second (*PC2*) principal components of the apo-peptide at T_m.

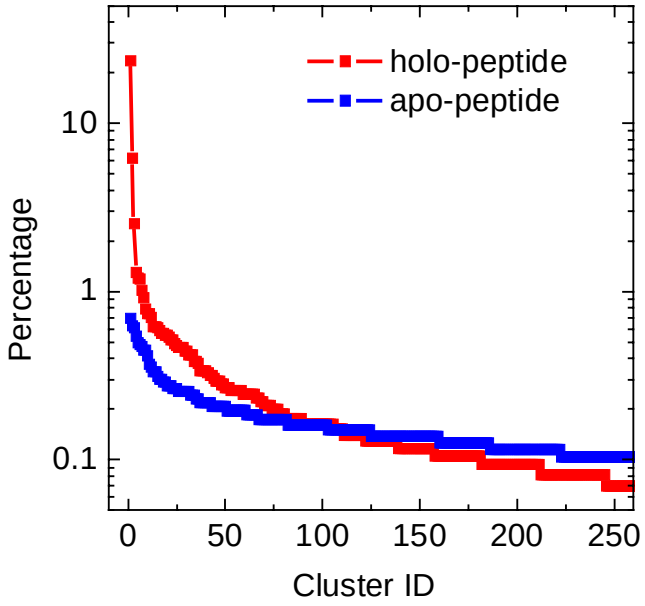

Figure S11 Percentages of the structures in each cluster for the holo-peptide (red) and apo-peptide (blue) sampled at T_m. The conformational clustering is based on the *rmsd*. The cut off of the *rmsd* in assigning the clusters is 3.5 Å.

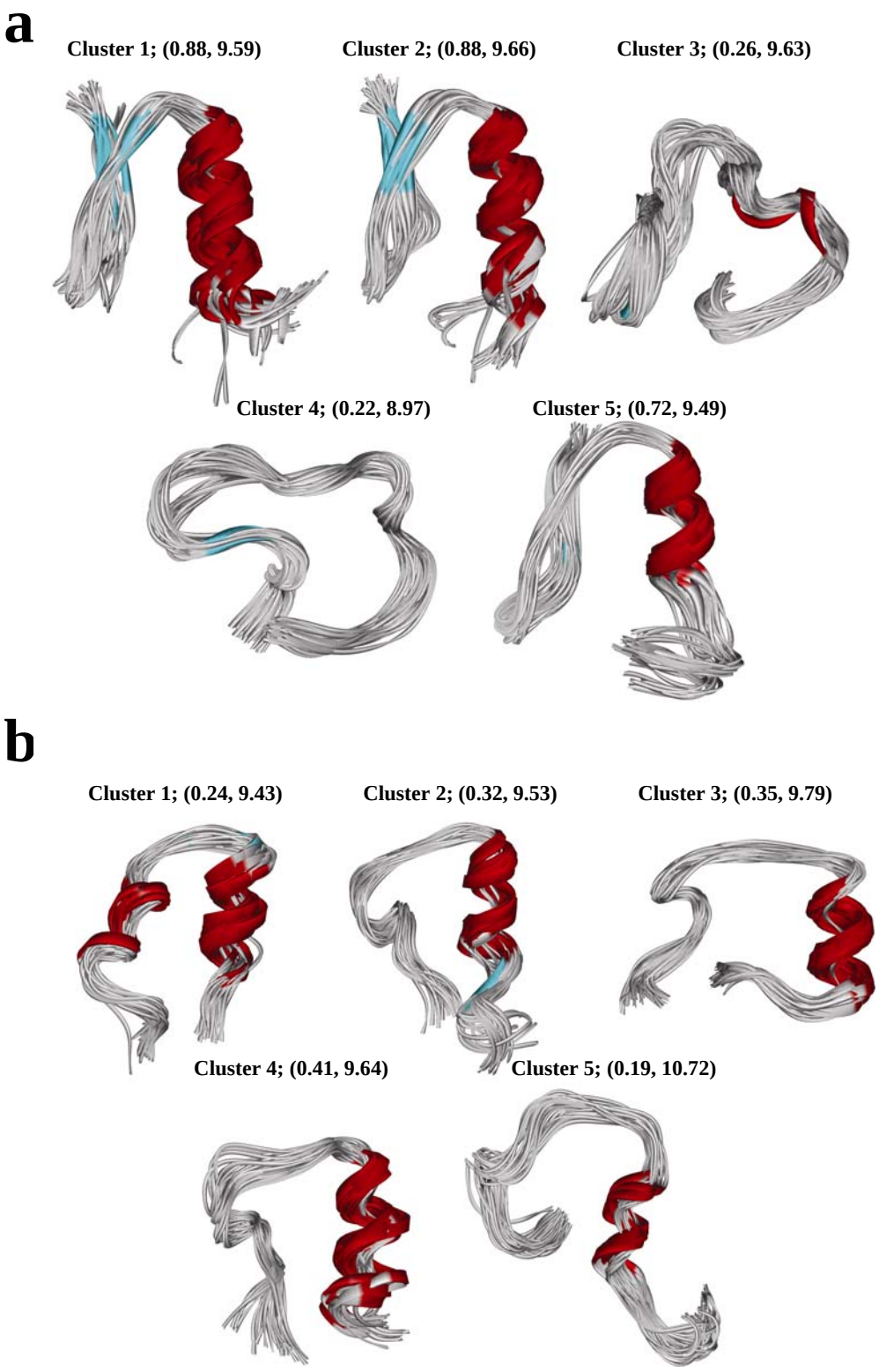

Figure S12 Structures of the five most probable clusters in **Figure S11** for the holo-peptide (a) and apo-peptide (b). Only 30 structures for each cluster are plotted. The cluster ID and the corresponding reaction coordinates (Q, R_g) are also presented.

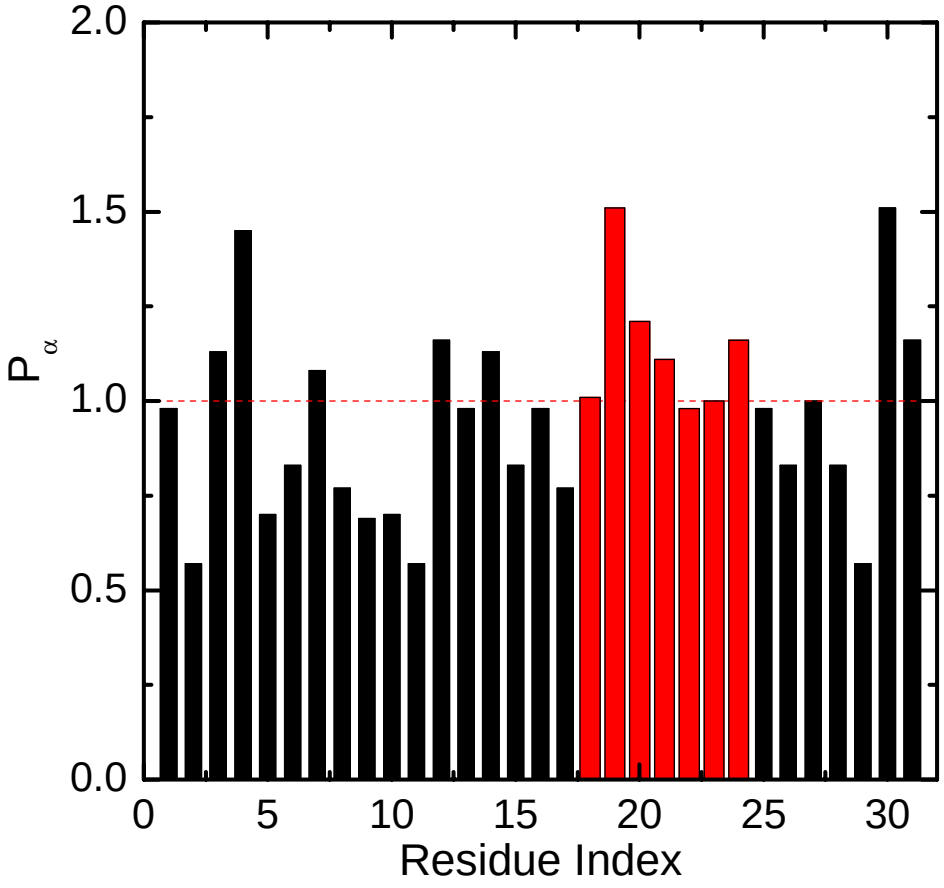

Figure S13 α-helix propensity of each amino acid of the zinc-finger Sp1f2 according to the statistical results of Chou and Fasman[1]. The segment Gln18-Thr24 (red bars) which locates at the first two helical turns of the C-terminal α-helix has higher α-helix propensity compared to other amino acid segments. This high α-helix propensity of segment Gln18-Thr24 results in the high probability for sampling the α-helix conformation even without zinc binding to the Histidines.

(1) Chou, P. Y.; Fasman, G. D. *Adv. Enzymol.* **1978**, **47**, 45–148.

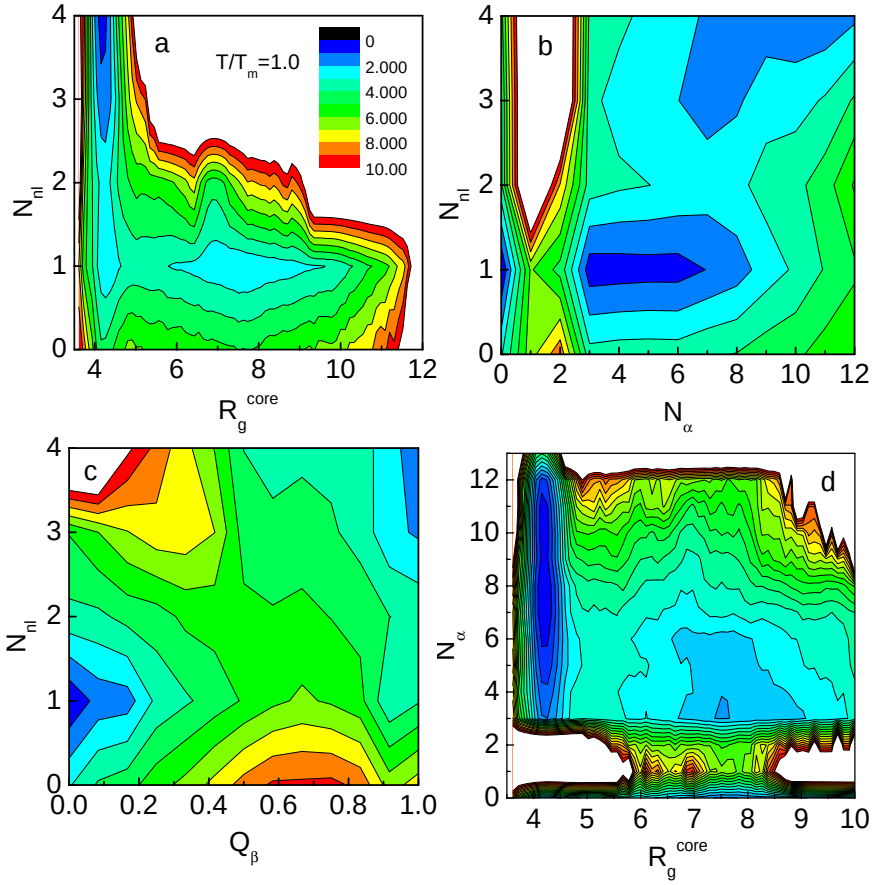

Figure S14 Free energy landscape projected onto the reaction coordinates (N_{nl}, R_g^{core}) (a), (N_{nl}, N_α) (b), (N_{nl}, Q_β) (c) and (N_α, R_g^{core}) (d) at T_m. The unit of the free energy is k_BT_m. Note that the barriers between 0 and 3 along the N_α axis in the left bottom of (b) and right bottom of (d) are resulted from the definition of the helical residues, which demands three or more consecutive residues satisfying the dihedral constraints defined in **simulation details**. One can see that the zinc binding plays crucial role on the formation of the hydrophobic core, as well as the folding and stabilization of the component secondary structures. **Figure S14d** also indicates that the packing of the hydrophobic residues can contribute to the folding and stabilization of the C-terminal α-helix.

Table S1. The parameters A_i and B_i used to calculate the transferred charges of the liganding atoms, as well as the model systems used to derive them by quantum chemical calculations.

Atom	Model system	A_i (e/Å)	B_i (e)
S_γ (Cys)	$CH3S^-$	-0.432	1.244
N_ε (His)	Imidazole	-0.272	0.734
O (backbone, Gln)	HCONH2	-0.485	1.232
N (backbone)	HCONH2	-0.232	0.627
O (Glu, Asp)	$HCOO^-$	-0.224	0.570
O (Ser, Thr, Tyr)	CH3OH	-0.123	0.320
O (Water)	H_2O	-0.100	0.266

Notes: The parameters A_i and B_i are calculated by combining the quantum chemical data and the AMBER ff03 parameters (The strategy for deriving the parameters are employed from Ref.[1] except that the parameters for the vdW radii of liganding atoms are taken from AMBER ff03 force field). In short, it is assumed that the transferred charge is negligible when the distance between the Zn(II) and liganding atom is larger than the sum of their vdW radii. In calculating the transferred charge at equilibrium distance, the model systems are optimised at B3-LYP/6-31+G* level. In this work, the side-chain of cysteine and histidine are modelled by methylthiolate and imidazole, respectively. The backbone and the side-chain of glutamine are modelled by formamide. The side-chains of aspartic acid and glutamic acid are modelled by $HCOO^-$. The side-chain of serine, threonine and tyrosine are modelled by methanol. In optimising the system containing the N of backbone, a constraint is added to achieve convergence. The coordination number for water molecule is six. For other cases, the coordination numbers are four.

Table S2. Coordination number (CN), bond length and backbone *rmsd* calculated using the AMBER ff03 forced field with the modifications implemented in this work (ff03-Modified), the CHARMM force field with the modifications by Lim and coworker (Charmm-Lim), and the AMBER ff02 polarizable force field with charge transfer (ff02-Modified) at T=298 K. For comparison, the experimentally determined coordination geometry of Sp1f2 (Sp1f2-NMR) and the Zif268 (Zif268-X ray) are also presented.

Methods	CN	Zn(II)-S (Å)	Zn(II)-N (Å)	*rmsd* (Å)
Zif268-X ray	4	2.29	2.04	
Charmm-Lim	4	2.31	2.04	1.32
Sp1f2-NMR	4	2.25	2.07	1.57
ff02-Modified	4	2.26	2.17	1.94
ff03-Modified	4	2.20	2.08	1.91

Notes: In this table, the results of CHARMM force field with the modifications by Lim and coworker (Charmm-Lim) are taken from Ref. [1] (T=300K). The results for the experimental structure of classical zinc-finger Zif268 and Sp1f2 are taken from Ref. [2] and Ref. [3], respectively. The *rmsd* of the sp1f2 in Ref.[3] is the averaged value of 20 NMR structures relative to the mean structure. The *rmsd* values are defined for backbone atoms in this table. During the calculations using the AMBER ff02 polarizable force field, the charge transfer is included with the similar manner as for the AMBER ff03 force field except that the parameters in determining the charge transfer is taken from ff02 force field. The Van der Waals radii of sulfur atom in ff02 force field is slightly modified, and the value in Ref. [4] is used. The polarizability of the Zn(II) is taken from Ref. [5]. With these parameters, the experimental coordination structure of the classical zinc-finger can be reproduced better.

(1) Sakharov, D. V.; Lim, C. *J. Am. Chem. Soc.* **2005**, 127, 4921–4929.

(2) Elrod-Erickson, M.; Rould, M. A.; Nekludova, L.; Pabo, C. O. *Structure (London)* **1996**, 4, 1171–1180.

(3) Narayan, V. A.; Kriwacki, R. W.; Caradonna, J. P. *J. Biol. Chem.* **1997**, 272, 7801–7809.

(4) Bredenberg, J.; and Nilsson, L. *Int. J. Quan. Chem.* **2001**, 83, 230–244.

(5) Sen, K. D, Bartolotti, L. J. *Phys. Rev.* **1992**, A45, 2076–2078.

Table S3. Secondary structure prediction of the individual amino acid of sp1f2 based on APSSP2[1]. The segment Ser17-Arg22, which locates at the first two helical turns of the C-terminal α-helix, is predicted to be helix with high probability of correct prediction.

A. A	S. S.	P. C. P.	A. A.	S. S.	P. C. P.
Phe3	C	0.8	Asp18	H	0.6
Met4	C	0.8	Glu19	H	0.6
Cys5	C	0.9	Leu20	H	0.6
Thr6	C	1.0	Gln21	H	0.6
Trp7	C	0.8	Arg22	H	0.5
Ser8	C	0.6	His23	C	0.5
Tyr9	C	0.8	Lys24	C	0.5
Cys10	C	0.9	Arg25	E	0.6
Gly11	C	0.8	Thr26	E	0.4
Lys12	C	0.8	His27	C	0.4
Arg13	C	0.5	Thr28	C	0.8
Phe14	C	0.5	Gly29	C	1.0
Thr15	C	0.6	Glu30	C	0.9
Arg16	C	0.7	Lys31	C	1.0
Ser17	H	0.6			

Note: A. A. stands for amino acid; S. S. stands for the predicted secondary strucutre; P. C. P. stands for probability of correct prediction; H stands for Helix; E stands for Strand; C stands for Coil.

(1) Raghava, G. P. S. **2000**, CASP4: 75–76.

The complete list of authors for Reference 33 and 45

(33) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, J. A.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Baboul, A. G.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Andres, J. L.; Gonzalez, C.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. (1998) *GAUSSIAN 98*; Gaussian, Inc., Pittsburgh PA.

(45) Case, D. A.; Darden, T. A.; Cheatham III, T. E.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Merz, K. M.; Wang, B.; Pearlman, D. A.; Crowley, M.; Brozell, S.; Tui, V.; Gohlke, H.; Mongan, J.; Hornak, V.; Cui, G.; Beroza, P.; Schafmeister, C.; Caldwell, J. W.; Ross, W. S.; Kollman, P. A. (2004) *AMBER 8*, University of California, San Francisco.