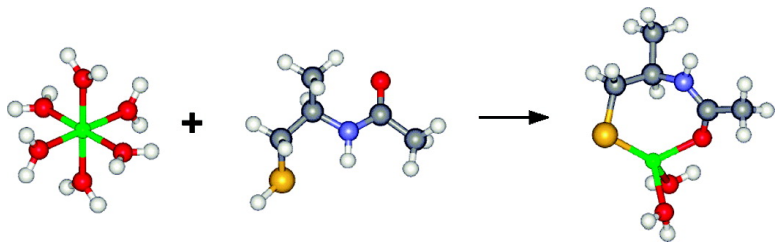
Article

All-Electron Calculations of the Nucleation Structures in Metal-Induced Zinc-Finger Folding: Role of the Peptide Backbone

Todor Dudev, and Carmay Lim

J. Am. Chem. Soc., **2007**, 129 (41), 12497-12504• DOI: 10.1021/ja073322c • Publication Date (Web): 21 September 2007



J|A|C|S
ARTICLES
Published on Web 09/21/2007

All-Electron Calculations of the Nucleation Structures in Metal-Induced Zinc-Finger Folding: Role of the Peptide Backbone

Todor Dudev[†] and Carmay Lim*[,†,‡]

Contribution from the Institute of Biomedical Sciences, Academia Sinica, Taipei 115, Taiwan, and Department of Chemistry, National Tsing Hua University, Hsinchu 300, Taiwan

Received May 21, 2007; E-mail: carmay@gate.sinica.edu.tw

Abstract: Although the folding of individual protein domains has been extensively studied both experimentally and theoretically, protein folding induced by a metal cation has been relatively understudied. Almost all folding mechanisms emphasize the role of the side-chain interactions rather than the peptide backbone in the protein folding process. Herein, we focus on the thermodynamics of the coupled metal binding and protein folding of a classical zinc-finger (ZF) peptide, using all-electron calculations to obtain the structures of possible nucleation centers and free energy calculations to determine their relative stability in aqueous solution. The calculations indicate that a neutral Cys first binds to hexahydrated Zn^{2+} via its ionized sulfhydryl group and neutral backbone oxygen, with the release of four water molecules and a proton. Another nearby Cys then binds in the same manner as the first one, yielding a fully dehydrated Zn^{2+}. Subsequently, two His ligands from the C-terminal part of the peptide successively dislodge the Zn-bound backbone oxygen atoms to form the native-like Zn-$(Cys)_2(His)_2$ complex. Each successive Zn complex accumulates increasingly favorable and native interactions, lowering the energy of the ZF polypeptide, which concomitantly becomes more compact, reducing the search volume, thus guiding folding to the native state. In the protein folding process, not only the side chains but also the backbone peptide groups play a critical role in stabilizing the nucleation structures and promoting the hydrophobic core formation.

Introduction

Folding of a protein to its unique three-dimensional (3D) structure is crucial for its proper functioning in vivo. Elucidating how a single polypeptide chain of L-amino acid (aa) residues self-assembles spontaneously under physiological conditions into a functional native conformation remains one of the most important unresolved problems in molecular biology. Several protein folding mechanisms have been proposed, which all capitalize on Levinthal's finding that a given polypeptide chain cannot sample the entire conformational space *randomly*/*unguided* to achieve its native fold in real time.[1] Thus, a *guided* conformational search, which limits the number of possible protein structures, is the heart of each protein folding model.

Many plausible protein folding models have attempted to resolve Levinthal's paradox, including (i) the framework model, (ii) the hydrophobic-collapse model, (iii) the "nucleation−condensation" model, and (iv) the folding funnel. To narrow the conformational search, the framework model hypothesizes a stepwise mechanism involving a hierarchical assembly: first, local secondary structure elements form independently in different segments of the polypeptide chain, and then these subunits diffuse until they collide to form the respective tertiary structure (the "diffusion-collision" model[2]). The hydrophobic-collapse model[3,4] also hypothesizes a stepwise mechanism involving the formation of kinetic intermediates, where not all the secondary structures and native contacts are fully formed. Unlike the framework model, the initial drive toward protein folding is the rapid collapse of the polypeptide chain around hydrophobic aa side chains, sequestering them from the surrounding water, while the polar and charged residues are left on the solvent-exposed surfaces of the protein; the resulting compact collapsed state (molten globule), containing secondary structure but suboptimally packed side chains, then rearranges to form the native protein conformation. The "nucleation−condensation" model[5−8] postulates that a small number of key residues form spontaneously a nucleus, around which the rest of the polypeptide chain condenses to form the native conformation; i.e., secondary structure formation is concomitant with tertiary structure. In contrast to a single dominant pathway in each of the aforementioned models, the "new view" of protein folding postulates the energy surface of protein folding as a funnel and assumes many paths, from a high-entropy, unfolded

[†] Academia Sinica.
[‡] National Tsing Hua University.

(1) Levinthal, C. In *Mossbauer Spectroscopy in Biological Systems*; Debrunner, P., Tsibris, J. C. M., Munck, E., Eds.; University of Illinois Press: Urbana, IL, 1969; p 22−24.
(2) Karplus, M.; Weaver, D. L. *Protein Sci.* **1994**, *3*, 650−668.
(3) Dill, K. A.; Bromberg, S.; Yue, K. Z.; Fiebig, K. M.; Yee, D. P.; Thomas, P. D.; Chan, H. S. *Protein Sci.* **1995**, *4*, 561−602.
(4) Ptitsyn, O. B. *Nat. Struct. Biol.* **1996**, *3*, 488−490.
(5) Kanehisa, M. I.; Tsong, T. Y. *J. Mol. Biol.* **1978**, *124*, 177−194.
(6) Fersht, A. R. *Curr. Opin. Struct. Biol.* **1997**, *7*, 3−9.
(7) Fersht, A. R. *Structure and mechanism in protein science: a guide to enzyme catalysis and protein folding*; W.H. Freeman: New York, 1999.
(8) Fersht, A. R. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 1525−1529.

state lacking intramolecular interactions to a low-entropy, folded state with native intramolecular interactions. A common accepted feature of all the folding mechanisms is that the side-chain interactions play a dominant role in the folding process. This view, however, has been challenged recently by Rose et al.,[9] who suggested that it is the energetics of the backbone hydrogen bonds that dominates the polypeptide self-assembly process.

Although many proteins fold into their native conformation before binding to their functional partner (another protein, DNA/RNA, or low-molecular-weight substrate), some biologically important proteins that are involved in critical aspects of biological regulation and signal transduction are not fully folded until they bind to their cognate partner; i.e., folding of the polypeptide chain proceeds in concert with substrate binding.[10,11] A metal cation could be considered as an example of the simplest substrate that assists/induces the folding process. With its ability to interact favorably with some functional groups from the polypeptide chain and form a stable complex with strong coordinative bonds, a metal cation may, in some proteins, act as a nucleation center around which the rest of the protein folds.[12] Perhaps the best-known example of metal-induced protein folding is the group of zinc-finger (ZF) proteins, small transcription factors involved in nucleic acid binding, gene expression regulation, reverse transcription, and virus assembly, which require Zn^{2+} for folding. The ZF peptide is largely unstructured in the absence of metal cations, but Zn^{2+} binding results in fast folding.[13−19]

Among the ZF protein family, the mechanism of metal-induced folding in the so-called "classical" ZFs, first identified in transcription factor IIIA,[20,21] has been studied the most. "Classical" ZF proteins have a modular structure and consist of several independently folding domains (fingers), each comprising about 30 aa residues and a Zn^{2+} tetrahedrally bound to two Cys and two His. The "classical" ZF domain adopts a $\beta\beta\alpha$ fold, with two antiparallel β-strands forming a β-sheet in the N-terminal region and an α-helix in the C-terminal region. It has the following characteristic sequence:[22] X_2-**C**-$X_{2,4}$-**C**-X_{12}-**H**-X_{3-5}-**H**, where X is any aa and the two Cys and two His in bold are conserved Zn^{2+} ligands.[14] Experimental studies have revealed the following features about the coupled metal binding and ZF folding process: (i) The overall reaction between an unfolded ZF and the hydrated Zn^{2+} to form a tetrahedral [Zn-(Cys_2-His_2)]^0 core is enthalpically driven; the unfavorable entropy associated with ZF folding is largely compensated by the favorable entropy associated with water release from Zn^{2+} upon Cys/His binding.[23] (ii) Between one and two protons appear to be released upon metal binding to the ZF peptide.[23] (iii) Metal-induced ZF folding is a stepwise reaction, where the two Cys residues bind first to Zn^{2+}, followed by the two His.[24,25] (iv) In the absence of Zn^{2+}, the 27-mer ZF3 peptide corresponding to the mouse Zif268 third ZF domain adopts a β-sheet-rich conformation, but binding of Zn^{2+} to Cys and His induces a partial transition from β-sheet to α-helix.[25] Swapping the two Cys ligands for His yields a ZF3(HHCC) peptide that cannot fold into the ZF structure.[25] (v) The highly conserved hydrophobic residues (Tyr, Phe, and Leu) play an important role in stabilizing the hydrophobic core of the folded protein.[26,27] (vi) Non-transition-metal ions such as Mg^{2+}, Ca^{2+}, and Ba^{2+} do not induce folding;[28−30] transition metal ions such as Ni^{2+} and Cu^{2+} lead to a non-native ZF fold,[20,31,32] but Co^{2+}, like Zn^{2+}, can induce folding to the native 3D structure.[18,33]

Despite the wealth of information accumulated on the metal-induced folding of "classical" ZFs, many intriguing questions about how the Zn^{2+} can induce folding to the native 3D structure and the steps involved in the ZF folding process remain elusive. Although the initial step of the metal-induced folding is known to involve the binding of Cys to Zn^{2+}, the structure and relative stability of possible nucleation centers are not known. In other words, does each metal ligand bind to Zn^{2+} simply by successively replacing a metal−O(water) bond with a metal−S(Cys^-) or a metal−N(His^0) bond? Do the backbone peptide groups participate in concert with the Cys/His side chain in initially trapping Zn^{2+}? Does Zn^{2+} become fully dehydrated, thus forming a hydrophobic core, only after it is bound to the two Cys and two His ligands?

Our goal is to determine the structures and relative stability of possible nucleation centers in the ZF folding process to enhance our understanding of the molecular mechanism of coupled metal (substrate) binding and protein folding. Density functional theory (DFT), in conjunction with the continuum dielectric method (CDM), was used to model the stepwise binding of Zn^{2+} to two Cys and two His in model ZF peptides (see Methods). The structures of the Zn^{2+} complexed with one, two, three, and four aa ligands were fully optimized at the S-VWN/(6-31+G*/GEN) level. Subsequently, the enthalpy, entropy, and free energy of each reaction in aqueous solution were evaluated. This allows us to assess the relative stability of different nucleation centers and suggest the energetically most favorable nucleation complexes along the ZF folding pathway. To the best of our knowledge, this is the first all-electron study of the nucleation structures and nucleation process in metal-

(9) Rose, G. D.; Fleming, P. J.; Banavar, J. R.; Maritan, A. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 16623−16633.
(10) Eliezer, D.; Palmer, A. G., III. *Nature* **2007**, *447*, 920−921.
(11) Sugase, K.; Dyson, H. J.; Wright, P. E. *Nature* **2007**, *447*, 1021−1025.
(12) Wilson, C. J.; Apiyo, D.; Wittung-Stafshede, P. *Q. Rev. Biophys.* **2004**, *37*, 285−314.
(13) Parraga, G.; Horvath, S.; Eisen, A.; Taylor, W. E. *Science* **1988**, *241*, 1489−1492.
(14) Frankel, A. D.; Berg, J. M.; Pabo, C. O. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 4841−4845.
(15) Krizek, B. A.; Amann, B. T.; Kilfoil, V. J.; Merkle, D. L.; Berg, J. M. *J. Am. Chem. Soc.* **1991**, *113*, 4518−4523.
(16) Michael, S. F.; Kilfoil, V. J.; Schmidt, M. H.; Amann, B. T.; Berg, J. M. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 4796−4800.
(17) Eis, P. S.; Lakowicz, J. R. *Biochemistry* **1993**, *32*, 7981−7993.
(18) Berg, J. M.; Godwin, H. A. *Annu. Rev. Biophys. Biomol. Struct.* **1997**, *26*, 357−371.
(19) Cox, E. H.; McLendon, G. L. *Curr. Opin. Chem. Biol.* **2000**, *4*, 162−165.
(20) Hanas, J. S.; Hazuda, D. J.; Bogenhagen, D. F.; Wu, F.-H.; Wu, C.-W. *J. Biol. Chem.* **1983**, *258*, 14120−14125.
(21) Miller, J.; McLachlan, A. D.; Klug, A. *EMBO J.* **1985**, *4*, 1609−1614.
(22) Bairoch, A.; Butcher, P.; Hofmann, K. *Nucleic Acids Res.* **1997**, *25*, 217−221.
(23) Blasie, C. A.; Berg, J. M. *Biochemistry* **2002**, *41*, 15068−15073.
(24) Shi, Y.; Beger, R. D.; Berg, J. M. *Biophys. J.* **1993**, *64*, 749−753.
(25) Miura, T.; Satoh, T.; Takeuchi, H. *Biochim. Biophys. Acta* **1998**, *1384*, 171−179.
(26) Jasanoff, A.; Weiss, M. A. *Biochemistry* **1993**, *32*, 1423−1432.
(27) Qian, X.; Weiss, M. A. *Biochemistry* **1992**, *31*, 7463−7476.
(28) Zawia, N. H.; Sharan, R.; Brydie, M.; Oyama, T.; Crumpton, T. *Dev. Brain Res.* **1998**, *107*, 291−298.
(29) Razmiafshari, M.; Zawia, N. H. *Toxicol. Appl. Pharmacol.* **2000**, *166*, 1−12.
(30) Razmiafshari, M.; Kao, J.; d'Avignon, A.; Zawia, N. H. *Toxicol. Appl. Pharmacol.* **2001**, *172*, 1−10.
(31) Waldschmidt, R.; Schneider, H. R.; Seifart, K. H. *Nucleic Acids Res.* **1991**, *19*, 1455−1459.
(32) Thiesen, H.-J.; Bach, C. *Biochem. Biophys. Res. Commun.* **1991**, *176*, 551−557.
(33) Lachenmann, M. J.; Ladbury, J. E.; Dong, J.; Huang, K.; Carey, P.; Weiss, M. A. *Biochemistry* **2004**, *43*, 13910−13925.

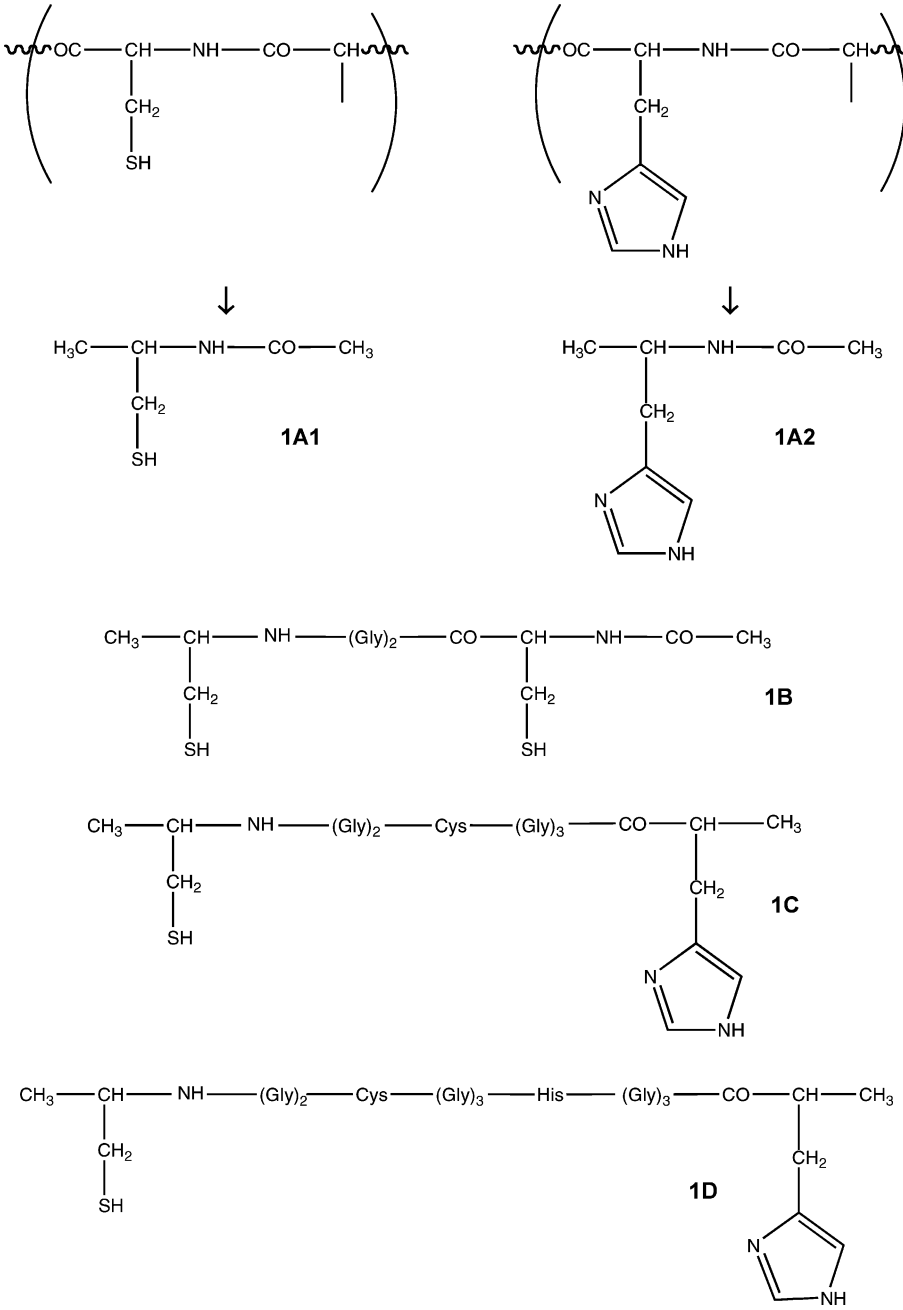


Figure 1. Models used in the calculations.

induced protein folding, providing insight into the molecular mechanism of coupled metal (substrate) binding and folding.

Methods

DFT Calculations. (a) Geometries. For each metal complex, several different conformers were initially generated using molecular graphics (GaussView)[34] and then subjected to full geometry optimization using the Gaussian 03 program.[35] The S-VWN functional[36,37] and the 6-31+G* basis set[38] were chosen for geometry optimization using the standard Berny algorithm,[39] as it reproduces the experimentally observed metal−ligand bond distances in a number of metal−ligand complexes within experimental error.[40] Due to the large size of the Cys-$(Gly)_2$-Cys-$(Gly)_3$-His-$(Gly)_3$-His dodecapeptide-like molecule (Figure 1, **1D**) involved in the last binding reaction, a lower level basis set, designated as GEN (comprising of the 3-21+G* basis set for H, O, S, C, and N, and the 6-31+G* basis set for Zn), was used in evaluating the optimized geometries of the largest complexes. For each fully optimized structure, the respective S-VWN/6-31+G* or S-VWN/GEN vibrational frequencies were computed to verify that the molecule was at the minimum of its potential energy surface. No imaginary frequency was found in any of the metal complexes.

(b) Gas-Phase Free Energies. Based on the fully optimized S-VWN/(6-31+G*/GEN) geometries, the electronic energies, E_{elec}, were evaluated using the B3-LYP functional in conjunction with the large 6-311++G(2d,2p) basis set. The latter was chosen from among several

(34) *GaussView 3.09*; Gaussian, Inc.: Pittsburgh, PA, 2000−2003.
(35) Frisch, M. J.; et al. *Gaussian 03*, rev. B.03; Gaussian, Inc.: Pittsburgh, PA, 2003.
(36) Slater, J. C. *Quantum theory of molecules and solids. Vol. 4: The self-consistent field for molecules and solids*; McGraw-Hill: New York, 1974.
(37) Vosko, S. H.; Wilk, L.; Nusair, M. *Can. J. Phys.* **1980**, *58*, 1200−1211.
(38) Hariharan, P. C.; Pople, J. A. *Theor. Chim. Acta* **1973**, *28*, 213−222.
(39) Peng, C. Y.; Ayala, P. Y.; Schlegel, H. B.; Frisch, M. J. *J. Comput. Chem.* **1996**, *17*, 49−56.
(40) Dudev, T.; Lim, C. *J. Am. Chem. Soc.* **2006**, *128*, 1553−1561.

Scheme 1

$$\begin{array}{ccc} & \Delta G^{1} & \\ \text{Reactants } (\varepsilon = 1) & \rightarrow & \text{Products } (\varepsilon = 1) \\ \Delta G_{solv}{}^{x}\,(\text{Reactants}) \downarrow & & \downarrow \Delta G_{solv}{}^{x}\,(\text{Products}) \\ \text{Reactants } (\varepsilon = 80) & \rightarrow & \text{Products } (\varepsilon = 80) \\ & \Delta G^{80} & \end{array}$$

other basis sets because it reproduces reasonably well the experimental gas-phase free energy of deprotonating H_2S, modeling the Cys sulfhydryl group: $\Delta G^{calc} = 344.2$ kcal/mol, while $\Delta G^{exp} = 345.4 \pm 4$,[41] 345.6 ± 2,[42] and 347.1 ± 2 kcal/mol.[43] The thermal energy, including zero-point energy (E_T), work (PV), and the translational, rotational, and vibrational entropy (S) corrections were evaluated using standard statistical mechanical formulas[44] with the S-VWN/(6-31+G*/GEN) frequencies scaled by an empirical factor of 0.9833.[45] The differences ΔE_{elec}, ΔE_T, ΔPV, and ΔS between the products and reactants were used to compute the reaction free energy in the gas phase at room temperature, $T = 298.15$ K, according to the following expression:

$$\Delta G^{1} = \Delta E_{elec} + \Delta E_{T} + \Delta PV - T\Delta S \qquad (1)$$

Continuum Dielectric Calculations. As the reactions are considered to occur on solvent-exposed surfaces of the unfolded peptide chain, the free energy in aqueous solution, ΔG^{80}, was calculated according to Scheme 1, where ΔG^{1}, the gas-phase free energy, was computed using eq 1, and ΔG^{80}_{solv}, the solvation free energy, was estimated by solving Poisson's equation using finite difference methods.[46,47] Thus, the reaction free energy in aqueous solution, ΔG^{80}, was computed from

$$\Delta G^{80} = \Delta G^{1} + \Delta G^{80}_{solv}(\text{products}) - \Delta G^{80}_{solv}(\text{reactants}) \qquad (2)$$

The solvation free energies were evaluated using the MEAD (macroscopic electrostatics with atomic detail) program,[48] as described in previous works.[40,48] The effective solute radii, which were obtained by adjusting the CHARMM (version 22)[49] van der Waals radii to reproduce the experimental hydration free energies of the Zn^{2+} cation and model ligand molecules, are as follows (in Å): $R_{Zn} = 1.40$, $R_S = 2.19$, $R_C = 1.88$, $R_N(HCONH_2) = 1.75$, R_N(imidazole) $= 1.80$, $R_O(HCONH_2) = 1.78$, $R_O(H_2O-Zn) = 1.70$, $R_H = 1.468$, and $R_H(H_2O-Zn) = 1.08$. These effective solute radii consistently overestimate the experimental hydration free energies of the metal cation and ligand molecules by less than 1 kcal/mol (see Supporting Information, Table 1). The experimental values of ΔG^{80}_{solv} for H^+ (−264.0 kcal/mol[50]) and/or H_2O (−6.3 kcal/mol[51]) were used in computing the solution free energies.

Results

Step 1: The First ZF Ligand Binds Preferentially via Its Side-Chain Atom and Its Backbone Oxygen to Zn^{2+}. To elucidate how the first Cys/His ligand binds to Zn^{2+}, we examined Zn^{2+} bound to Cys and His, modeled by **1A1** and **1A2**, respectively, in Figure 1. The cysteine-like (**1A1**) and histidine-like (**1A2**) ligands, which contain all potential ligating entities (viz., the side-chain S/N and the backbone O/N), are sufficient to describe the binding mode of the first Cys/His ligand to Zn^{2+} and the initial nucleation structures, assuming that the other aa residues in the ZF peptide do not participate in metal binding. These ligands can bind Zn^{2+} in the following three modes: (1) monodentately via the side-chain S(Cys^-)/N(His) (Figures 2 and 3, **2A** and **3A**); (2) bidentately via the side-chain S(Cys^-)/N(His) and the backbone *nitrogen* (**2B** and **3B**); and (3) bidentately via the side-chain S(Cys^-)/N(His) and the backbone *oxygen* (**2C** and **3C**). The interactions between the hydrated Zn^{2+} and Cys/His ligands are considered to take place at solvent-accessible surfaces of the unfolded ZF peptide (see Methods). The chelation bidentate metal coordination of a side-chain S(Cys^-)/N(His) and a backbone oxygen yields the most stable structure. Relative to the free energy of the (S,O)-bidentate seven-membered-ring structure (**2C**), the formation free energies of the S-monodentate (**2A**) and (S,N)-bidentate five-membered-ring (**2B**) complexes are higher by 14.7 and 15.5 kcal/mol, respectively (Figure 2). Likewise, relative to the free energy of the (N,O)-bidentate eight-membered-ring complex (**3C**), the formation free energies of the N-monodentate (**3A**) and (N,N)-bidentate six-membered-ring (**3B**) structures are higher by 3.6 and 19.3 kcal/mol, respectively (Figure 3). Compared to the backbone nitrogen, the more electronegative backbone oxygen has higher affinity for Zn^{2+}, as it can donate more charge to the metal ion ($q^{donated}_{S,O} - q^{donated}_{S,N} = 0.15$e) and can form a larger ring structure that is presumably less strained. Furthermore, it can displace another Zn-bound water molecule from the S/N-monodentate complex to form a more stable seven- or eight-membered-ring structure.

Zn^{2+} Prefers Cys to His. To evaluate whether Zn^{2+} prefers to bind to Cys or His, the free energy of Zn^{2+} binding to Cys/His in aqueous solution to form the lowest energy Zn-Cys (**2C**) or Zn-His complexes (**3C**) was computed, as described in Methods. In aqueous solution, Zn^{2+} is modeled as $[Zn(H_2O)_6]^{2+}$, as it is experimentally determined to be hexahydrated.[52] Upon binding to a protein ligand, Zn^{2+} is assumed to be tetracoordinated, as it is found to be tetracoordinated in all 3D structures of structural Zn proteins.[53] In the unfolded metal-free ZF peptide at a physiological pH of 7, Cys side chains with typical pK_a values between 8 and 9[54,55] would be protonated, while His side chains with pK_a values close to 7[56] could be neutral or protonated. However, upon binding to Zn^{2+}, the neutral Cys[57] and positively charged His (see Supporting information) would be deprotonated.

The solution free energies, ΔG^{80}, for hexahydrated Zn^{2+} binding to a neutral Cys (−12.1 kcal/mol, Figure 4a), a protonated His (−3.1 kcal/mol, Figure 4b), and a neutral His (−6.4 kcal/mol, number in parentheses in Figure 4b) in aqueous solution show that Zn^{2+} prefers binding to Cys rather than His.

(41) Rosenstock, H. M.; Draxl, K.; Steiner, B. W.; Herron, J. T. *J. Phys. Chem. Ref. Data* **1977**, *6*, Suppl. 1.
(42) Cumming, J. B.; Kebarle, P. *Can. J. Chem.* **1978**, *56*, 1−9.
(43) Bartmess, J. E.; McIver, R. T., Jr. *Gas phase ion chemistry*; Academic Press: New York, 1979; Vol. 2.
(44) McQuarrie, D. A. *Statistical Mechanics*; Harper and Row: New York, 1976.
(45) Wong, M. W. *Chem. Phys. Lett.* **1996**, *256*, 391−399.
(46) Gilson, M. K.; Honig, B. *Biopolymers* **1986**, *25*, 2097−2119.
(47) Lim, C.; Bashford, D.; Karplus, M. *J. Phys. Chem.* **1991**, *95*, 5610−5620.
(48) Bashford, D. In *Scientific Computing in Object-Oriented Parallel Environments*; Ishikawa, Y., Oldehoeft, R. R., Reynders, V. W., Tholburn, M., Eds.; Springer: Berlin, 1997; Vol. 1343, pp 233−240.
(49) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187−217.
(50) Tissandier, M. D.; Cowen, K. A.; Feng, W. Y.; Gundlach, E.; Cohen, M. H.; Earhart, A. D.; Coe, J. V.; Tuttle, T. R., Jr. *J. Phys. Chem. A* **1998**, *102*, 7787−7794.
(51) Ben-Naim, A.; Marcus, Y. *J. Chem. Phys.* **1984**, *81*, 2016−2027.
(52) Marcus, Y. *Chem. Rev.* **1988**, *88*, 1475 - 1498.
(53) Li, Y.-M.; Lim, C., manuscript in preparation.
(54) Dawson, R. M. C.; Elliot, D. C.; Elliot, W. H.; Jones, K. M. *Data for Biochemical Research*, 3rd ed.; Clarendon Press: Oxford, 1995.
(55) Bombarda, E.; Morellet, N.; Cherradi, H.; Spiess, B.; Bouaziz, S.; Grell, E.; Roques, B. P.; Mely, Y. *J. Mol. Biol.* **2001**, *310*, 659−672.
(56) Stryer, L. *Biochemistry*, 4th ed.; W.H. Freeman and Co.: New York, 1995.
(57) Dudev, T.; Lim, C. *J. Am. Chem. Soc* **2002**, *124*, 6759−6766.

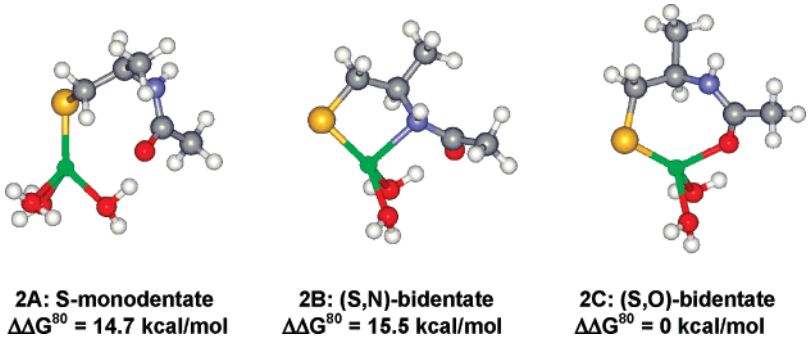


Figure 2. Fully optimized S-VWN/6-31+G* geometries of Zn-Cys complexes and their free energies relative to the free energy of the (S,O)-bidentate structure (in kcal/mol).

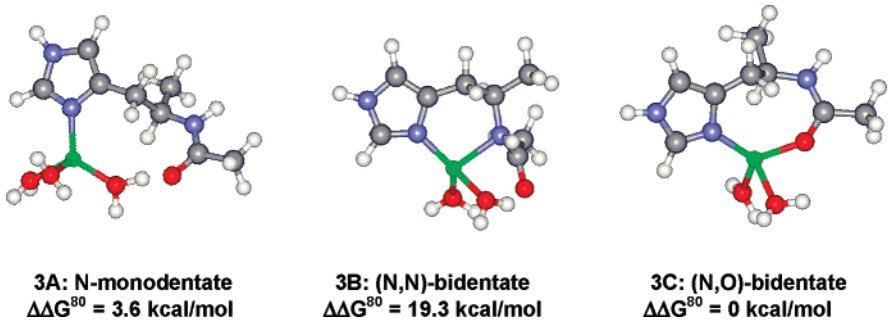


Figure 3. Fully optimized S-VWN/6-31+G* geometries of Zn-His complexes and their free energies relative to the free energy of the (N,O)-bidentate structure (in kcal/mol).

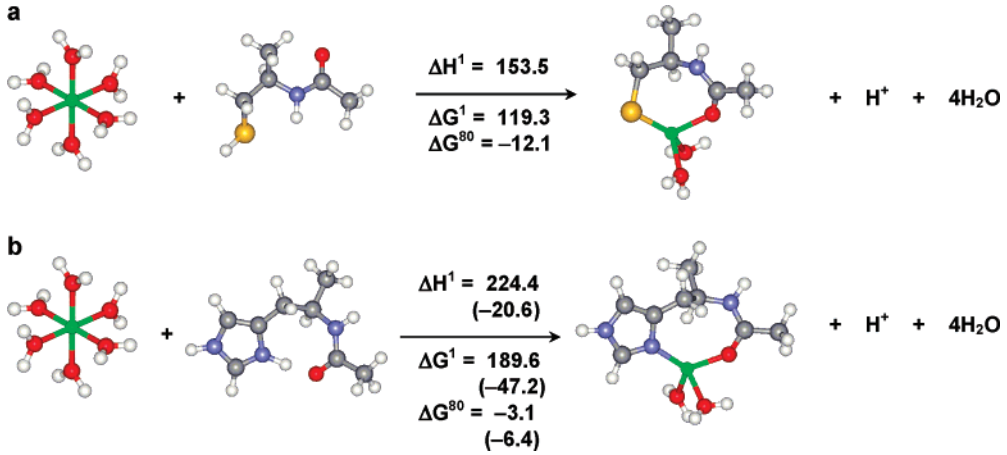


Figure 4. Gas-phase enthalpy (ΔH^1) or free energy (ΔG^1) and solution free energy (ΔG^{80}) for the binding of hexahydrated Zn^{2+} to protonated (a) Cys and (b) His in aqueous solution (in kcal/mol). The numbers in parentheses correspond to the ΔH^1, ΔG^1, and ΔG^{80} values for the binding of a *deprotonated* His^0 to Zn^{2+} in aqueous solution.

More direct support that Zn^{2+} binds first to Cys and then to His comes from a 3.4 ns molecular dynamics simulation of the Zif268 Cys_2-His_2 ZF peptide in TIP3P[58] water at 500 K, using the CHARMM force field,[49] which shows that the His ligands detach from Zn^{2+} faster than the respective Cys side chains during the unfolding process (see Supporting Information, Figure 1; D. Sakharov and C. Lim, unpublished results). This is in line with the experimental observation that Zn^{2+} coordinates preferentially to Cys prior to His in metal-assisted ZF folding.[25]

If the ZF peptide were partially folded such that the Cys/His ligands were solvent-inaccessible, Zn^{2+} would be predicted *not* to bind to protonated Cys/His (large positive ΔG^1, Figure 4), whereas it could still bind to neutral His ($\Delta G^1 = -47.2$ kcal/mol). In fact, neutral His binds Zn^{2+} with a more favorable free energy gain than protonated His. Note that, for both Zn-Cys and Zn-His reactions, the release of a proton and/or four water molecules upon metal binding contributes favorably to the reaction entropy, the $T\Delta S^1$ values ranging from 27 (for neutral His) to 34 kcal/mol (neutral Cys, Figure 4a, and protonated His, Figure 4b).

Step 2: Zn^{2+} Binds Preferentially to the Second Cys Side-Chain Sulfur and Backbone Oxygen. As Zn^{2+} coordinates preferentially to Cys rather than His and the two Cys ligands are separated by only 2−4 residues (see sequence motif in Introduction), Zn^{2+} will likely bind another Cys rather than His, which is separated from the nearest Cys by 12 residues. To elucidate the binding mode of the second Cys ligand, it was bound to Zn^{2+} (1) monodentately via its sulfhydryl S atom (**5A**)

(58) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926−923.

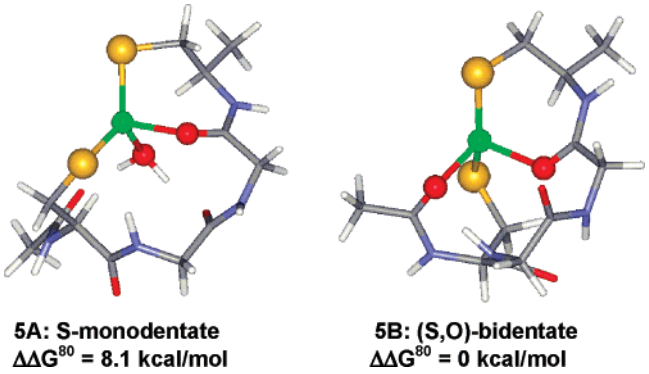


Figure 5. Fully optimized S-VWN/6-31+G* geometries of Zn-(Cys-Gly-Gly-Cys) complexes and the respective free energy relative to that of the (S,O)-bidentate structure (in kcal/mol).

or (2) bidentately via its side-chain S and backbone O (**5B**), while the first Cys is bound to Zn^{2+} in the preferred (S,O)-bidentate mode (Figure 5). The **Cys**-$X_{2,4}$-**Cys** segment of the ZF peptide (see Introduction) was modeled by the Cys-Gly-Gly-Cys peptide (Figure 1, **1B**), where a minimalist $(Gly)_2$ spacer was used to decrease computational cost. As for the first Cys binding, the (S,O)-bidentate mode is more stable than the respective S-monodentate mode. However, the free energy difference between the Zn-Cys_2 structures **5A** and **5B** (8.1 kcal/mol) is smaller than that between the single Cys structures **2A** and **2C** (14.7 kcal/mol).

Zn^{2+} Is Dehydrated upon Binding the Second Cys. To evaluate whether binding of a second Cys to the Cys-bound Zn^{2+} is favorable or not, the free energy for binding of a second neutral Cys to Zn^{2+} with concomitant proton release was computed. The resulting free energy, $\Delta G^{80} = -4.0$ kcal/mol (corresponding to the reaction in Figure 6a but with the free Cys ligand protonated rather than deprotonated), indicates that binding of a neutral Cys to the $[Zn\text{-}Cys]^+$ complex is thermodynamically favorable in aqueous solution. However, it is accompanied by a smaller free energy gain, as compared to binding of the first Cys^0 to hydrated Zn^{2+} (Figure 4a). This is probably because (i) the positive charge on Zn^{2+} is greatly reduced by charge transfer from the first Cys^-, thus reducing Zn's ability to stabilize the second Cys^-;[57] (ii) compared to the binding of the first Cys^0 to hydrated Zn^{2+}, fewer water molecules are liberated when the second Cys^0 binds to Zn^{2+} (2 vs 4 H_2O), resulting in a less favorable entropic gain [the net translational, rotational, and vibrational entropy change, $T\Delta S^1$, for the reaction involving the first Cys^0 shown in Figure 4a (34 kcal/mol) is less positive than that (23 kcal/mol) for the binding of the second Cys^0 to Zn^{2+}]; (iii) the **5B** complex, with three intramolecular rings, is expected to be more strained than the corresponding **2C** complex, with only one ring. Notably, at this stage, the last two water molecules from Zn's hydration sphere are released to bulk solution, thus creating a hydrophobic core around Zn^{2+}.

The $[Zn\text{-}Cys]^+$ Complex Binds Preferentially to a Deprotonated Cys. Prior to binding Zn^{2+}, the second Cys may be deprotonated due to its proximity to the cationic $[Zn\text{-}Cys]^+$ complex. Binding of a deprotonated Cys^- to the $[Zn\text{-}Cys]^+$ complex results in a higher free energy gain ($\Delta G^{80} = -33.7$ kcal/mol, Figure 6a) than binding of the corresponding neutral Cys with concomitant proton release. This finding is consistent with (i) isothermal titration calorimetry experiments showing that about one proton is released during the metal binding process and (ii) NMR experiments showing that the two His

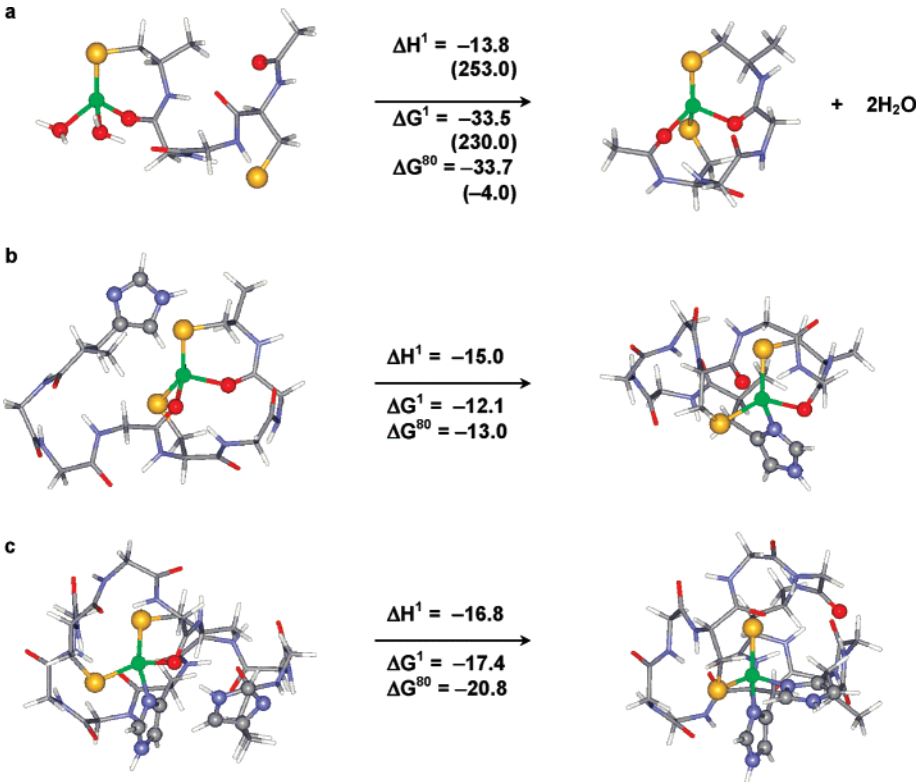


Figure 6. Gas-phase enthalpy (ΔH^1) or free energy (ΔG^1) and solution free energy (ΔG^{80}) for the binding of (a) the second Cys, (b) first His, and (c) second His to Zn^{2+} (in kcal/mol). The numbers in parentheses correspond to the ΔH^1, ΔG^1, and ΔG^{80} values for the binding of a *protonated* Cys^0 to Zn^{2+} with the concomitant release of a proton to aqueous solution.

ligands in the free ZF peptide have a pK_a of ~6.5,[15] indicating that they are predominantly neutral rather than protonated at pH 7, and therefore no His protons need to be released upon binding Zn^{2+}.

Step 3: His Can Displace a Zn-Bound Carbonyl Oxygen. Although the backbone nitrogen binds Zn^{2+} less tightly than the respective backbone oxygen (see above), can the His side-chain nitrogen dislodge the peptide backbone oxygen from Zn^{2+} in the dehydrated Zn-Cys_2 complex, **5B**? The two His ligands are separated from the Cys ligands by at least 12 residues. Ab initio energy and frequency calculations of such a long and floppy peptide chain, however, would currently be computationally prohibitive with the basis set employed in this work. As our interest is in the thermodynamic (rather than the kinetic) outcome of the competition between Cys/His ligands for Zn^{2+}, we first assumed a short spacer between the His and Cys ligands and subsequently checked the sensitivity of the Zn-His interaction energy on the spacer length. Using a $(Gly)_3$ spacer, the difference in electronic energy (at the B3LYP/6-311++G(2d,-2p) level) between the metal-bound His ligand (Figure 6b, right) and the corresponding free His with Zn^{2+} bound by two Cys (Figure 6b, left) is −15.1 kcal/mol, which is similar to that obtained using a much longer $(Gly)_9$ spacer (−15.8 kcal/mol) at the same B3LYP/6-311++G(2d,2p) level.

Hence, the Cys-$(Gly)_2$-Cys-$(Gly)_3$-His ligand (Figure 1, **1C**) with the two Cys side-chain S^- and backbone O atoms tetracoordinated to Zn^{2+} was used to evaluate if neutral His could bind Zn^{2+}. The negative gas-phase and solution free energies (Figure 6b) show that neutral His can compete with the backbone oxygen atom for Zn^{2+} in a solvent-inaccessible or -accessible site. Note that even though a longer spacer may not affect the thermodynamic outcome of the binding of His to the dehydrated Zn-Cys_2 complex, it would be expected to affect the corresponding kinetics by decreasing the probability of finding a ZF peptide conformation where the His is within attacking distance of Zn^{2+}.

Why the His Side-Chain N Binds Tighter than the Peptide Backbone O to Zn^{2+}. That displacing a Zn-bound carbonyl oxygen by His is energetically favorable is consistent with the higher affinity of His for Zn^{2+} relative to that of the backbone carbonyl oxygen found in previous work.[59] An energy component analysis of the interactions between Zn^{2+} and imidazole (modeling the His side chain) and formamide (modeling the backbone peptide group) has shown that, compared to formamide, the higher polarizability and better charge-donating ability of imidazole contribute to its higher affinity for Zn^{2+}.[59] This is reflected in the respective polarization and charge-transfer energy components, which are by 7.1 and 10.1 kcal/mol, respectively, in favor of imidazole, while the net charge transferred from imidazole to Zn^{2+} (0.50e) is greater than that from $HCONH_2$ (0.37e).[59]

Step 4: Binding of the Second His to Zn^{2+} Is Thermodynamically Even More Favorable than That of the First His. Since the two His in ZF peptides are separated by 3−5 residues, a minimalistic model of the end segment with the two His separated by three Gly was used to verify that the last His ligand can indeed bind to Zn^{2+}. The free energy for the *free* His of the Cys-$(Gly)_2$-Cys-$(Gly)_3$-His-$(Gly)_3$-His ligand, which is bound to Zn^{2+} via two Cys and a His (Figure 6c, left structure), binding to Zn^{2+} to yield the native-like Zn-Cys_2-His_2 complex (Figure 6c, right structure) was evaluated as described in Methods. The resulting negative gas-phase and solution free energies given in Figure 6c show that the second neutral His binds to Zn^{2+} with a higher free energy gain than the first neutral His (−20.8 vs −13.0 kcal/mol). The binding site acquires its "classical" Zn-Cys_2-His_2 configuration.

(59) Garmer, D. R.; Gresh, N. *J. Am. Chem. Soc.* **1994**, *116*, 3556−3567.

Discussion

Although previous work had suggested that metal-induced ZF folding is a stepwise reaction, the initial step of which involves first the binding of Cys followed by His to Zn^{2+},[25] it was not known (i) if each of the metal ligands binds to Zn^{2+} via their side chains by successive water displacement from Zn^{2+}, (ii) if the backbone peptide groups also bind to Zn^{2+}, and (iii) when Zn^{2+} becomes dehydrated. The calculations herein show that Zn^{2+} coordinates preferentially to Cys rather than His (Figure 4), suggesting a Cys→Cys→His→His pathway for metal-induced ZF folding, as observed experimentally. They also shed light on the above questions by characterizing the nucleation complexes formed in the course of metal-induced ZF folding.

Figure 7 illustrates the steps involved in the Zn-induced ZF folding process. In the absence of metal ions, the ZF peptide is largely unstructured, especially at the N-terminus, so that the two Cys ligands are solvent-exposed—if the N-terminus were folded such that the two Cys were buried, it would not be thermodynamically favorable for the solvent-inaccessible Cys to bind Zn^{2+} (positive ΔG^1, Figure 4a). In the presence of Zn^{2+}, the solvent-exposed Cys coordinates tetrahedrally to Zn^{2+} via the S atom from the ionized sulfhydryl group and preferentially the O rather than the N atom from an adjacent backbone group. Four water molecules and a proton are released to bulk solution to form the first nucleation center, **7A**, contributing significantly to the reaction entropy. The second Cys, which is close (in sequence) to the first Cys and may be deprotonated as it approaches the positively charged Zn-Cys complex, binds Zn^{2+} in the same (S,O)-bidentate mode, liberating the remaining two Zn-bound water molecules. Thus, in two reaction steps, Zn^{2+} is completely dehydrated, forming the hydrophobic core **7B**. At this stage, the metal cation has immobilized two Cys and two backbone groups, reducing significantly the conformational freedom of that part of the peptide chain, while the rest of the peptide is free to explore conformational space. When one of the two His ligands from the C-terminal part of the peptide comes sufficiently close to the Zn-$(Cys)_2$ complex, the His side chains can successively dislodge the Zn-bound backbone oxygen atoms to form the native-like Zn-$(Cys)_2(His)_2$ complex, **7D**. Binding of the two His to Zn^{2+} further limits the degrees of freedom of the respective segments of the peptide chain.

Of the four nucleation structures shown in Figure 7, forming the Zn-$(Cys)_2$(His) complex **7C** from the Zn-$(Cys)_2$ complex **7B** is probably rate-limiting. This is because, in the consensus classical ZF sequence (see Introduction), the two Cys ligands are separated by 2−4 residues, and thus it is easy to encounter a conformation where the second Cys is close enough to the single Cys complex **7A** to displace two water molecules and form **7B**. Likewise, as the two His ligands are separated by 3−5 residues, it is also relatively easy to find a conformation where the second His is in the vicinity of the Zn-$(Cys)_2$(His)

complex **7C** to displace a backbone oxygen and form **7D**. However, because of the long 12-mer spacer between the Cys and His ligands, we hypothesize that it is rarer to encounter a conformation where one of the His ligands is close enough to the Zn-$(Cys)_2$ complex to displace a backbone oxygen and bind to Zn^{2+}. All-atom folding/unfolding MD simulations of a classical Zn-finger peptide support a high energy barrier for the binding of the first His to Zn^{2+} and a similar folding process, in which the Zn-bound water oxygen atoms are first replaced by two Cys sulfur atoms and two nearby backbone oxygen atoms, which are then replaced by the His nitrogen atoms (Wei Wang, private communication).

The nucleation structures depicted in Figure 7 appear to be consistent with available experimental findings. Replacing a His ligand with Cys in the classical ZF peptide would yield nucleation structures like **7C** and/or **7D** with the Zn−N(His) bond replaced by the Zn−S(Cys^-) bond. Thus, there are ZFs with Zn^{2+} bound to three or four Cys (CCCH and CCCC binding motifs, respectively). Replacing one of the Cys/His ligands with Gly/Ala in the classical ZF peptide would also yield nucleation structures similar to those depicted in Figure 7, but with a Zn−S(Cys^-) or Zn−N(His) bond replaced by a Zn−O(Gly) bond. Thus, ZF peptides in which Gly/Ala was substituted for one of four Zn ligands, yielding **G**CHH, C**G**HH, CC(**G**/**A**)H, and CCH(**G**/**A**) mutants, retained the ability to bind Zn^{2+}.[60] On the other hand, swapping the Cys and His ligands in the classical ZF peptide would yield nucleation structures vastly different from those depicted in Figure 7, as Zn^{2+} would be predicted to bind first to the C-terminal solvent-exposed Cys residues and then the N-terminal His ligands. This would therefore lead to ZF misfolding, in accord with the experimental finding that the HHCC ZF peptide leads to a misfolded structure.[25] Using a metal ion that prefers aa residues or a coordination geometry different from Zn^{2+} to induce ZF folding would also yield nucleation structures that do not resemble those depicted in Figure 7, and would lead to ZF misfolding as well.

In summary, ZF peptides apparently follow a nucleation folding pathway where Zn^{2+} acts as a nucleus around which the peptide folds. The Zn^{2+} cation assists ZF peptide folding by gradually reducing the magnitude of accessible conformational space via successively coordinating to a set of peptide ligands (two Cys/backbone carbonyls and then two His), forming complexes of increasing stability with increasing native contacts. By successively replacing Zn−O(water) bonds with Zn−S(Cys^-)/O(backbone) bonds and Zn−O(backbone) bonds with Zn−N(His) bonds, each successive Zn complex accrues more and more favorable as well as native interactions, while the ZF becomes increasingly compact—this lowers both the energy and entropy of the ZF polypeptide, reducing the search volume, thus guiding folding to the native state. The results provide insights into why certain mutations lead to protein misfolding and suggest that backbone peptide groups play an important role in stabilizing the nucleation structures and promoting the hydrophobic core formation.

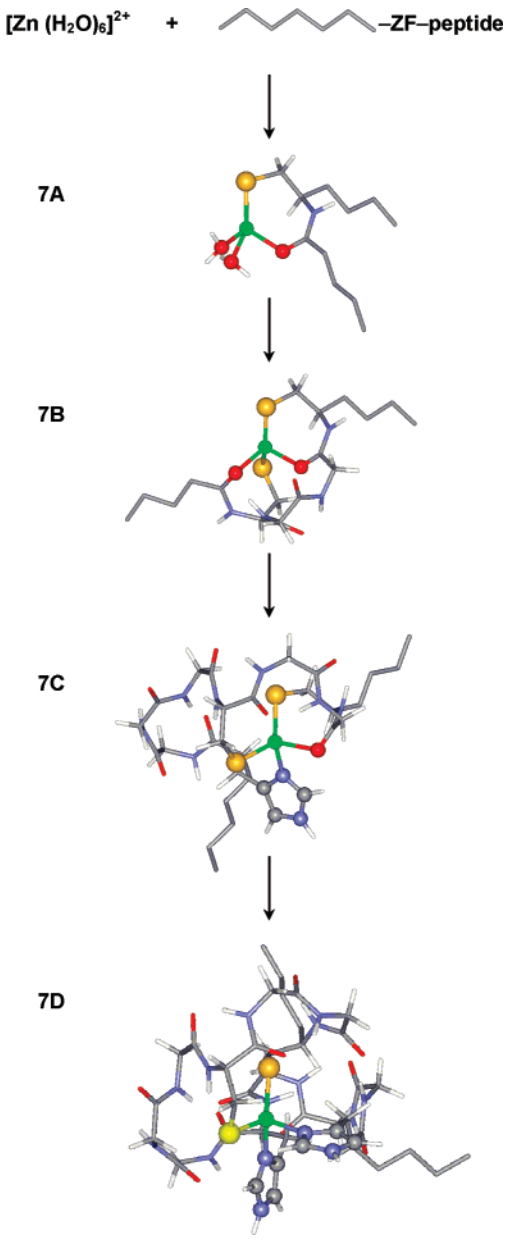


Figure 7. Schematic representation of nucleation complexes along the ZF folding process. The nucleation structures involving the Zn-bound ligands correspond to the fully optimized geometries shown in Figure 6, with the "zigzag chains" added to represent the rest of the unfolded or partially folded polypeptide chain.

Acknowledgment. Supported by the Institute of Biomedical Sciences, Academia Sinica, and the NSC (contract no. NSC 94-2113-M-001-018).

Supporting Information Available: Calibration of solvation free energy calculations, complete ref 35, free energy of deprotonation of zinc-bound protonated imidazole, ZF unfolding simulations, and atomic coordinates of the fully optimized Zn^{2+} complexes. This material is available free of charge via the Internet at http://pubs.acs.org.

JA073322C

(60) Nomura, A.; Sugiura, Y. *Inorg. Chem.* **2002**, *41*, 3693−3698.