Single-Chain versus Dimeric Protein Folding: Thermodynamic and Kinetic Consequences of Covalent Linkage

ARTICLE *in* JOURNAL OF THE AMERICAN CHEMICAL SOCIETY · AUGUST 2001

Impact Factor: 12.11 · DOI: 10.1021/ja015990i · Source: PubMed

1 AUTHOR:

Huan-Xiang Zhou
Florida State University
259 PUBLICATIONS **10,163** CITATIONS

SEE PROFILE

Single-Chain versus Dimeric Protein Folding: Thermodynamic and Kinetic Consequences of Covalent Linkage

Huan-Xiang Zhou

Department of Physics, Drexel University
Philadelphia, Pennsylvania 19104

Received April 10, 2001
Revised Manuscript Received June 4, 2001

The construction of a single-chain protein by linking the C terminal of one subunit with the N terminal of another in an otherwise dimeric protein has been considered as a strategy for increasing protein stability.[1,2] In this Communication we investigate the role of the covalent linkage in the stability of the folded protein and the implication for the folding kinetics.

The class of dimeric proteins under consideration have hydrophobic cores formed by side chains from both subunits, such that they become unfolded when the subunits dissociate. In addition, the covalent linkage is assumed to be flexible and not form specific contacts with the rest of the single-chain protein. In this class are a number of well-characterized proteins, including the gene V dimer of bacteriophage f1,[1] the Arc repressor of bacteriophage P22,[2] and the coiled-coil region of the yeast transcription factor GCN4.[3]

Our primary goal is to relate the folding stability of the dimeric protein to that of the single-chain version (see Figure 1). Let the two subunits be denoted as A and B. If the partition functions of subunit α in the unfolded and folded states are u_α and f_α, respectively, then the equilibrium constant for the dimeric protein is:[4]

$$K^{\mathrm{d}} = f_{\mathrm{A}} f_{\mathrm{B}} f_{\mathrm{int}} / u_{\mathrm{A}} u_{\mathrm{B}} \tag{1}$$

where f_{int} arises from the interaction between the two subunits in the folded state. In the simplest case where the two subunits are modeled as spheres interacting via a central symmetric potential $U(r)$, one has $f_{\mathrm{int}} = \int \exp[-U(r)/k_{\mathrm{B}}T] 4\pi r^2 \, \mathrm{d}r$,[4] where k_{B} is Boltzmann's constant and T is the absolute temperature. Note that in defining f_α one assumes that the subunit has the structure in the folded state (which by itself, i.e., without the other subunit, may not be stable).

To derive a simple expression for the equilibrium constant for the single-chain version, we assume that the linkage is structureless and modeled as a polymer chain. In the unfolded state the end-to-end distance d of the linkage is able to sample all possible values. On the other hand, in the folded state d is restricted to small fluctuations around a fixed value d_0. Let the probability density for the end-to-end vector be $p(d_0)$ when $d = d_0$, then the equilibrium constant for the single-chain protein is:

$$K^{\mathrm{s}} = f_{\mathrm{A}} f_{\mathrm{B}} f_{\mathrm{int}} p(d_0) / u_{\mathrm{A}} u_{\mathrm{B}} \tag{2}$$

In writing eq 2 we assume that (1) for the unfolded chain the partition function is the product of three factors, u_{A}, u_{B}, and a corresponding quantity for the linkage and (2) in the folded state the linkage behaves the same way as it does in the unfolded state,

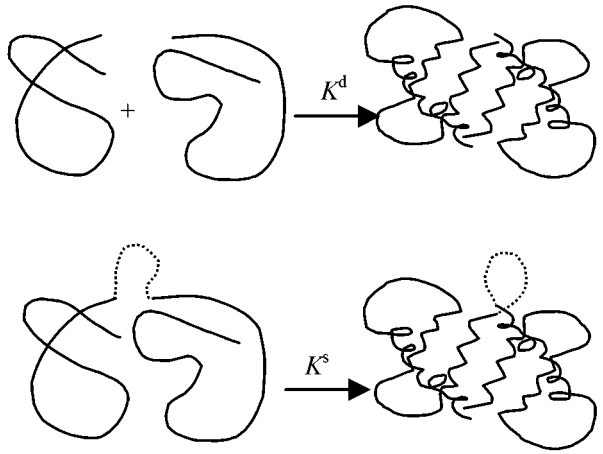

Figure 1. The folding of the dimeric protein (with equilibrium constant K^{d}) and the single-chain version (with equilibrium constant K^{s}). The dotted curve represents the covalent linkage.

except that its end-to-end distance is fixed at d_0. In particular the latter assumption means that the interaction between the linkage and the rest of the protein is negligible. Comparison of eqs 1 and 2 leads to:

$$K^{\mathrm{s}}/K^{\mathrm{d}} = p(d_0) \tag{3}$$

Recently Zhou[5] found that unstructured loops in proteins can be modeled very well by the worm-like chain[6] with a persistence length $l_{\mathrm{p}} = 3$ Å. For this polymer model one has[7]

$$p(d) = (3/4\pi l_{\mathrm{p}} l_{\mathrm{c}})^{3/2} \exp(-3d^2/4l_{\mathrm{p}}l_{\mathrm{c}})(1 - 5l_{\mathrm{p}}/4l_{\mathrm{c}} + 2d^2/l_{\mathrm{c}}^2 - 33d^4/80l_{\mathrm{p}}l_{\mathrm{c}}^3 - 79l_{\mathrm{p}}^2/160l_{\mathrm{c}}^2 - 329d^2l_{\mathrm{p}}/120l_{\mathrm{c}}^3 + 6799d^4/1600l_{\mathrm{c}}^4 - 344ld^6/2800l_{\mathrm{p}}l_{\mathrm{c}}^5 + 1089d^8/12800l_{\mathrm{p}}^2l_{\mathrm{c}}^6) \tag{4}$$

where $l_{\mathrm{c}} = Lb$, with L the number of residues forming the linkage, and $b = 3.8$ Å, the nearest $C_\alpha - C_\alpha$ distance.

The $K^{\mathrm{s}}/K^{\mathrm{d}}$ ratio calculated from eqs 3 and 4 for the Arc repressor, the gene V protein, and GCN-p1 is shown in Table 1. These results agree with the experimental values to within a factor of 2. Considering the fact that the calculated results do not involve any adjustable parameters and the values of $K^{\mathrm{s}}/K^{\mathrm{d}}$ span 2 orders of magnitude, the agreement is very satisfactory.

We also examined the prediction of eq 3 on two cases in which some of the requirements for the use of eq 3 are not met. *Streptomyces* subtilisin inhibitor (SSI; pdb code 2sic) is folded only as a homodimer. Tamura and Privalov[8] constructed a single-chain version by mutating Asp83, located within the dimer interface, to Cys and cross-linking with a disulfide bond. The link is not between the C terminal of one subunit and the N terminal of the second and the Asp to Cys mutation itself may affect the stability of the folded state. Nonetheless, the fact remains that the $C_\alpha - C_\alpha$ distance between the two Cys83 residues is restricted to around $d_0 = 6.6$ Å in the folded state and will have a much wider range in the folded state. Using the distribution of the $C_\alpha - C_\alpha$ distances of disulfide bonded Cys in proteins as a model for $p(d)$, we estimated $p(d_0) \approx 1.7$ M. Tamura and Privalov

(1) Liang H.; Sandberg, W. S.; Terwillinger, T. C. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 7010−7014.
(2) Robinson, C. R.; Sauer, R. T. *Biochemistry* **1996**, *35*, 13878−13844.
(3) Moran, L. B.; Schneider, J. P.; Kentsis, A.; Reddy, G. A.; Sosnick, T. R. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 10699−10704.
(4) (a) Hill, T. L. *An Introduction to Statistical Thermodynamics*; Dover: New York, 1986; pp 177−180. (b) Shoup, D.; Szabo, A. *Biophys. J.* **1982**, *40*, 33−39. (c) Zhou, H.-X. *Biopolymers*. In press.
(5) Zhou, H.-X. *J. Phys. Chem. B.* In press.
(6) Kratky, O.; Porod, G. *Recl. Trav. Chim.* **1949**, *68*, 1106−1122.
(7) (a) Gobush, W.; Yamawaka, H.; Stockmayer, W. H.; Magee, W. S. *J. Chem. Phys.* **1972**, *57*, 2839−2843. (b) Yamawaka, H.; Stockmayer, W. H. *J. Chem. Phys.* **1972**, *57*, 2843−2854.
(8) Tamura, A.; Privalov, P. L. *J. Mol. Biol.* **1997**, *273*, 1048−1060.

10.1021/ja015990i CCC: $20.00
Published on Web 06/14/2001

Table 1. Experimental and Calculated Ratio of K^s and K^d

protein[a]	condition	K^d (M^{-1})	K^s	linkage[b]	K^s/K^d (mM)	$p(d_0)$ (mM)[c]
Arc repressor (1myk)	4.19 M urea; $T = 298$ K	10^3	2.44	A50−B7; $L = 25$; $d_0 = 29.9$ Å	2.44	3.99
Gene V protein (1gvp)	2.6 M Gdn·HCl; $T = 298$ K	10^5	1.51×10^4	A86−B1; $L = 8$; $d_0 = 12.4$ Å	151	74
GCN4-p1 (2zta)	4 M Gdn·HCl; $T = 283$ K	353	85	A2−B2; $L = 9$; $d_0 = 6.3$ Å	241	129

[a] Entry in parentheses is the PDB code. [b] A*i*−B*j* refers to the two end residues of the covalent linkage (e.g., residue 50 of subunit A). These were typically selected as the last unstructured residue of subunit A and the first unstructured residue of subunit B. Additional linking residues were introduced in the experiments, giving rise to the listed total residue L for the covalent linkage. The end-to-end distance d_0 was calculated between residues A*i* and B*j* from the PDB structure. [c] $p(d)$ has the unit of density (i.e., number of ends per unit volume). It is converted to the unit of molarity by dividing by the Avogadro number.

found $K^d = 2.4 \times 10^4$ M^{-1} and $K^s = 4.3 \times 10^3$ at $T = 353$ K, thus $K^s/K^d = 18$ M. While the value of $p(d_0)$ estimated for the present case is about an order of magnitude higher than those for the gene V protein and GCN4-p1, it is still an order of magnitude lower than the experimental value of K^s/K^d. The additional stabilization of ∼1.6 kcal/mol may be attributed in part to the elimination of the repulsion between the two negatively charged Asp83 residues in the disulfide bonded mutant.

In the second case we compare the folding of the intact chymotrypsin inhibitor 2 (CI2) and the association/folding of a dimeric version in which the peptide bond between Met40 and Glu41 is broken. A main difference with the proteins listed in Table 1 is that the linkage in CI2, taken to be the reactive site loop consisting of residues 35 to 44, has extensive hydrophobic and electrostatic interactions with the rest of the protein.[9] The equilibrium constant for the dimeric version at $T = 298$ K and without denaturant was measured by Ladurner at al.[9c] to be $K^d = 2.44 \times 10^7$ M^{-1} and the equilibrium constant for intact CI2 is $K^s = 3.74 \times 10^5$ when extrapolated to the same condition.[9a] One thus has $K^s/K^d = 15$ mM. The end-to-end distance d for the linkage (from PDB entry 2ci2) is $d_0 = 20.7$ Å, eq 4 thus predicts $K^s/K^d = 12$ mM. The close agreement with the experimental result is perhaps coincidental, but it does serve to demonstrate the importance of accounting for the fact that d samples a wide range of values in the unfolded state of the single-chain protein but is restricted to d_0 in the folded state.

For the proteins listed in Table 1, the specific interactions maintaining the folded state are identical in the dimeric and single-chain versions. The two respective unfolding rate constants k_u^d and k_u^s, dictated by the breaking of some of these specific interactions, are thus expected to be not too different. This is indeed the case. Specifically, for the Arc repressor $k_u^d = 18$ s^{-1} and $k_u^s = 25$ s^{-1} at 7 M urea, for the gene V protein $k_u^d = 3.5 \times 10^{-3}$ s^{-1} and $k_u^s = 2.5 \times 10^{-3}$ s^{-1} at 5 M Gdn·HCl, and for GCN4-p1 $k_u^d = 1.9$ s^{-1} and $k_u^s = 0.4$ s^{-1} at 4 M Gdn·HCl. A hallmark for the covalent linkage serving just as a tether (as opposed to an active participant in stabilizing the folded structure) should be the tracking of k_u^d by k_u^s. More recent variants of the covalent linkage for the Arc repressor designed by Robinson and Sauer[10] led to substantial slowing down of the unfolding process (with $k_u^s = 0.7$ s^{-1}, i.e., one 26th of k_u^d, at 7 M urea). This slowing down indicates an active role for the covalent linkage (and consequently the breakdown of eq 3).

As a consequence of the result $k_u^d \approx k_u^s$, the folding rate constants k_f^d and k_f^s of the dimeric and single-chain proteins can be related via $k_f^s/k_f^d = p(d_0)$. This allows us to estimate the limit of the folding rate of a single-chain protein. The folding rate constant of the dimeric protein is limited by the diffusional approach of the two subunits. When the folded dimer is stereospecific and long-range electrostatic interactions are absent, the diffusion-limited rate constant is $\sim 10^6$ M^{-1} s^{-1}.[11] If $p(d_0)$ is of the order of magnitude of 0.1 M (see Table 1), then $k_f^s \leq 10^5$ s^{-1}. This estimate complements one previously proposed by Eaton et al.[12] Recent experimental results on the folding of two small proteins approach this limit.[13]

In summary, we explored the connection between the stability of dimeric proteins and that of their single-chain versions. We restricted ourselves to the simplest case where the covalent linkage acts merely as a tether. In this case we can relate the ratio of the two respective equilibrium constants to the probability density for the end-to-end distance of the covalent linkage to be at the value in the folded structure. We further argue that the unfolding rates of the dimer and the single-chain protein should be similar. On the basis of these results we estimate a rate limit of 10^5 s^{-1} for the folding of single-chain proteins.

Acknowledgment. This work was supported in part by the National Institutes of Health (GM-58187).

JA015990I

(9) (a) Itzaki, L. S.; Otzen, D. E.; Fersht, A. R. *J. Mol. Biol.* **1995**, *254*, 260−288. (b) Daggett, V.; Li, A.; Itzhaki, L. S.; Otzen, D. E.; Fersht, A. R. *J. Mol. Biol.* **1996**, *257*, 430−440. (c) Ladurner, A. G.; Itzhaki, L. S.; de Prat Gay, G.; Fersht, A. R. *J. Mol. Biol.* **1997**, *273*, 317−329.

(10) Robinson, C. R.; Sauer, R. T. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 5929−5934.

(11) Zhou, H.-X. *Biophys. J.* **1997**, *73*, 2441−2445.

(12) (a) Hagen, S. J.; Hofrichter, J.; Szabo, A.; Eaton, W. A. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 11615−11617. (b) Eaton, W. A.; Munoz, V.; Hagen, S. J.; Jas, G. S.; Lapidus, L. J.; Henry, E. R.; Hofrichter, J. *Annu. Rev. Biophys. Biomol. Struct.* **2000**, *29*, 327−359.

(13) (a) Burton, R. E.; Huang, G. S.; Daughety, M. A.; Calderone, T. L.; Oas, T. G. *Nat. Struct. Biol.* **1997**, *4*, 305−310. (b) Wittung-Stafshede, P.; Lee, J. C.; Winkler, J. R.; Gray, H. B. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 6587−6580.