Modified self-consistent harmonic approach to thermal fluctuational disruption of disulfide bonds in proteins

Y. Z. Chen
Department of Computational Science, National University of Singapore, Lower Kent Ridge Road, Singapore 119260
(Received 21 June 1999; revised manuscript received 9 August 1999)

Disulfide bonds (covalent bonds between two sulfur atoms on separate amino acid cysteines) play an important role in the folding stability of many proteins. The determination of the dynamic stability of disulfide bonds can therefore facilitate an understanding of the mechanism of protein stability. We have examined whether a modified self-consistent harmonic approach, which has been successful in modeling hydrogen bond breaking dynamics in DNA and covalent bond disruption in drug-DNA systems, is useful in computing thermal fluctuational disruption probability of disulfide bonds in proteins. Our results on a number of protein x-ray crystal structures showed that the computed disruption probabilities are consistent with observations. The free energies derived from computed probabilities are comparable to the observed values from protein engineering experiments. The method presented here has a potential application in analyzing disulfide bond dynamics in a variety of biological processes. [S1063-651X(99)13611-0]

PACS number(s): 87.15.By, 87.10.+e, 63.20.Dj, 63.70.+h

I. INTRODUCTION

Disulfide bonds are covalent bonds between two sulfur atoms on separate amino acid cysteines in a number of proteins. Figure 1 illustrates a disulfide bond in a protein called BPTI. These bonds are believed to play a stabilizing role in globular proteins (spherelike proteins) that bind or act on external coat of cells [1,2]. Deletion of disulfide bonds from these proteins results in unfolding (a transition from a folded globular structure to an unfolded random coil conformation). Reintroduction of these disulfide bonds leads to the refolding of these proteins back to their original globular structure. These findings prompted significant interest in exploring the introduction of disulfide bonds to enhance protein stability [2]. Moreover, disulfide bonds have been found to be involved in the initial stages of biosynthesis of some proteins inside cells [3]. Therefore, these bonds have been considered to be an excellent experimental probe of the intermediate conformations in the folding and unfolding processes of these proteins [4].

Due to this great interest, much effort has been devoted to the study of various aspects of disulfide bonds in protein folding stability. Early works showed that disulfide bonds can reduce the entropy of the unfolded state and thus increase the relative stability of the folded state [5]. The computed entropy changes were found to be comparable to the observed total free energies [5]. More recently it has become clear that the effect of the disulfide bond is not only on entropic changes in the unfolded state. Several studies have shown that disulfide bond induced torsion strain (protein chain distortion induced by rotation of chemical bonds) is significant [2,6]. This strain contributes a negative free energy change with a magnitude comparable to the total free energy change for the removal of disulfide bonds. A thermodynamic analysis indicated that, to a large extent, disulfide bonds stabilize a protein enthalpically in the folded state [7]. There are also examples that show that the introduction of disulfide bonds does not always stabilize protein [8].

These experimental and theoretical findings indicated that the effect of disulfide bonds is complex, and likely involves many factors of comparable energy including disulfide bond covalent bonding between two amino acids, disulfide bond

FIG. 1. A disulfide bond in the protein BPTI. The main chain of this protein is represented by a ribbon, and the side chain of amino acids is represented by a line drawing. The disulfide bond is indicated by an arrow. This bond cross-links two ends of a folded protein chain, and plays an important role in stablizing the fold. There are two additional disulfide bonds in this protein which are not displayed explicitly.

induced torsion strain in protein chain, and entropic effects. Therefore, a study of disulfide bond breaking dynamics (events prior to bond disruption induced conformation change) is of significance in probing one of the factors involved in the overall effect of disulfide bonds in protein.

The method used in this study is a modified self-consistent harmonic approach (MSHA) which was developed for the study of thermal fluctuational hydrogen bond breaking in DNA in the premelting [9] and melting temperature regimes [10]. This method was later extended to the study of thermal fluctuational covalent bond breaking between a drug and a DNA [11]. In this work, the parameters for disulfide bonds were derived by means of the same formalism as that for covalent bonds between an anticancer drug cisplatin and DNA [11]. These parameters were derived from general data from molecular orbital calculations, neutron diffraction study, and other existing parameters widely used in molecular dynamics simulation of biomolecules. The usefulness of our method and the algorithm of these parameters in modeling disulfide bonds in proteins were tested by comparison of computed thermal fluctuational disulfide bond disruption probabilities with experimentally estimated disulfide bond breaking free energy.

II. THEORETICAL METHODS

A. Protein structures, potential functions, and parameters

The proteins considered in the present work are BPTI (Protein DataBank Id: 4PTI), ribonuclease T1 (9RNT), α-lactalbumin (1ALC), and ribonuclease A (3RN3). The xyz coordinates of these proteins are from the Protein DataBank (http://www.resb.org/pdb/). No further conformation optimization was made on these structures.

The internal motions of each protein are described by the Hamiltonian

$$
\begin{aligned}
H=&\sum_{\text{atoms}}\frac{P^2}{2m}+\sum_{\text{bond-stretch}}\frac{1}{2}K_r(r-r_{\text{eq}})^2\\
&+\sum_{\text{bond-angle-bending}}\frac{1}{2}K_\theta(\theta-\theta_{\text{eq}})^2\\
&+\sum_{\text{bond-rotation}}\frac{V_n}{2}[1+\cos(n\phi-\gamma)]\\
&+\sum_{\text{S-bond}}[V_0(1-e^{-a(r-r_0')})^2-V_0]\\
&+\sum_{\text{H-bond}}[V_0(1-e^{-a(r-r_0')})^2-V_0]\\
&+\sum_{\text{Nonbonded}}\left[\frac{A_{ij}}{r_{ij}^{12}}-\frac{B_{ij}}{r_{ij}^{6}}+\frac{q_iq_j}{\epsilon_{ij}r_{ij}}\right],
\end{aligned}
\tag{1}
$$

where bond stretch terms are for bond vibration motions, bond angle bending terms are for motions that change the angle between two neighboring bonds, bond rotation terms are for the rotation of bonds around their neighboring bonds, and S- and H-bond terms are for disulfide and hydrogen bond interactions, respectively. K_r and r_{eq} are the covalent bond (other than disulfide bond) force constant and equilibrium bond length; K_θ and θ_{eq} are bond angle bending force constant and equilibrium bond angle; V_n, n, and γ are bond rotation (torsion) parameters; A_{ij} and B_{ij} are nonbonded van der Waals potential parameters between ith and jth atom; q_i is the partial charge of ith atom and ϵ_{ij} is the dielectric constant between ith and jth atom; and V_0, a, and r_0' are parameters for the More potential describing either H or S bonds. The variable r is the atom-atom distance (hydrogen donor and acceptor distance for H bonds), θ is the bond angle (the angle between two bonds connected by a common atom), and ϕ is the torsion angle (the angle between the two planes formed by each pair of bonds linearly connected to each other).

Except for the S- and H-bond terms, the potential terms and parameters in Eq. (1) are the same as that of a widely used software package for molecular dynamics simulations of proteins, DNAs, RNAs, and other organic molecules [12]. This software package is named AMBER, which stands for Assisted Model Building with Energy Refinement. We used a Morse potential to empirically describe disulfide bond and hydrogen bond interactions. The Morse potential was introduced to empirically model hydrogen bonds in organic molecules [13], and has been shown to describe hydrogen bond disruption dynamics of DNA fairly well [9,10]. The Morse potential was proposed to empirically model covalent bonds [14,15], and has been successfully used to model the dynamics of covalent bond breaking of anticancer drug cisplatin from DNA [11]. The parameters for both S and H bonds in proteins were derived from the same respective formalism for covalent bonds and H bonds in the modeling of cisplatin-DNA complexes [11].

For S bonds, the potential depth $V_0=14$ kcal/mol is from self-consistent molecular orbital calculations [16]. The parameter $a=3.44$ Å^{-1} is determined by matching the calculated second derivative of the potential at potential minimum to the AMBER force constant [12]. r_0' is the potential minimum for the strained S bond. Disulfide bonds in proteins are strained due to strong cross bond forces, and this is evidenced by the observed variation of bond lengths (from 1.99 to 2.09 Å). r_0' can be divided into two terms $r_0'=r_0+\delta r$, where r_0 is the potential minimum for an unstrained S bond and δr is the strain induced by cross bond stress. $r_0=2.038$ Å is from a neutron diffraction study on L cystine [17,18]. The value of δr is determined by a variety of interactions across the bond. It can be determined empirically by equating the calculated thermal average bond length $\langle r\rangle$ with the observed length in the x-ray crystal structure. In this way $\delta r=r_x-r_0-dr$, where r_x is the bond length in the x-ray crystal structure and dr is the thermal expansion determined from the zero average force condition $\langle V'\rangle=0$. Notice that effects associated with forces other than the S bond are shifted into δr. These forces are relatively unchanged over the range of displacement for S-bond disruption, and only the S-bond potential need to be included in the zero force calculation.

Water molecules found in the x-ray crystal structures are included in our calculations. Moreover, distance dependent dielectric constants are used to mimic the solvent polarization effect.

TABLE I. Comparison between calculated disulfide bond disruption free energy ΔG_{theor} with the observed free energy ΔG_{expt} for the disulfide bond in BPTI [6], ribonuclease T1 [5], α-lactalbumin [20], and ribonuclease A [21]. ΔG_{expt} for the CYS6-CYS103 bond in ribonuclease T1 is deduced from the observed value for all the two disulfides in the protein. ΔG_{expt} for ribonuclease A is the per bond average from the observed value of 19 kcal/mol for all four disulfide bonds. The calculated mean bond length $\langle r \rangle$ and disruption probability P are also included.

Protein	Disulfide bond	$\langle r \rangle$ Å	P	ΔG_{theor} kcal/mol	ΔG_{expt} kcal/mol
BPTI	SG CYS5–SG CYS55	2.050	3.347E-4	4.665	
	SG CYS14–SG CYS38	2.088	3.330E-4	4.669	5
	SG CYS30–SG CYS51	2.024	3.350E-4	4.665	
α-lactalbumin	SG CYS6 - SG CYS120	2.039	3.332E-4	4.668	3.12
	SG CYS28–SG CYS111	2.085	3.316E-4	4.671	
	SG CYS61–SG CYS77	2.063	3.306E-4	4.672	
	SG CYS73–SG CYS91	2.024	3.289E-4	4.675	
ribonuclease T1	SG CYS2–SG CYS10	2.034	3.312E-4	4.671	3.4
	SG CYS6–SG CYS103	1.989	3.290E-4	4.675	3.8-5.9
ribonuclease A	SG CYS26–SG CYS84	2.062	3.355E-4	4.664	4.75
	SG CYS40–SG CYS95	2.089	3.341E-4	4.666	4.75
	SG CYS58–SG CYS110	2.024	3.345E-4	4.665	4.75
	SG CYS65–SG CYS72	2.066	3.354E-4	4.664	4.75

B. Statistical algorithm for disulfide bond disruption probability

The algorithm for disulfide bond disruption is the same as that used in modeling cisplatin-DNA covalent bond disruption [11]. This approach is based on the Bogoliubov variational theorem which states that the free energy F of a system can be approximated by the solutions of an effective Hamiltonian [19] From the Bogoliubov inequality

$$F \leqslant F_0 + \langle H - H_0 \rangle, \tag{2}$$

one can self-consistently adjust the parameters of the trial Hamiltonian H_0 with respect to the true Hamiltonian H to find a trial system that minimizes the left-hand side terms and thus best approaches the true free energy. Here F_0 is the free energy of the trial Hamiltonian system. The following harmonic Hamiltonian H_0 is used as the trial Hamiltonian:

$$\begin{aligned} H_0 = & \sum_{\text{atoms}} \frac{P^2}{2m} + \sum_{\text{bond-stretch}} \frac{1}{2} K_r (r - r_{\text{eq}})^2 \\ & + \sum_{\text{bond-angle-bending}} \frac{1}{2} K_\theta (\theta - \theta_{\text{eq}})^2 \\ & + \sum_{\text{bond-rotation}} \frac{1}{2} K_\Phi (\Phi - \Phi_{\text{eq}})^2 + \sum_{\text{S-bonds}} \frac{1}{2} K_S (r - \langle r \rangle)^2 \\ & + \sum_{\text{H-bonds}} \frac{1}{2} K_H (r - \langle r \rangle)^2 + V_{\text{st}}, \end{aligned} \tag{3}$$

where V_{st} is the potential value at equilibrium positions. Since changes in bond rotational angles are generally small for motions prior to the disruption of a disulfide bond, the force constant K_Φ for bond rotation can be given by the second derivative of the relevant potential. The algorithm of the H-bond force constant K_H was given in an earlier publication [11].

The force constant of a s bond K_S is determined by minimization of free energy expansion in Eq. (2), which gives

$$K_S = (1 - P) \frac{\int_{r_c}^{\infty} dr \frac{d^2 V(r)}{dr^2} e^{-(r - \langle r \rangle)^2 / 2 \langle u^2 \rangle}}{\int_{r_c}^{\infty} dr e^{-(r - \langle r \rangle)^2 / 2 \langle u^2 \rangle}}, \tag{4}$$

where r_c is the inner-bound cutoff determined from $V(r_c) = 2V_0$. Our analysis indicates that the calculations are not sensitive to the exact choice of the cutoff [$V(r_c) = 0$ and $V(r_c) = 4V_0$ give similar results]. The scaling factor $(1 - P)$ is introduced to take into account the effect of disrupted bonds in the statistical ensemble and P is the disruption probability of the bond. $\langle u^2 \rangle$ is the mean square vibrational amplitude of the bond given by

$$\langle u^2 \rangle = \sum_l s_l^2, \tag{5}$$

where ω_l and l are the frequency and the index of the normal modes of the harmonic system in Eq. (3), respectively. T is the temperature, k_B is Boltzmann's constant, and $\hbar$ is Planck's constant divided by 2π.

The self-consistent Bogoliubov approach gives rise to statistical probability distribution functions of finding a S bond with a particular length. From these distribution functions one can determine the probability of finding a S bond fluctuating beyond a certain break-away point, i.e., the disruption probability of individual S bond. This probability is given by

TABLE II. Calculated bond disruption free energy ΔG_{theor} for the disulfide bonds in several proteins using different Morse potential parameters: (a) $V_0=14$ kcal/mol, $a=3.44$ Å^{-1}, and $r_0=2.038$ Å. (b) $V_0=14$ kcal/mol, $a=3.70$ Å^{-1}, and $r_0=2.038$ Å. (c) $V_0=17$ kcal/mol, $a=3.44$ Å^{-1}, and $r_0=2.038$ Å.

Protein	Disulfide bond	ΔG_{theor} (kcal/mol)		
		(a)	(b)	(c)
BPTI	SG CYS5–SG CYS55	4.665	5.078	5.298
	SG CYS14–SG CYS38	4.669	5.079	5.300
	SG CYS30–SG CYS51	4.665	5.077	5.298
α-lactalbumin	SG CYS6–SG CYS120	4.668	5.070	5.289
	SG CYS28–SG CYS111	4.671	5.074	5.291
	SG CYS61–SG CYS77	4.672	5.076	5.293
	SG CYS73–SG CYS91	4.675	5.079	5.297
ribonuclease T1	SG CYS2–SG CYS10	4.671	5.073	5.292
	SG CYS6–SG CYS103	4.675	5.079	5.296
ribonuclease A	SG CYS26–SG CYS84	4.664	5.040	5.283
	SG CYS40–SG CYS95	4.666	4.984	5.286
	SG CYS58–SG CYS110	4.665	5.009	5.285
	SG CYS65–SG CYS72	4.664	5.041	5.283

$$P=\int_{L_{\max}}^{\infty} dr e^{-(r-\langle r\rangle)^2/2\langle u^2\rangle}, \tag{6}$$

where $L_{\max}$ is the maximum stretch length (breakaway point) of a S bond. It is determined as the inflection point of the potential function [where $V''=0$, which gives $L_{\max}=r_0'+(1/a)\ln 2$].

Given the S-bond disruption probability, the free energy change ΔG_{theor} associated with the disruption of a covalent bond can be deduced from the van't Hoff relation

$$\Delta G_{\text{theor}}=-RT\ln P. \tag{7}$$

In this work, the temperature T was taken as room temperature 293 K.

C. Computation procedure

The computation procedure is as follows: Starting from an initial set of force constants and given the structure of a protein, the equations of motions derived from the Hamiltonian in Eq. (3) is numerically solved to determine the normal modes. The calculated normal modes are then used to calculate $\langle u^2\rangle$ and P from Eqs. (5) and (6). These are then used to calculate the parameter r' using the method given in Sec. II A, and a set of force constants from Eq. (4). These calculated force constants are then used to restart another round of calculation. Such a process continues until every output force constant matches the input force constant, judged by the condition $\Delta K_S/K_S<0.01$. The self-consistently determined P's are then used to calculate ΔG_{theor} from Eq. (7). Our computations were carried out on a DEC α workstation. The diagonalization of the dynamic matrix from Eq. (3) was performed by the LAPACK routine DSYEV.

III. RESULTS AND DISCUSSIONS

The calculated s-bond disruption free energy ΔG_{theor} for the proteins BPTI, ribonuclease T1, α-lactalbumin, and ribonuclease A are given in Table I along with the calculated mean bond lengths and disruption probabilities. For comparison the observed free energy ΔG_{expt} for the removal of each of these disulfide bonds in these proteins are also included in Table I. In the relevant experiments, individul disulfide bonds were removed without substantial change of the overall protein conformation. Thus the observed free energy roughly measures that of the bond disruption before a considerable conformation change. It can thus be compared with our results.

We found that all disulfide bonds have very similar ΔG_{theor} near $\sim$4.66 kcal/mol. The narrow range of ΔG_{theor} results from the small variation in bond length (less than 0.1 Å) in the proteins studied. The length of a disulfide bond is not the only determinant of ΔG_{theor}. For instance, in ribonuclease T1, ΔG_{theor} of the CYS6-CYS103 bond with a length of 1.989 Å is smaller than that of the CYS2-CYS10 bond with a length of 2.034 Å. From Eqs. (6) and (7), we find that ΔG_{theor} depends on the vibrational mean square amplitude of the bond motion as well as on bond length. It is the synergistic effects of cross bond static forces and the amplitude of vibrational motion of the bond end atoms that determines the calculated ΔG_{theor}'s.

From Table I we find that our calculated ΔG_{theor} for several disulfide bonds is comparable to the ΔG_{expt} estimated from protein engineering experiments. These include the CYS14-CYS38 bond in BPTI [6], the CYS6-CYS120 bond in α-lactalbumin [20], and the CYS2-CYS10 bond in ribonuclease T1 [5]. In addition, the ΔG_{theor} for the CYS6 - CYS103 bond in ribonuclease T1 and that for the four disulfide bonds in ribonuclease A are also comparable with the average value of ΔG_{expt} [5,21].

In our study we used a Morse potential to describe a disulfide bond. The Morse potential was suggested as a potential for a covalent bond [14] as well as for a H bond [13]. In this work, the potential parameters were determined by a simple scheme based on data from molecular orbital calculations and existing parameters widely used in molecular dynamics simulation of biomolecules. The statistical mechanical algorithm for computing the bond disruption proability was from a well established method based on the Bogoliubov variational theorem. No attempt was made to fit the parameters to reproduce experimental data. Our results for disulfide bonds in proteins were in fair agreement with experiments. This, coupled with the fair agreement between computation and observations for drug-DNA covalent bonds, indicates the usefulness of the MSHA in modeling the dynamics of covalent bond disruption in biological systems. Our analysis indicated that the calculated ΔG_{theor}'s are not sensitive to a small variation of the parameters as shown in Table II. Further refinement of the parameters can give more accurate ΔG_{theor}'s.

IV. CONCLUSIONS

A modified self-consistent harmonic approach has been shown to be useful in deriving thermal fluctuation bond disruption of disulfide bonds in proteins. Results on several proteins are in fair agreement with observations. The method presented here has potential application in analyzing disulfide bond disruption probability in unfolded, intermediate, and folded states to facilitate an analysis of the effect of disulfide bond formation and disruption on protein stability in biological processes. In order to probe the effect of disulfide bond breaking (particularly entropic effect) fully, a method for the analysis of conformational change following disulfide bond disruption needs to be developed. A combination of the MSHA with Monte Carlo or with molecular dynamics may be useful in this regard.

[1] J. M. Thornton, J. Mol. Biol. **151**, 261 (1981).
[2] R. Wetzel, Trends Biochem. Sci. **12**, 478 (1987).
[3] R. B. Freedman, Curr. Opin. Struct. Biol. **5**, 85 (1995).
[4] D. P. Goldenberg, Biochemistry **27**, 2481 (1988).
[5] C. N. Pace, G. R. Grimsley, J. A. Thomson, and B. J. Barnett, J. Biol. Chem. **263**, 11 820 (1988).
[6] D. P. Goldenberg, L. S. Bekeart, D. A. Laheru, and J. D. Zhou, Biochemistry **32**, 2835 (1993).
[7] A. J. Doig and D. H. Williams, J. Mol. Biol. **217**, 389 (1991).
[8] D. P. Goldenberg and T. E. Creighton, J. Mol. Biol. **179**, 527 (1984).
[9] Y. Z. Chen, Y. Feng, and E. W. Prohofsky, Biopolymers **31**, 139 (1991).
[10] Y. Z. Chen and E. W. Prohofsky, Phys. Rev. E **49**, 873 (1994).
[11] Y. Z. Chen, Yong-Li Zhang, and E. W. Prohofsky, Phys. Rev. E **55**, 5843 (1997).
[12] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, Jr., D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman, J. Am. Chem. Soc. **117**, 5179 (1995).
[13] N. C. Baird, Int. J. Quantum Chem. Quantum Biol. Symp. **1**, 49 (1974).
[14] B. R. Gelin, in *Computer Simulation of Biomolecular Systems*, edited by W. F. van Gunsteren, P. K. Weiner, and A. J. Wilkinson (ESCOM, Leiden, 1993), Vol. 2, pp. 127–146.
[15] N. T. Yu, Methods Enzymol. **130**, 350 (1986).
[16] W. J. Pietro, M. M. Francl, W. J. Hehre, D. J. DeFrees, J. A. Pople, and J. S. Binkley, J. Am. Chem. Soc. **104**, 5039 (1982).
[17] D. D. Jones, I. Bernal, M. N. Frey, and T. F. Koetzle, Acta Crystallogr., Sect. B: Struct. Crystallogr. Cryst. Chem. **30**, 562 (1974).
[18] H. E. Van Wart, L. L. Shipman, and H. A. Scheraga, J. Phys. Chem. **79**, 1428 (1975).
[19] H. B. Callen, in *Thermodynamics and an Introduction to Thermostatics*, 2nd ed. (Wiley, New York, 1985), pp. 433–435.
[20] M. Ikeguchi, S. Sugai, M. Fujino, T. Sugawara, and K. Kuwajima, Biochemistry **31**, 12 695 (1992).
[21] Y. Konishi, T. Ooi, and H. A. Scheraga, Biochemistry **21**, 4741 (1982).