Biochemistry and Biophysics Reports 9 (2017) 295–301

Contents lists available at ScienceDirect

Biochemistry and Biophysics Reports

journal homepage: www.elsevier.com/locate/bbrep

Thermal stability of pepsin: A predictive thermodynamic model of a multi-domain protein

Ali Asghar Rastegari[a,*], Behnaz Buzari[b], Abdol-Khalegh Bordbar[b]

[a] Department of Molecular and Cell Biochemistry, Falavarjan Branch, Islamic Azad University, Isfahan 84517-31167, Iran
[b] Laboratory of Biophysical Chemistry, Department of Chemistry, University of Isfahan, Isfahan 81746-73441, Iran

ARTICLE INFO

Keywords:
Multimeric proteins
Thermal denaturation
Protein unfolding
Differential Scanning Calorimetric
Sequential transition

ABSTRACT

Pepsin is generally used in the preparation of $F(ab)_2$ fragments from antibodies. The antibodies that are one of the largest and fastest growing categories of bio- pharmaceutical candidates. Differential scanning calorimetric is principally suitable method to follow the energetics of a multi-domain, fragment to perform a more exhaustive description of the thermodynamics in an associating system. The thermodynamical models of analysis include the construction of a simultaneous fitting of a theoretical expression. The expression depending on the equilibrium unfolding data from multimeric proteins that have a two-state monomer. The aim of the present study is considering the DSC data in connection with pepsin going through reversible thermal denaturation. Afterwards, we calculate the homology modeling identification of pepsin in complex multi-domain families with varied domain architectures. In order to analyze the DSC data, the thermal denaturation of multimer proteins were considered, the "two independent two-state sequential transitions with domains dissociation model" was introduced by using of the effective ΔG concept. The reversible unfolding of the protein description was followed by the two-state transition quantities which is a slower irreversible process of aggregation. The protein unfolding is best described by two non-ideal transitions, suggesting the presence of unfolding intermediates. These evaluations are also applicable for high throughput investigation of protein stability.

1. Introduction

Pepsin is commonly used in the preparation of $F(ab)_2$ fragments from antibodies. Antibodies and antibody-derived molecules are one of the largest and fastest growing classes of bio- pharmaceutical candidates and products. Differential scanning calorimetry (DSC) is the most direct experimental technique to resolve the energies out of conformational transitions of biological macromolecules. The stability of multi-domain proteins is commonly investigated by DSC. One of the great advantages of DSC is that it can detect fine-tuned of interactions between the individual domains of a protein. Followed by measuring the temperature dependence of the partial heat capacity, a basic thermodynamic property, DSC gives immediate access to the thermodynamic mechanism that governs a conformational equilibrium. The way proteins work imposes constraints on their function. Knowing the sources of the protein stability is essential to recognize their structure and function. One of the method for quantifying the stability of a protein is to populate the native and unfolded states by physical and chemical means. Then, the transitions measured by DSC or fluorescence, and absorption spectroscopy were evaluated [1,2]. During the last two decades DSC has significantly contributed to the development of our current understanding of the energetics and thermodynamic properties of protein folding-unfolding transitions [3]. By scanning microcalorimetry, it was shown that thermal transition is connected purely to the denaturation of protein molecules in the crystal and it is not accompanied by the crystal disintegration into separate molecules [4].

At critical temperatures and higher, along with typical loosening of the protein globuleis is observed along with it oligomeric molecules undergo dissociation process into subunits. Subsequently these smaller components may associate with each other and oligomeric structures. This association may result in irreversible denaturation and deviant physicochemical behavior [5]. The denaturation unfolding process is strongly dependent on the heating rate. As it is expected, the unfolding process is kinetically controlled by the presence of an irreversible reaction. CD signal on heating of proteins, constructing ellipticity quantities at wavelength distinct is very applicable for following the denaturation process [6]. A large body of research on the thermodynamics of small monomeric and single domain proteins has indicated that the hydrophobic effect and loss of conformational entropy

Abbreviations: DSC, Differential Scanning Calorimetry; T_m, Thermal midpoint; ΔH^{cal}, Calorimetric enthalpies; ΔH^{VH}, Van't Hoff enthalpy

* Corresponding author.
E-mail addresses: aarastegari@gmail.com, rastegari@iaufala.ac.ir (A.A. Rastegari), bordbar@chem.ui.ac.ir (A.-K. Bordbar).

http://dx.doi.org/10.1016/j.bbrep.2017.01.005
Received 14 July 2016; Received in revised form 18 December 2016; Accepted 17 January 2017
Available online 25 January 2017

are the major determinants of stability in the native state [7]. The unfolding transitions of several small globular proteins are usually highly cooperative. They closely follow a two-state mechanism under equilibrium conditions [8,9]. However, in some cases, stable intermediates have been detected and partially characterized [10,11]. Much of the disagreement between theoretical and experimental results of thermal denaturation of proteins derives from the necessity of using some models to interpret the thermodynamic data for proteins [1]. In these cases that have no easy experimental methods, applying an empirical formula can be a rational. For overcoming to the above difficulties, the present study uses of the effective ΔG concept, the thermodynamics of a multi-domain protein which undergo thermal denaturation is studied as a function of protein concentration. ΔGe_{ff} is a valuable factor that provides an natural increase of the stability of multimeric proteins; one can determine the section of unfolded protein from ΔGe_{ff} of a multimeric protein as easily as doing so from the ΔG° of a monomeric protein [8]. In this regards, we try to discuss and introduce "the two independent two-state transitions with a domain dissociation model".

2. Methods

DSC measurements were carried out on a MicroCall "MC-2" Differential Scanning Calorimeter (Micro All Inc., Northampton, MA) with cell volumes 1.14 mL, at heating rates 1.5 °C min^{-1}. DSC scans were obtained in a temperature range from 283.15 to 373.15 K. During the measurements, the protein concentrations was 30 μM, and pH ranged from 1.0 to 4.0 [12]. The buffers used for acid denaturation of pepsin are at different pH values KCl/HCl (pH 0.8–1.4), Gly/HCl (pH 1.6–3.0), sodium acetate (pH 3.5–5.0) at a concentration of 20 mM [13]. Pepsin, lyophilized powder (≥2500 units mg^{-1} protein) as well as Aldrich. Degassing during the calorimetric measurements was prevented by additional constant pressure of 1 atm over the liquids in the cells. At first, the solvent was placed in both the sample and reference compartments. A DSC curve corresponding to solvent vs. solvent run was used as an instrumental baseline. The calorimetric data was corrected for calorimetric baseline (by subtracting solvent-solvent scan).

In this study we used the Swiss-Model template library (SMTL) [14,15] database to analyze the domain organization of proteins. The SCOP [16] and CATH [17] families corresponding to multi-domain proteins have been used to identify single domain homologues. A control dataset of non-redundant single domain proteins obtained from completely different families has also been formed. All the structures have been optimized at 298 K by using the optimizing tool available in FoldX (version 3.0) [18]. Then free energy computations were done.

3. Results and discussion

3.1. The thermodynamic models

We clearly pointed out that this manuscript is the continuation of previously published work, as described in the framework of this model. The previously published paper were also clearly identified (reference [12]) at the manuscript.

The framework used for fitting DSC data offers four models. They all of them use the Levenberg-Marquardt non-linear least-square method [19], there was a conflict with the number of parameters involved:

1) Two-state with zero ΔC_p (parameters: Thermal midpoint (T_m), Calorimetric enthalpies (ΔH^{cal})); 2) Non-two state with zero ΔC_p (parameters: T_m, ΔH^{cal}, Van't Hoff enthalpy (ΔH^{VH})); 3) Two-state with non-zero ΔC_p (parameters: T_m, ΔH^{cal}, ΔC_p, BL0, BL1) and 4) Dissociation with non-zero ΔC_p (parameters: T_m, ΔH^{cal}, ΔC_p, BL0, BL1, n- number of multimers).

With exception of parameter number (4), one or more transitions can be used to fit the models, each transition has a relevant parameter set, for example in the case of two overlying transitions can be used to fit to one or more transitions. In the case of multiple transitions, each transition has its own complete parameter set, e.g., if pattern 1 is used to fit two overlapping transitions there will be two independent parameters sets (T_{m1}, ΔH^{cal1}) and (T_{m2}, ΔH^{cal2}). These parameters specify the thermal midpoint (T_m) and heat change (ΔH^{cal}) for each transition. BL0 and BL1 parameters define the slope and intercept of the low-temperature baseline segment, they are not repeated and appears only once in each model. While all four models use calorimetric heat change, only non-two state with zero ΔC_p model has a van't Hoff heat change (ΔH^{VH}). The ΔH^{cal} is determined only by the area under transition peak, while the van't Hoff heat is determined only by the shape (ΔC_p^{max} at transition midpoint). The transition sharpness is associated with ΔH^{VH} largernes. The relationship between ΔH^{cal} and ΔH^{VH} can sometimes provide insights not accessible from ΔH^{cal} model alone. If a protein is composed of two identical domains, which unfold independently with the same T_m and ΔH^{cal}, then the ratio of H^{cal}/H^{VH}, will be 2, while it would be 1 if the protein had a single domain. If, on the other hand, the protein dimerized and dimer underwent only a single coupled transition then, the ratio will be 0.5, etc.

In the model 1, it is possible to define that overlapping transitions are either independent or sequential in nature, for example, if two architectural domains are interacting strongly, it is possible to assume that their transitions will be coupled in a sequential manner. However, the independent model might better describe two transitions that are thoroughly uncoupled from one another. In practice, this option is often not critical because the sequential and independent models lead essentially identical effects whenever the T_m's of two transitions are separated by a couple of degrees or more. The mathematical derivations for each pattern has been introduced in the previous article [19]. Generally is the aim, the objective is using the simplest model which produces a good framework for the data. Therefore, if data is described by two-state model using two transitions, it would be preferred over a two-state model using three transitions or a non-two state model having two transitions. In this study, a noble two independent two-state transitions with domains dissociation model is introduced, as the fifth fitting model.

3.2. The two independent two-state transitions with domains dissociation model

First, for investigating of thermal multi-domains protein stability, the modified Gibss-Helmholts equation is determined. The fundamental assumption in this model is the equilibrium reversibility of thermal denaturation process in multimeric protein between the folded and unfolded states. As the inherent difficulty in the treatment and analysis of their equilibrium behavior in experimental scope, the use of the Gibss-Helmholts empirical modified function in terms of the number of domains can be a cross cut to thermodynamic purposes.

Eq. (1) shows the equilibrium between n-identical domain protein and its unfolded monomer without any intermediate state is shown in:

$$F_n \leftrightarrow nU \tag{1}$$

The reaction equilibrium constant is as follows:

$$K_D = \frac{[U]^n}{[F_n]} \tag{2}$$

The definition of unfolded protein fraction, f_D is expressed as follow:

$$f_D = \frac{[U]}{P_t} \tag{3}$$

Where, P_t is the total protein concentration in domain units. As fallow K_{unf} and ΔG^o can be expressed as functions of f_D:

$$[F_n] = \frac{P_t(1-f_D)}{n}, \quad [U] = P_t f_D \tag{4}$$

$$K_D = \frac{n f_D^n P_t^{n-1}}{1-f_D} \tag{5}$$

$$\Delta G_D^o = -RT \ln \frac{n f_D^n P_t^{n-1}}{1-f_D} \tag{6}$$

In order to determine the modified ΔH_D and ΔS_D functions in terms of the temperature, applying corresponding prameters (ΔH_{Deff},ΔG_{Deff}) will be a useful approach. In the following way:

$$\Delta H_{Deff} = \Delta H_{Geff} + \Delta C_P (T - T_G) \tag{7}$$

and

$$\Delta H_D = RT^2 \frac{d \ln K_D}{dT} = RT^2 \frac{d \ln\left(\frac{n f_D^n P_t^{n-1}}{1-f_D}\right)}{dT} \tag{8}$$

$$\Delta H_{Deff} = RT^2 \frac{d \ln K_{Deff}}{dT} = RT^2 \frac{d \ln\left(\frac{n f_D}{1-f_D}\right)}{dT} \tag{9}$$

The values of ΔH_D and ΔH_{Deff} at T_G (Temperature where $\Delta G_D = 0$) (ΔH_G and ΔH_{Geff}, respectively), can be obtained as follow:

$$\Delta H_G = \lim_{T \to T_G} \Delta H_D = \lim_{T \to T_G} \left(RT^2 \frac{d \ln K_D}{dT} \right) = RT_G^2 \times \left(\frac{n + (1-n) f_D}{f_D (1-f_D)} \times \left(\frac{df_D}{dT} \right)_{T_G} \right) \tag{10}$$

$$\Delta H_{Geff} = \lim_{T \to T_G} \Delta H_{Deff} = \lim_{T \to T_G} \left(RT^2 \frac{d \ln K_{Deff}}{dT} \right) = RT_G^2 \times \left(\frac{1}{f_D (1-f_D)} \times \left(\frac{df_D}{dT} \right)_{T_G} \right) \tag{11}$$

The following equation can be easily extracted by comparing Eqs. (10) and (11).

$$\Delta H_{Geff} = \Delta H_G \times \left[\frac{1}{(n + (1-n) f_D)} \right] \tag{12}$$

Substitution of the above mentioned equation into Eq. (7), the modified function for ΔH_{Deff} in terms of the number of domains can be obtained by:

$$\Delta H_{Deff} = \frac{\Delta H_G}{n + (1-n) f_D} + \Delta C_P (T - T_G) \tag{13}$$

By following the similar trend for ΔH_{Deff}, the corresponding statement for ΔS_{Deff} can be obtained as follow:

$$\Delta S_D = \frac{\Delta H_D - \Delta G_D}{T} = RT \frac{d \ln K_D}{dT} + R \ln K_D = RT \frac{d \ln\left(\frac{n f_D^n P_t^{n-1}}{1-f_D}\right)}{dT} + R \ln \left(\frac{n f_D^n P_t^{n-1}}{1-f_D} \right) \tag{14}$$

On the other hand, it can be rewritten as:

$$\Delta S_{Deff} = \Delta S_{Geff} + \Delta C_P \ln \frac{T}{T_G} = RT \frac{d \ln\left(\frac{f_D}{1-f_D}\right)}{dT} + R \ln \left(\frac{f_D}{1-f_D} \right) \tag{15}$$

Subsequently, it can be shown that:

$$\Delta S_G = \lim_{T \to T_G} \Delta S_D = RT_G \left(\frac{n + (1-n) f_D}{f_D (1-f_D)} \times \left(\frac{df_D}{dT} \right)_{T_G} \right) + R \ln \left(\frac{f_D}{1-f_D} \right) + RT \left(n f_D^{n-1} P_t^{n-1} \right) \tag{16}$$

and

$$\Delta S_{Geff} = \lim_{T \to T_G} \Delta S_{Deff} = RT_G \left(\frac{1}{f_D (1-f_D)} \times \left(\frac{df_D}{dT} \right)_{T_G} \right) + R \ln \left(\frac{f_D}{1-f_D} \right) \tag{17}$$

so

$$\Delta S_{Geff} = \Delta S_G + RT_G \left(\frac{n + (1-n) f_D}{f_D (1-f_D)} \times \left(\frac{df_D}{dT} \right)_{T_G} \right) - R \ln (n f_D^{n-1} P_t^{n-1}) \tag{18}$$

and

$$\Delta S_{Deff} = \Delta S_D + RT \left(\frac{n + (1-n) f_D}{f_D (1-f_D)} \times \left(\frac{df_D}{dT} \right) \right) + R \ln (n f_D^{n-1} P_t^{n-1}) \tag{19}$$

Therefore:

$$\Delta S_{Deff} = \Delta S_G + RT_G \frac{(1-n)}{f_D} \left(\frac{df_D}{dT} \right)_{T=T_G} - R \ln (n f_D^{n-1} P_t^{n-1}) + \Delta C_P \ln \left(\frac{T}{T_G} \right) \tag{20}$$

The modified function of ΔS_D in terms of the number of domains (Eq. (21)) derived by substitution of Eq. (20) into Eq. (15).
ΔG_{Deff} can be defined as follow:

$$\Delta G_{Deff} = -RT \ln \left(\frac{f_D}{1-f_D} \right) \tag{21}$$

The modified function of ΔG_{Deff} in terms of the number of domains (Eq. (23)) can be obtained by substitution of Eq. (22) into Eq. (6).

$$\Delta G_{Deff} = \Delta G_D + RT \ln (n f_D^{n-1} P_t^{n-1}) \tag{22}$$

The modified Gibss-Helmholtz equation (Eq. (23)) for multi-domain proteins undergo thermal denaturation can be obtained by substitution of corresponding equation for ΔH_{Deff}(Eq. (13)), ΔS_{Deff} (Eq. (20)) and ΔG_{Deff} (Eq. (21)), into famous Gibbs Eq. ($\Delta G_{Deff} = \Delta H_{Deff} + T\Delta S_{Deff}$).

$$\Delta G_D = \Delta H_G \left(\frac{1}{n + (1-n) f_D} - \frac{T}{T_G} \right) + \Delta C_P \left[(T_G - T) + T \ln \frac{T}{T_G} \right] - RTT_G \times \frac{(1-n)}{f_D} \times \left(\frac{df_D}{dT} \right)_{T_G} \tag{23}$$

This is a well-defined function for thermodynamic stability of multi-domain proteins that undergo reversible thermal denaturation. The theoretical results of ΔH_{Deff} and ΔG_{Deff} from in this study (Eqs. (8)–(23)) provided in Table 1 to be compared with experimental data [12]. As mentioned earlier, the values of $\Delta H_{Deff}^{\mathrm{mod}\ el}$ arising from present model are in good agreement with the calorimetric ΔH^{cal} values. They can fit a two-state transition process. According to the DSC thermogram, it was also determined that the low pH-induced denaturation of pepsin affects the domains of the protein differently in comparison with heat induced denaturation. This difference indicates the presence of thermal fluctuations in the native conformation of pepsin [12]. It is also consistent with the non-ideal unfolding observed in DSC experiments. The thermal denaturation determinations of ΔG_{Deff} would confirm the assumption that DSC transition peaks can indeed be evaluated by thermodynamic transition patterns. Moreover, it allows the analysis of protein unfolding transitions by the use of thermodynamic variation in state models. The analysis of the transitions by a two-state transition model, requires that the DSC results, in terms of T_m and the unfolding

Table 1
The theoretical thermodynamic parameters at T_G for the thermal unfolding of pepsin at different pH values within the framework of this study.

pH	T_G (K)	ΔH^{cal} (kcal mol^{-1})	$\Delta H_{Deff}{}^{model}$ (kcal mol^{-1})	$\Delta G_{Deff}{}^{model}$ (kcal mol^{-1})
The first transition				
1.0	–	–	–	–
2.0	305.55 ± 0.2	128 ± 5	105.7	117.96
3.0	322.15 ± 0.2	76 ± 4	70.12	−0.9
4.0	320.05 ± 0.2	44 ± 3	24.43	45.8
The second transition				
1.0	332.65 ± 0.2	210 ± 5	149.48	0.006
2.0	339.85 ± 0.2	133 ± 3	128.52	−2.2
3.0	348.55 ± 0.2	117 ± 4	43.45	−49.43
4.0	349.85 ± 0.2	108 ± 2	80.28	−56.8

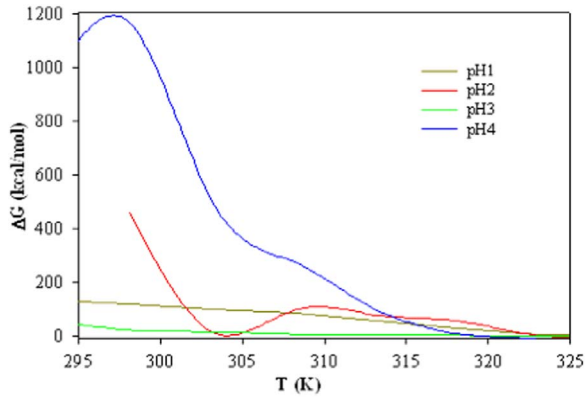

Fig. 1. Temperature dependence of the Gibbs energy difference in the native and denatured states of Pepsin at various pH values as a function of temperature.

enthalpy, as is indeed observed for the multimeric proteins. Extensive calorimetric studies on small globular domains have demonstrated that these proteins typically show an ideal two-state behavior. The exothermic enthalpy variation observed in a DSC experiment is accredit to unfolding of a part of the protein molecule. In base of the theoretical values of ΔG_{Deff} in this study, the stability of arrangement in pepsin can be predicted at four mentioned pH (1–4) can be predicted. All experimental data suggests that at pH 4 the stability of pepsin is enhanced, as illustrated in Fig. 1, the positive big ΔG_{Deff} value at pH 4 in T=298.15 K is evident on this suggestion. At the temperature of the first peak observed (and for low heating rates), it was observed that the concentration of unfolded pepsin is high; accordingly, so is the rate of aggregation was also high. At those situations, most pepsin molecules will be thoroughly unfolded before being incorporated in the aggregations. At higher temperatures (low heating rates) the unfolding takes place at a low speed, which leads to a low concentration of (partially) unfolded pepsin molecules. The rate of aggregation is respectively slower and it could well be that at such low aggregation rates pepsin molecules are incorporated in the aggregation before they obtain efficient time for complete unfolding.

3.3. Multi-domain homology

The route and method of denaturation affect the structure of the formed aggregates. We comput the problem of homology modeling identification is computed in complex multi-domain families that have varied domain architectures. The Swiss-Model template library (SMTL) was performed along with Blast [14] and HHBlits [15] for evolutionary related structures matching of target sequence Sus scrofa (Pig) Pepsin A. Overall 549 templates were found, some of them are mentioned in Table 2. The templates with the highest quality have then been selected for building a model. Models are built based on the target-template alignment using Promod-II (Fig. 2). Coordinates are first coneserved between the target and the template, are copied from the template to the model. Insertions and deletions are remodeled using a fragment library (Fig. 3). Side chains are then reconstructed. Eventually, the geometry of the resulted model is equalized by using a force field. In case loop modeling with ProMod-II [20] does not give gratifying results, an alternative model is built with Modeller [21]. The global and per-residue model quality is appraised using the QMEAN scoring function [22]. For enhancing the performance, weights of the individual QMEAN terms have been trained specifically for Swiss Model. Ligands present in the template structure (N – Ethoxycarbonyl - L-leucyl-N-[(1R,2S,3S)-1-(cyclohexylmethyl) - 2,3- dihydroxy-5-methyl-hexyl] -L- leucinamide) are transferred by homology to the model. When the following criteria are met (Gallo -Casserino, to be published): (a) The ligands are annotated as biologically relevant in the template library, (b) the ligand is in contact with the model, (c) the ligand is not clashing with the protein, (d) the residues in contact with the ligand are conserved between the target and the template. If any of these four principal is not satisfied, a certain ligand will not be included in the model. The model summary contains some information on why and which ligand has not been covered (Fig. 4). Homo oligomeric structure of the target protein is predicted based on the analysis of pairwise interfaces of the identified template structures. For each relevant interface between polypetide chains (interfaces with more than 10 residue-residue interactions), the Qscore Oligomer [23] is predicted by considering features such as similarity to target and frequency of observing this interface in the identified templates (Kiefer, Bertoni, Biasini, to be published). The prediction is performed with a random forest regressor using these features as input parameters to predict the probability of conservation for each interface (Fig. 5). The Qscore Oligomer of the whole complex is then calculated as the weight averaged Qscore Oligomer of the interfaces. The oligomeric state of the target is predicted to be the same as in the template. The similarity is when Qscore Oligomer is predicted to be higher or equal to 0.5.

Furthermore, it can be estimate the correlation between the change in heat capacity (ΔC_p) and the surface area, depends on unfolding ΔASA_{unf} for a set of domain proteins. It is widely accepted that the value of ΔH^{cal} is dependent on the change in solvent accessible to non polar surface area upon unfolding (ΔASA_{unf}), which is directly proportional to the number of residues in a protein. Such data has been assembled for a variety of template domain proteins, and the slope of the plot of ΔC_p vs. number of residues is 0.020 kcal - mol^{-1} deg^{-1} residues^{-1}. Interestingly, the slope of a plot of ΔC_p vs. number of residues for the set of globular proteins is 0.015 kcal - mol^{-1} deg^{-1} residues^{-1}, which is significantly smaller than for template domain proteins (Fig. 6). This observation implies that the surface area exposed to upon unfolding of a globular protein is smaller than expected from the number of residues in the protein. It can be rationalized by two unusual features of domain proteins: 1) their unfolded state is more compact than predicted by the typical self avoiding random walk treatment of the denatured state, and 2) their elongated native state results in a greater surface area to volume ratio than for a globular domain protein. These effects both will contribute to the observed deviation from the behavior of domain proteins. The change in accessible surface area/energy while going from folded to compact unfolded is less than that in going from folded to extended unfolded. We observe a small deviation for domain proteins with 20 or fewer repetitions (Fig. 7). Therefore, it can be concluded that the predominant cause of the deviation in the ΔC_p values is the nature of the unfolded state.

3.4. Thermodynamic domain interfaces

Domain interfaces in several multi-domain proteins are involved in some important function, i.e., they take part in binding or catalysis or act as hinges to facilitate conformational transitions and it may not be

Table 2
The templates for the homology modeling of Pepsin A.

Template	Seq identity	Oligo-state	Found	Method	Resolution	Seq similarity	Coverage
2PSG.1.A	99.46	homo-dimer	HHblits	X-ray	1.80 Å	0.61	0.96
1PSA.1.A	100.00	homo-dimer	BLAST	X-ray	2.90 Å	0.61	0.85
1YX9.1.A	100.00	monomer	BLAST	X-ray	3.00 Å	0.61	0.85
1F34.1.A	100.00	hetro-oligomer	BLAST	X-ray	2.45 Å	0.61	0.85
3PEP.1.A	99.69	monomer	BLAST	X-ray	2.30 Å	0.61	0.85
4PEP.1.A	99.69	monomer	BLAST	X-ray	1.80 Å	0.61	0.85

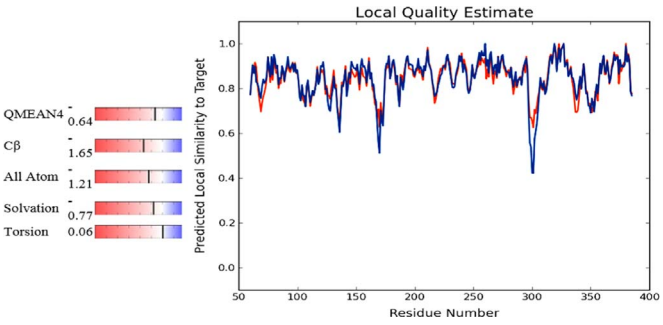

Fig. 2. Local similarity quality estimate in target vs, residue number.

possible to tune the functional interfaces to promote folding. We computationally investigated the role of interface domain interaction in full length protein that contains two domain of protein Sus scrofa (Pig) Pepsin A and we found that an altogether domain insertion, promotes unfolding. Strong interface domain interaction is involved in unfolding of multi-domain protein. The simulations were performed close to the unfolding temperature, T_G, where multiple transitions occur between the equally populated folded and unfolded ensembles. As a result the best sampling of the transition region is achieved. The presence of a single free energy barrier separating the native and unfolded ensembles at T_G implies that the protein unfolds. If the different domains of a multi-domain protein fold at different T_Gs, partially unfolded states get populated at temperatures between the lowest and the highest domain specific T_Gs. Upon mutation, a domain specific decrease of T_G can result in the incomplete folding of that domain at the T_G of the whole protein and the population of partially unfolded states in the folded ensemble. This process in reduced folding cooperativity. Unfolding is usually deduced from the heat capacity curve using the ratio of the van't Hoff enthalpy (ΔH^{VH}) to the calorimetric enthalpy (ΔH^{cal}). We computed the stability (ΔG_{Deff}) of

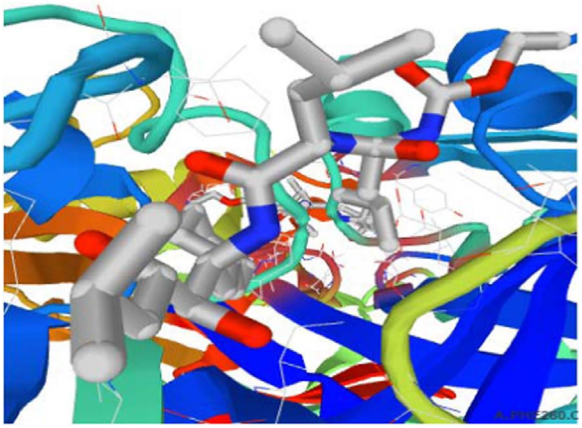

Fig. 4. The geometry of the model with ligand by using a force field.

each of the proteins/domains in these templates using the energy function (Fig. 8). It was found that the domains of multi-domain proteins considered in isolation (7.65 ± 0.86 Kcal.mol-1). Examination of unfolding with DSC indicated that at least two unfolding transitions exist. At and the higher temperature (and lower enthalpy) transition correlated with low aggregation. This result revealed that the least conformationally stable regions of a multi-domain protein are not necessarily the most aggregation.

We make changes to Pepsin A which alters the free energetic balance between domains and in turn the unfolding. However, since the domains of Pepsin A are of unequal in size. The largest contribution to $C_v(T_G)$ comes from the unfolding of core which is not perturbed much in simulations. Thus, the $\Delta H^{VH}/\Delta H^{cal}$ is not a sensitive measure of the unfolding of Pepsin A and its mutants. Therefore, we use the height of the free energy barrier was used at T_G and the "unfoldedness" of the protein in this fold ensemble to infer the degree of unfolding. We

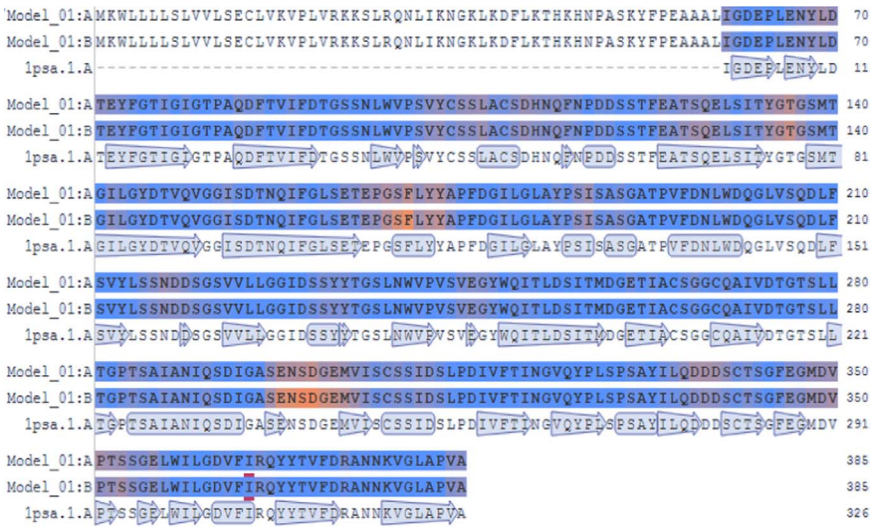

Fig. 3. Fragment library of coordinates which are conserved between the target and the template.

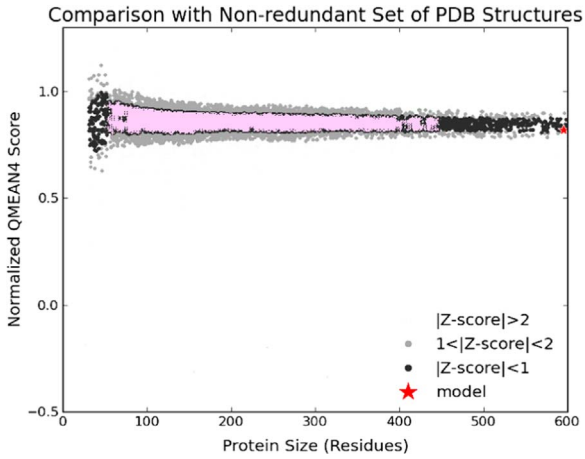

Fig. 5. Normalize scores for a number of residues non-redundant set of PDB structure.

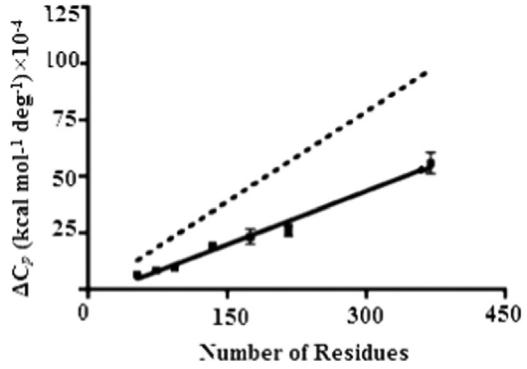

Fig. 6. Linear relationship between ΔC_p and the number of residues in the protein. ΔC_p is plotted versus number of residues for the proteins acid protease (solid symbols). The solid black line is linear fit of the data, giving a slope of 0.020 kcal mol^{-1} deg^{-1} $residues^{-1}$. The dotted black shows the collected data for ΔC_p versus number of residues for many globular domain proteins. The slope of this line is 0.015 kcal mol^{-1} deg^{-1} $residues^{-1}$.

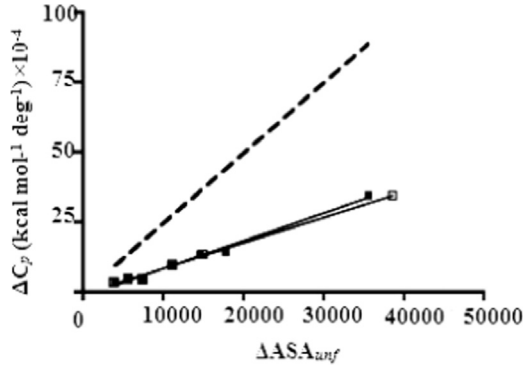

Fig. 7. Correlation between the change in heat capacity (ΔC_p) and the surface area exposed to unfolding ΔASA_{unf} for a set of globular domain proteins (dashed line) and for acid protease proteins (symbols). The filled squares represent the ΔASA_{unf} calculated using ASA_N from the acid protease crystal structures in the unfolded state. The empty squares represent the ΔASA_{unf} calculated using ASA_N from the acid protease number of residues and the surface area exposed expected for globular domain protein of that size in the unfolded state. The solid black lines show the linear regression fit to the data.

define unfoldedness is defined as the ratio of the population of a mutant at the value of the reaction coordination where Pepsin A is unfolded to the population of Pepsin A at the value of the reaction coordination where Pepsin A is unfolded. This definition intrinsically assumes that the value of the reaction coordination where Pepsin A is unfolded is greater than or equal to the amount of the reaction coordinate where the mutants are unfolded.

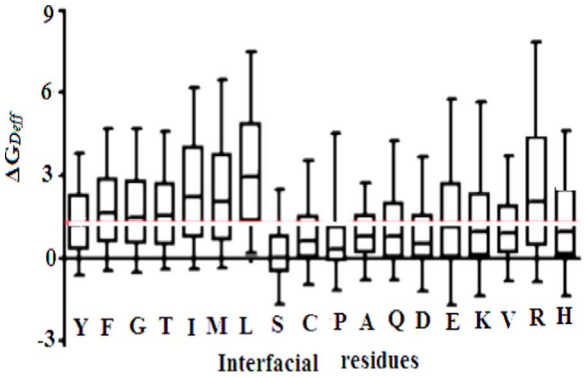

Fig. 8. Effect of interfacial residues on the energy of association between the interacting domains. Each of the interfacial residues of the 450 interfaces was mutated to Alanine and the change in the interaction energies between the two domains was computed. All residues except Glycines and Alanines were mutated in this analysis.. It is observed that the medians values of ΔG for these mutations are above the 1.5 kcal mol^{-1} threshold.

The presence of multiple domains in proteins can cause interactions between partially unfolded domains and in turn to increment unfolding. However, several multi-domain proteins unfold unstablility in vitro. Unfolding, the all or no folding of a protein with the population of few intermediates, diminishes partly folded states.

4. Conclusions

It was shown that the use of effective ΔG provides some valuable information about the stability of multimeric proteins. ΔG_{Deff} is a useful parameter that gives an intuitive appreciation of the stability of multimeric proteins, one can calculate the fraction of unfolded protein from ΔG_{Deff} of a multimeric protein can be calculated by the $\Delta G°$ of a monomeric protein. The DSC thermogram can be fit to the $\Delta H_{Deff}^{mod\ el}$ for a two-state process, and the unfolding process. The reversible unfolding of the protein described by the two-state transition quantities is followed by a slower irreversible process of aggregation.

This model can be used to analyze the changes in the structural and functional properties of a number of large multimeric proteins subjected to broad range temperature variations. The description of the thermal denaturation of multimeric proteins may demand more complex models.

Acknowledgments

The authors are greatfully acknowledge from financial support by Isfahan University and Falavarjan Branch, Islamic Azad University (Grant no. 51723890822022). Authors also appreciate the Swiss-Model for the homology modeling project "PEPA_PIG P00791 Pepsin A(3.4.23.1)", and for help in evolutionary related structures matching the target sequence.

Appendix A. Transparency document

Transparency document associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.bbrep.2017.01.005.

References

[1] V.G. Panse, C.P. Swaminathan, J.J. Aloor, A. Surolia, R. Varadarajan, Unfolding thermodynamics of the tetrameric chaperone, SecB, Biochem. 39 (2000) 2362–2369.
[2] P.L. Privalov, S.A. Putekhin, Scanning microcalorimetry in studying temperature-induced changes in proteins, Methods Enzymol. 131 (1986) 4–51.
[3] G.D. Manetto, C. La Rosa, D.M. Grasso, D. Milardi, Evaluation of thermodynamic properties of irreversible protein thermal unfolding measured by DSC, J. Therm. Anal. Cal. 80 (2005) 263–270.
[4] A.A. Makarov, N.G. Esipova, D.R. Monaselidze, G.N. Mgeladze, G.V. Madzhagaladze, L.P. Kuranova, A.L. Grebenko, Microcalorimetry of thermal

denaturation of pepsin and trypsin in the crystal state, Biochim. Biophys. Acta 434 (1976) 286–289.
[5] V.G. Artiukhov, O.V. Basharina, G.A. Vashanov, M.A. Nakvasina, O.V. Putintseva, Thermal denaturation of oligomeric proteins: structural and functional modifications, the sequence of steps, Biofizika 49 (4) (2004) 617–630.
[6] S.R. Tello-Solis, A. Hernández-Arana, Effect of irreversibility on the thermodynamic characterization of the thermal denaturation of Aspergillus saitoi acid proteinase, Biochem. J. 311 (1995) 969–974.
[7] J. Ramprakash, V. Doseeva, A. Galkin, W. Krajewski, L. Muthukumar, S. Pullalarevu, E. Demirkan, O. Herzberg, J. Moult, F.P. Schwarz, Comparison of the chemical and thermal denaturation of proteins by a two-state transition model, Anal. Biochem. 374 (2008) 221–230.
[8] C. Park, S. Marqusee, Analysis of the stability of multimeric proteins by effective ΔG and effective m-values, Protein Sci. 13 (2004) 2553–2558.
[9] P.S. Kim, R.L. Baldwin, Specific intermediates in the folding reactions of small proteins and the mechanism of protein folding, Annu. Rev. Biochem. 51 (1982) 459–489.
[10] P.L. Privalov, Stability of proteins: small globular proteins, Adv. Protein Chem. 33 (1979) 167–241.
[11] A. Hernández-Arana, M. Soriano-Garcia, Detection and characterization by circular dichroism of a stable intermediate state formed in the thermal unfolding of papain, Biochim. Biophys. Acta 954 (1988) 170–175.
[12] A.A. Rastegari, B. Behnaz Buzari, V. Pavelkic, K. Gopcevic, M. Petkovic, A.K. Bordbar, Thermal denaturation of pepsin at acidic media: using DSC, MALDI TOF MS and PAGE techniques, Thermochim. Acta 568 (2013) 165–170.
[13] G. Fiorentino, I.D. Giudice, L. Petraccone, S. Bartolucci, P.D. Vecchio, Conformational stability and ligand binding properties of BldR, a member of the MarR family, from Sulfolobus solfataricus, Bioch. Bioph. Acta 1844 (2014) 1167–1172.
[14] S.F. Altschul, T.L. Madden, A.A. Schaffer, J. Zhang, Z. Zhang, W. Miller, D.J. Lipman, Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, Nucleic Acids Res. 25 (17) (1997) 3389–3402.
[15] M. Remmert, A. Biegert, A. Hauser, J. Soding, HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment, Nat. Methods 9 (2012) 173–175.
[16] A. Andreeva, D. Howorth, J.M. Chandonia, S.E. Brenner, T.J. Hubbard, C. Chothia, A.G. Murzin, Data growth and its impact on the SCOP database: new developments, Nucleic Acids Res. 36 (2008) D419–D425.
[17] L.H. Greene, T.E. Lewis, S. Addou, A. Cuff, T. Dallman, M. Dibley, O. Redfern, F. Pearl, R. Nambudiry, A. Reid, I. Sillitoe, C. Yeats, J.M. Thornton, C.A. Orengo, The CATH domain structure database: new protocols and classification levels give a more comprehensive resource for exploring evolution, Nucleic Acids Res. 35 (2007) D291–D297.
[18] J. Schymkowitz, J. Borg, F. Stricher, R. Nys, F. Rousseau, L. Serrano, The FoldX web server: an online force field, Nucleic Acids Res. 33 (2005) W382–W388.
[19] DSC Data Analysis in Origin, Tutorial Guide. The Calorimetry Experts, Version 5.0, 1998, p. 66.
[20] N. Guex, M.C. Peitsch, SWISS-model and the Swiss-PdbViewer: an environment for comparative protein modeling, Electrophoresis 18 (1997) 2714–2723.
[21] A. Sali, T.L. Blundell, Comparative protein modelling by satisfaction of spatial restraints, J. Mol. Biol. 234 (1993) 779–815.
[22] P. Benkert, M. Biasini, T. Schwede, Toward the estimation of the absolute quality of individual protein structure models, Bioinformatics 27 (2011) 343–350.
[23] V. Mariani, F. Kiefer, T. Schmidt, J. Haas, T. Schwede, Assessment of template based protein structure predictions in CASP9, Proteins 79 (Suppl 10) (2011) S37–S58.