High-quality thermodynamic data on the stability changes of proteins upon single-site mutations

Fabrizio Pucci,[*] Raphaël Bourgeas,[†] and Marianne Rooman[‡]

Department of BioModeling, BioInformatics & BioProcesses,

Université Libre de Bruxelles, CP 165/61,

Roosevelt Avenue 50, 1050 Brussels, Belgium

Abstract

We have set up and manually curated a dataset containing experimental information on the impact of amino acid substitutions in a protein on its thermal stability. It consists of a repository of experimentally measured melting temperatures (T_m) and their changes upon point mutations (ΔT_m) for proteins having a well-resolved X-ray structure. This high-quality dataset is designed for being used for the training or benchmarking of in silico thermal stability prediction methods. It also reports other experimentally measured thermodynamic quantities when available, *i.e.* the folding enthalpy (ΔH) and heat capacity (ΔC_P) of the wild type proteins and their changes upon mutations ($\Delta\Delta H$ and $\Delta\Delta C_P$), as well as the change in folding free energy ($\Delta\Delta G$) at a reference temperature. These data are analyzed in view of improving our insights into the correlation between thermal and thermodynamic stabilities, the asymmetry between the number of stabilizing and destabilizing mutations, and the difference in stabilization potential of thermostable versus mesostable proteins.

---

[*]Electronic address: `fapucci@ulb.ac.be`

[†]Electronic address: `rbourgeas@ulb.ac.be`

[‡]Electronic address: `mrooman@ulb.ac.be`

I. INTRODUCTION

The availability of a complete and well-curated dataset for training and testing purposes is a basic prerequisite for the development of any knowledge-based bioinformatics prediction tool. Here we present a repository containing thermal and thermodynamic stability data on experimentally characterized single-site protein mutants for which an X-ray structure is available, which we have set up by screening the literature and freely accessible databases. This dataset is meant to be as complete as possible, and to contain as much as possible noise- and error-free data. The amount and quality of the data used to set up a predictor are indeed two fundamental requirements for getting reliable predictions.

We have used this dataset to design and test our method for predicting the change in melting temperature of proteins upon point mutations [1], and to compare its performance with that of other existing tools. This dataset is intended as a common benchmark for training and validating different in silico tools for protein stability prediction and rational design.

There is an increasing interest for the development of reliable stability predictors that can be used for rationally designing protein mutants with improved properties, and there is thus concomitantly an increasing need for high-quality and easily accessible datasets. Indeed, the design of new enzymes and other proteins that remain stable and active in unusual environments or at temperatures that differ from their physiological temperatures would allow the optimization of a wide series of biotechnological processes in many sectors such as agro-food, biopharmaceuticals and environment [2, 3].

Another asset of this dataset is that it can serve as a basis for large-scale analyses in view of improving our understanding of the factors that modulate the thermal resistance and other thermodynamic properties of proteins and their variation upon mutation. As a matter en fact, we performed in this paper some analyses that yield some interesting insights.

II. BRIEF THEORETICAL REVIEW OF PROTEIN STABILITY

The protein folding transition is thermodynamically characterized by a change in free energy, enthalpy, entropy and heat capacity. Assuming protein folding to be a two-state transition and the change in heat capacity ΔC_p to be temperature independent, the folding

2

free energy $\Delta G(T) = G^{folded}(T) - G^{unfolded}(T)$ can be written as

$$\Delta G(T) = \Delta H_m \left(1 - \frac{T}{T_m}\right) - \Delta C_p \left((T_m - T) + T \operatorname{Log}\left[\frac{T}{T_m}\right]\right), \tag{1}$$

where T_m is the melting temperature and ΔH_m the enthalpic change measured at T_m. Note that with these conventions, ΔH_m and ΔC_p are negative; the folding free energy $\Delta G(T)$ is negative under physiological conditions and is positive for temperatures above T_m. The protein stability curve has thus an inverse bell shape, as shown in Figure 1.
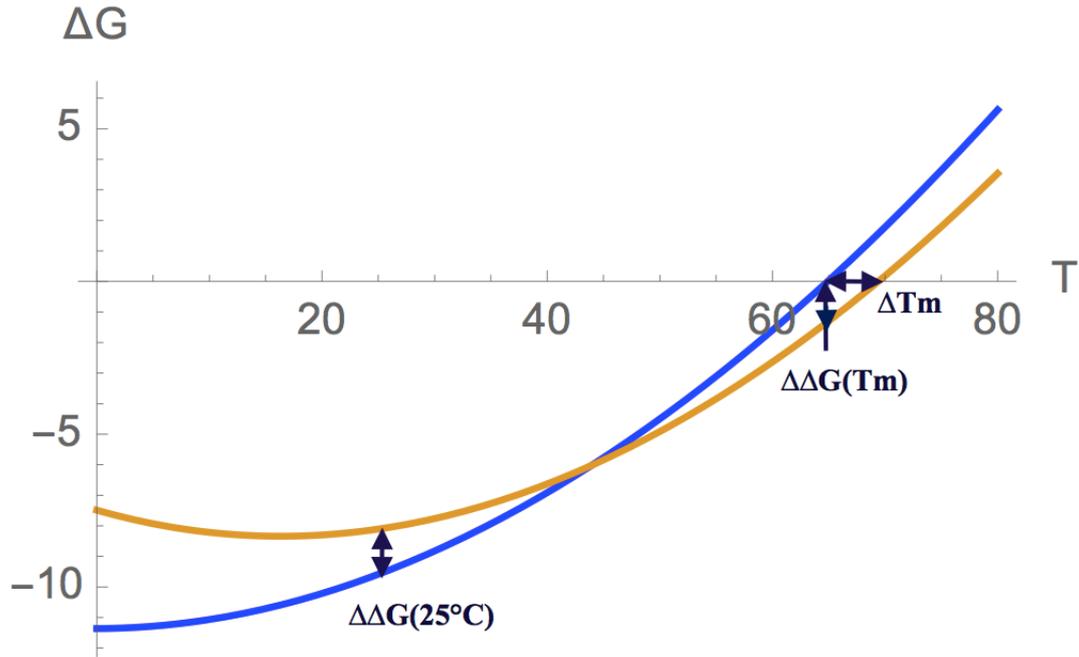

FIG. 1: Folding free energy in kcal/mol as a function of the temperature for human lysozyme (PDB code 1LZ1) (blue curve) and its mutant R21A (green curve). While there is an anticorrelation between ΔT_m and $\Delta\Delta G(T_r)$ at $T_r = T_m$, there is a correlation between them at $T_r = 25°C$. This is an example of mutation that thermally stabilizes the protein and thermodynamically destabilizes it.

When one or several residues of the wild type protein are substituted, the change in protein stability due to the mutation can be characterized by a temperature descriptor, namely the change in melting temperature upon mutation:

$$\Delta T_m = T_m^{mut} - T_m^{wild}, \tag{2}$$

3

which measures the change in thermal stability. It can also be characterized by a free energy descriptor, *i.e.* the change in folding free energy a the reference temperature T_r (usually chosen to be the room temperature $T_r = 298$ K):

$$\Delta\Delta G(T_r) = \Delta G^{mut}(T_r) - \Delta G^{wild}(T_r), \tag{3}$$

which measures the change in so-called thermodynamic stability. Note that with these conventions, thermally stabilizing mutations have positive ΔT_m-values, and thermodynamically stabilizing mutations have negative $\Delta\Delta G(298K)$ values.

The dataset presented in this paper has been created in view of developing a predictor of protein thermal stability changes upon point mutations [1]. It thus contains all the point mutations with experimentally measured ΔT_m values that we have collected from the literature and satisfy certain criteria. The corresponding values of $\Delta\Delta G(T_r)$ and of the other thermodynamic quantities appearing in equation (1) are known only for a subset of the entries.

In order to understand the precise relation between thermal and thermodynamic stability changes, one needs to have independent experimental determinations of their respective descriptors ΔT_m and $\Delta\Delta G(T_r)$, or to know the values of all the thermodynamic quantities that appear in equation (1), *i.e.* T_m, ΔH_m and ΔC_P, for both the wild type and mutant proteins. Unfortunately all these informations are not always available or not sufficiently accurately measured. Under the assumption that the mutated protein is a perturbation of the wild type, some approximations can be made; for example it is quite reasonable to consider that the parameter

$$x = \frac{\Delta T_m}{T_m^{wild}} << 1, \tag{4}$$

with the temperature expressed in Kelvin, is small and thus that an expansion of equation (3) in powers of x can be performed. This yields:

$$\Delta\Delta G(T_r) = \frac{\Delta H_m T_r}{(T_m^{wild})^2}\Delta T_m + \Delta\Delta H_m \left[1 - \frac{T_r}{T_m^{wild}} + \frac{T_r}{(T_m^{wild})^2}\Delta T_m\right]$$
$$+ \Delta\Delta C_P \left[T_r - T_m^{wild} - T_r \log \frac{T_r}{T_m^{wild}}\right] + \mathcal{O}\left[x^2\right], \tag{5}$$

4

where $\Delta H_m = \Delta H^{wild}(T_m^{wild})$, $\Delta\Delta H_m = \Delta H^{mut}(T_m^{wild}) - \Delta H^{wild}(T_m^{wild})$ and $\Delta\Delta C_P = \Delta C_P^{mut} - \Delta C_P^{wild}$. As seen from this equation, the correlation between $\Delta\Delta G(T_r)$ and ΔT_m generically depends on an intricate combination of variations of thermodynamic quantities. If we assume $\Delta\Delta H_m \simeq 0$ and $\Delta\Delta C_p \simeq 0$, equation (5) reduces to:

$$\Delta\Delta G(T_r) \simeq \frac{\Delta H_m T_r}{(T_m^{wild})^2}\Delta T_m. \tag{6}$$

Under this (strong) assumption, we find thus a linear relation between $\Delta\Delta G(T_r)$ and ΔT_m; the proportionality coefficient is however protein-dependent. Note that ΔH_m is negative with our conventions, and that $\Delta\Delta G(T_r)$ and ΔT_m are thus anticorrelated.

On the other hand, at the reference temperature $T_r = T_m^{wild}$, equation (5) simplifies to:

$$\Delta\Delta G(T_m) = \frac{\Delta H^{mut}(T_m^{wild})}{T_m^{wild}}\Delta T_m + \mathcal{O}\left[x^2\right]. \tag{7}$$

The proportionality assumption between $\Delta\Delta G(T)$ and ΔT_m is thus valid at T_m. If moreover we assume $\Delta\Delta H_m \simeq 0$, this equation becomes the Becktel-Schellman formula [4]:

$$\Delta\Delta G(T_m) \simeq \Delta S_m \Delta T_m, \tag{8}$$

where $\Delta S_m = \Delta S^{wild}(T_m^{wild}) = \Delta H_m/T_m^{wild}$ is the entropic contribution at T_m^{wild}.

III. METHODS

A. Dataset design

We started collecting the mutations with experimentally measured ΔT_m values from the ProTherm database [5], and searched for additional entries by literature screening. Each entry (including those from ProTherm) was manually and carefully checked from the original literature to remove imprecisions and errors. We selected the mutations that satisfy the following criteria:

- Only single point mutations were included.

- Only mutations in proteins, whose three dimensional (3D) structures were experimentally solved by X-ray crystallography with a resolution of at most 2.5 Å, were considered.

- Only mutations that were experimentally characterized in monomeric proteins were taken into account, irrespective of the oligomeric state of the biological unit; this ensures that the measured T_m corresponds to the (un)folding transition and not to a change in quaternary state.

- Only wild-type and mutant proteins that are described in the reference articles as undergoing a two-state (un)folding transition were included.

- Destabilizing or stabilizing mutations by more than 20 °C were overlooked, as they are likely to induce important structural modifications.

When several experimental ΔT_m values were found in the literature for the same mutation, we chose the one measured at pH closest to seven and with the lowest concentration of additives; if more than one measurement in the same conditions was available, the average ΔT_m was taken.

In addition to the change in melting temperature upon mutation, other thermodynamic quantities associated to the mutation are reported in the dataset when available. These are the ΔC_P of the wild type protein and its change upon mutation $\Delta \Delta C_P$, $\Delta \Delta G(T_r)$ and the reference temperature T_r at which the measurement was performed, the ΔH_m of the wild type and

$$\overline{\Delta \Delta} H_m = \Delta H_m^{mut}(T_m^{mut}) - \Delta H_m^{wild}(T_m^{wild}) = \Delta \Delta H_m + \Delta C_P^{mut} \Delta T_m. \tag{9}$$

$\overline{\Delta \Delta} H_m$ refers to a quantity that is slightly different from $\Delta \Delta H_m$ appearing in equation (5): the former is computed at different T_m values whereas the latter is computed at T_m^{wild}; the difference is proportional to ΔT_m. We report in the dataset $\overline{\Delta \Delta} H_m$ rather than $\Delta \Delta H_m$ as these are the measured quantities. Note that for a given mutation these different quantities are not always measured in exactly the same experimentally conditions than the corresponding ΔT_m values.

When available, the $\Delta \Delta G$'s indicated in the dataset are the values that are measured by monitoring the (un)folding transition through chemical (de)naturation using urea or guanidinium chloride (GdmCl); the temperature at which the experiments were performed is also reported. If such data are not available but all the thermodynamic quantities in equation (1) are known for the wild type and the mutant proteins, they are used to evaluate $\Delta \Delta G$ at 25°C. Otherwise, approximations were made to evaluate $\Delta \Delta G$, and the corresponding entries

in the dataset are labeled by a subscript. The $\Delta\Delta G$ values obtained with the approximation consisting in considering $\Delta\Delta C_P \simeq 0$ are indicated with a subscript (b); the temperature at which they were estimated is equal to 25°C. When the stronger approximation consisting of supposing also $\Delta\Delta H_m \simeq 0$ (see equation (6)) is used to derive the value of $\Delta\Delta G$ from ΔT_m, we mark it with the subscript (a); the temperature at which this quantity is given is equal to T_m^{wild}. For the few entries whose values are labeled with a subscript (c), the $\Delta\Delta G(25°C)$ values are computed from ΔT_m using an empirical correlation between the two quantities computed on a subset of mutants of the same wild type protein [6]. Finally the $\Delta\Delta G$ values at T_m that are derived from the approximation (see ref. [7, 8]) $\Delta\Delta G(T_m) = \frac{\left(\Delta H_m^{wild} + \Delta C_P^{wild} \Delta T_m\right)\Delta T_m}{T_m}$ are labeled with the subscript (d).

The experimental techniques used for measuring the protein melting temperatures and other thermodynamic quantities are indicated in the dataset. These are differential scanning calorimetry (DSC), circular dichroism (CD), absorbance (Abs), and fluorescence.

The Protein DataBank (PDB) code [9] of the best resolved 3D X-ray structure of each wild type protein is specified in the dataset. For a few entries, the PDB code is labeled with a subscript. This means that the wild type structure of the protein whose ΔT_m was measured was unavailable, and that the structure of an almost identical protein was used instead, under the assumption that the impact of the modification on the structure is negligible. In particular, the 1bni$_{h102a}$ code means that the structure is obtained from the PDB structure 1bni with the His residue at position 102 manually substituted into an Ala. The same procedure is used for the PDB structures 1ycc$_{c102a}$, 1urp$_{l265c}$ and 5pti$_{m52l}$. The other PDB codes with subscripts, $i.e$ 1tpk$_r$, 1yu5$_{d1}$ and 1yu5$_{d2}$, refer to experimentally characterized proteins whose sequences have been manually truncated by a few residues compared to the original PDB structure. Note that we checked that the mutations or truncated residues in these pseudo-wild type proteins are all distant from the mutations whose ΔT_m was measured, so that they may be assumed as not interfering.

B. Data records

The dataset contains experimental information on 1,626 point mutations that have been introduced in about 93 proteins. This data was collected by screening the literature and databases, and carefully checked on the basis of the original articles. For each mutation, the

7

following informations are reported:

- The PDB [9] code of the 3D structure of the wild type protein (Column II).

- The chain name, residue number and residue name of the wild type and mutant amino acids (Columns III-VI).

- The experimental value of the change in melting temperature upon mutation (ΔT_m) using the convention of equation (2) (Column VII).

- The experimentally measured melting temperature (T_m) and the number of residues (N_r) of the wild type protein (Columns VIII and XIV, respectively).

- The experimental values of $\overline{\Delta\Delta H_m}$, which is the change in calorimetric enthalpy upon mutation measured at the mutant and wild type melting temperatures, respectively, as defined in equation (9) (Column IX), and of the ΔH_m of the wild type protein at T_m^{wild} (Column X), when available. The subscript (e) means that the reported values correspond to the van't Hoff enthalpy instead of the calorimetric enthalpy.

- The experimentally measured values of $\Delta\Delta C_P$ (Column XI) and the ΔC_P^{wild} of the wild type protein (Column XII), when available.

- The values of $\Delta\Delta G(T_r)$, using the conventions of equation (3) (Column XIII), and the reference temperature T_r (in degrees Celsius) at which they were measured or derived (Column XIV). Entries without superscripts are experimental or calculated from other measured thermodynamic quantities, whereas entries with superscripts (a), (b), (c) or (d) were obtained using different levels of approximations, as explained in the previous section.

- The resolution of the X-ray structure (in Å) (Column XVI).

- The name of the protein and its host organism (Columns XVII-XVIII)

- The bibliographic references (Column XIX).

- The pH and the experimental technique used for measuring ΔT_m (Columns XX-XXI).

8

IV.   RESULTS

We investigated some biophysical properties of the data reported in our dataset. First of all, the ΔT_m distribution obtained from all the entries is dominated by destabilizing mutations, as shown in Figure 2. The average T_m-value, $\langle \Delta T_m \rangle$, is indeed equal to -2.7°C, while the standard deviation and the kurtosis of the distribution are equal to 5.3°C and 4.0°C, respectively. The large majority of point mutations (about 70%) are thus destabilizing. Although this ΔT_m distribution is not built from the ensemble of possible mutations but rather from the subset of experimentally characterized mutations, we may nevertheless assume that it represents well the actual ΔT_m distribution of all possible mutations. The relative abundance of destabilizing mutations with respect to stabilizing ones can be interpreted as being due to the evolutionary force that tends to optimize the proteins for stability and thus to minimize the deleterious impact of random mutations. It must nevertheless be emphasized that all proteins are left with stability weaknesses [10] or frustrations [11], in particular in functional regions.
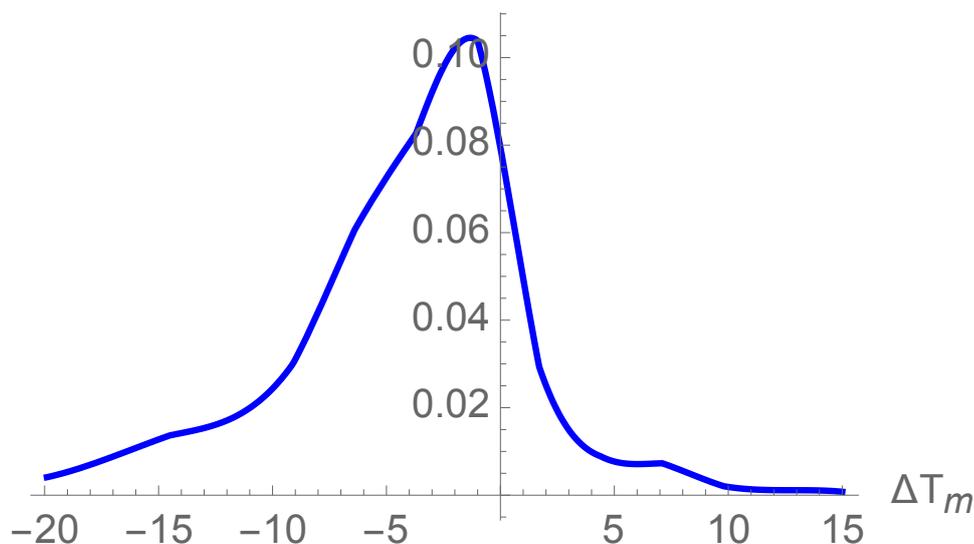

FIG. 2: Normalized ΔT_m distribution for the 1,626 mutations collected in the dataset.

The solvent accessibility of the mutated residues is an important feature that modulates the average stability changes. It is defined as the ratio between the solvent accessible surface of a residue X in a given structure and in the extended tripeptide Gly-X-Gly conformation, and has been computed using an in-house program [12]. In Figures 3a-c, we show the

9

experimental ΔT_m distribution as a function of the solvent accessibility of the mutated residues; three solvent accessibility (Acc) ranges are considered: Acc $< 15\%$ (core), $15\% <$ Acc $< 50\%$ (partially buried) and Acc $> 50\%$ (surface). The three distributions were found to be significantly different according to the 2-sample Kolmogorov-Smirnov (K-S) test (P-value $< 10^{-3}$). The mean $\langle \Delta T_m \rangle$ values of the distributions are equal to -4.3°C, -1.6°C and -1.1°C for the core, partially buried and surface mutations, respectively. As expected, the mutations in the core are on the average more destabilizing than those at the surface since core residues play a stronger role in the structural stability than surface residues, which also contribute to stability but to a lesser extent.
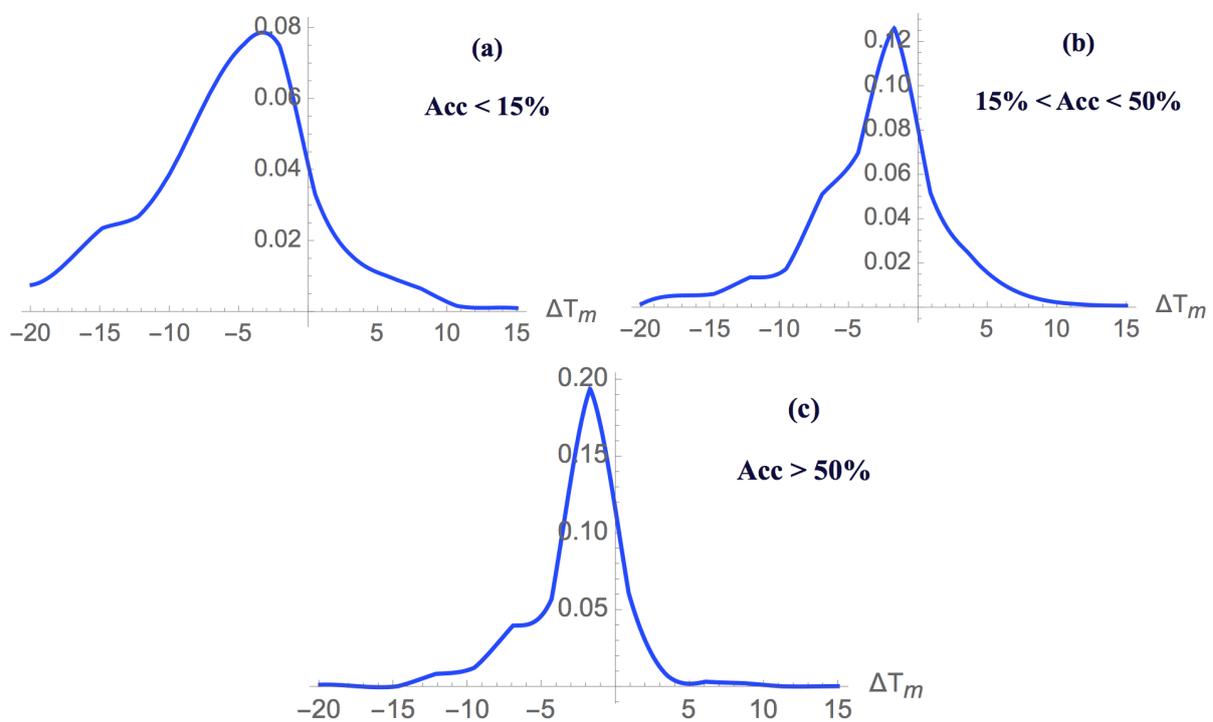

FIG. 3: Normalized ΔT_m distribution for : (a) 734 mutations in the protein core (Acc$<15\%$), (b) 513 mutations of partially buried residues ($15\%<$Acc$<50\%$) and (c) 379 mutations at the protein surface (Acc$>50\%$)

.

It is also informative to analyze the relation between ΔT_m and T_m. Indeed, it could be argued that it is "easier" to destabilize thermostable proteins or equivalently, to stabilize mesostable proteins. To check this hypothesis, we computed $\langle \Delta T_m \rangle$ separately for the mu-

10

tations introduced in thermostable proteins defined here as having a melting temperature higher than 65°C and those introduced in mesostable proteins with $T_m < 65$°C. We obtain a value of $\langle \Delta T_m \rangle = -3.6$°C for thermostable proteins and $\langle \Delta T_m \rangle = -2.1$°C for mesostable proteins. On the average, mutations in thermostable proteins are thus more destabilizing than in mesostable proteins. Moreover, the normalized ΔT_m distributions for mesostable and thermostable proteins are shown in Figure 4. They are statistically different according to the K-S test with a P-value $< 10^{-4}$, and the former appears to be shifted towards stabilizing mutations compared to the latter. This interesting result supports the view that the fraction of stabilizing mutations is larger in mesostable proteins than in thermostable proteins, and thus that the former are easier to stabilize than the latter, in agreement with the starting hypothesis.
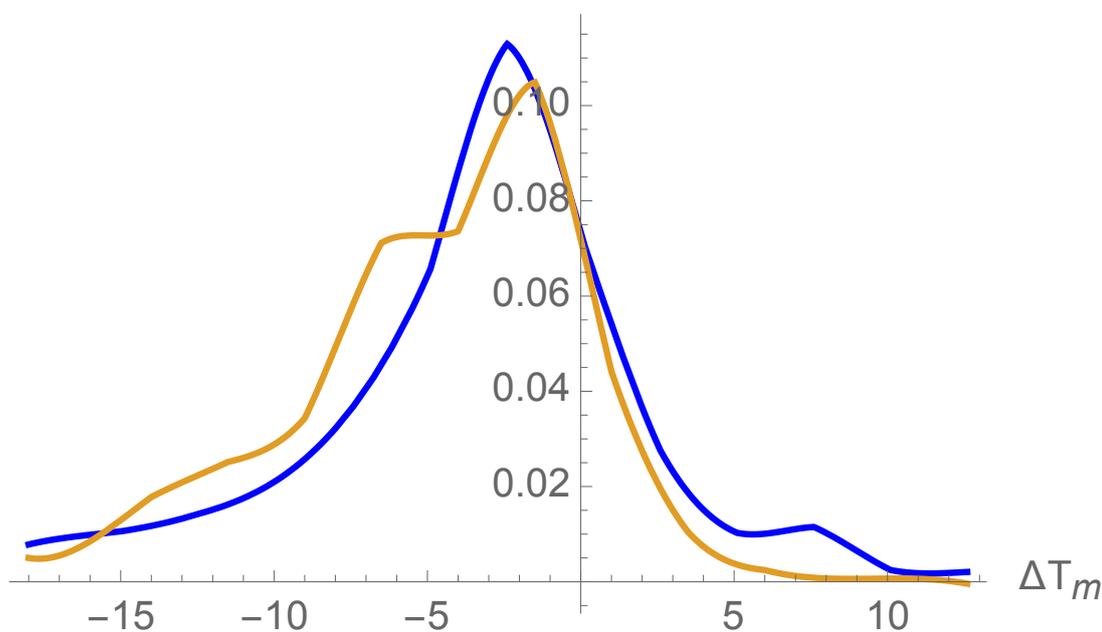

FIG. 4: Normalized ΔT_m distributions for mesostable (blue) and thermostable proteins (orange).

We also analyzed the other thermodynamic quantities reported in our dataset and first of all, the change in folding enthalpy upon point mutations. The normalized $\overline{\Delta\Delta H_m}$ distribution (see equation (9)) is plotted in Figure 5a; its mean value is $\langle \overline{\Delta\Delta H_m} \rangle = 7.3$ kcal/mol for the set of 993 entries for which this quantity has been measured experimentally. Hence, the mutations are on the average enthalpically destabilizing at T_m. Figure 5b shows the normalized distribution of the change in entropy upon mutation

$\overline{\Delta\Delta}S_m = \Delta S^{mut}(T_m^{mut}) - \Delta S^{wild}(T_m^{wild})$. The mean value of the distribution is found to be $\langle\overline{\Delta\Delta}S_m\rangle = 0.04$ kcal/(mol K).

Finally, we plotted the normalized distribution of $\Delta\Delta C_P$ and $\Delta\Delta G$ in Figures 5c-d. The mean values of these two distributions are positive: $\langle\Delta\Delta C_p\rangle = 0.08$ kcal/(mol K) for the set of 250 entries for which this value has been measured, and $\langle\Delta\Delta G\rangle = 0.89$ kcal/mol for 1,147 entries. Hence, the majority of mutant proteins have a less negative ΔC_P and a less negative ΔG than wild type proteins; the mutant proteins are thus on the average less thermodynamically stable than the wild type. Note that the asymmetric nature of the $\Delta\Delta G$ distribution is likely to cause biases in the prediction methods that use these data as learning set [13].

In summary, the large majority of the mutations are thermodynamically destabilizing (as measured by positive $\Delta\Delta G$) and thermally destabilizing (as measured by negative ΔT_m).
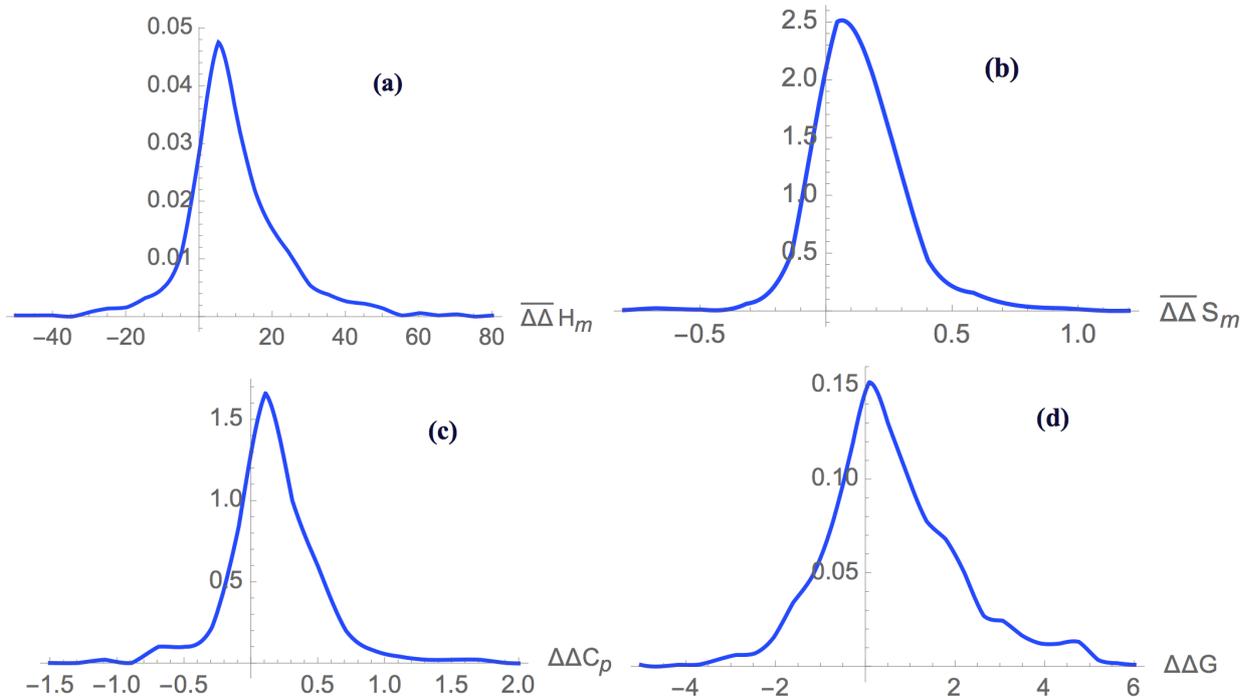

FIG. 5: Normalized distributions for : (a) $\overline{\Delta\Delta}H_m$ (in kcal/mol) and (b) $\overline{\Delta\Delta}S_m$ (in kcal/(mol K)) obtained from 993 mutations, (c) $\Delta\Delta C_p$ (in kcal/(mol K)) obtained from 250 mutations, and (d) $\Delta\Delta G$ (in kcal/mol) obtained from 1,147 mutations.

The next point we investigated is the correlation between the thermodynamic stability descriptor $\Delta\Delta G(T_r)$ and the thermal stability descriptor ΔT_m. Indeed, these two quantities

are often taken as equivalent stability measures even though this assumption is based on an approximation, as shown in equation (6). Nevertheless, this hypothesis seems *a priori* not totally unjustified, as the linear anticorrelation between the two quantities is in general quite good. In our dataset, the Pearson correlation coefficient r, computed on the 1,147 mutations for which both $\Delta\Delta G(T_r)$ and ΔT_m are available independently from the choice of T_r is equal to :

$$r_{[\Delta\Delta G(T_r),\Delta T_m]} = -0.82.$$

We must however notice that the temperature at which the $\Delta\Delta G$ measurements were performed is not always the same (as described in the "Dataset design" subsection of Methods); it is usually either 25°C or the T_m of the wild type protein.

For the subset of 461 mutations for which $\Delta\Delta G$ has been measured or computed at T_m^{wild}, the linear anti-correlation is close to perfect:

$$r_{[\Delta\Delta G(T_m),\Delta T_m]} = -0.94.$$

The anticorrelation does not reach -1 since the proportionality coefficient between $\Delta\Delta G(T_m)$ and ΔT_m is protein-dependent (see equations (6)-(7)). It is illustrated in Figure 6b.

In contrast, for the 449 mutations for which $\Delta\Delta G(25°C)$ has been directly measured or for which all thermodynamic quantities that allow using the full equation (5) have been measured (entries without subscript in the dataset), the anticorrelation is much lower:

$$r_{[\Delta\Delta G(25°C),\Delta T_m]} = -0.68, \tag{10}$$

as shown in Figure 6a. We would like to stress the value of this $\Delta\Delta G(25°C)$-ΔT_m anticorrelation coefficient can be expected to be close to the real one, as it has not been artificially improved by adding computed $\Delta\Delta G$'s that presuppose this anticorrelation.

For some entries, the two descriptors $\Delta\Delta G(25°C)$ and ΔT_m are correlated rather than anticorrelated. These signal interesting mutations that stabilize the protein thermally while destabilizing it thermodynamically at room temperature, or conversely, destabilize it thermally while stabilizing it thermodynamically. As an example of such an unusual behavior, we plotted in Figure 1 the full protein stability curve of the wild type human lysozyme and of the mutant R21A [14]; these are one of the entries of our dataset.
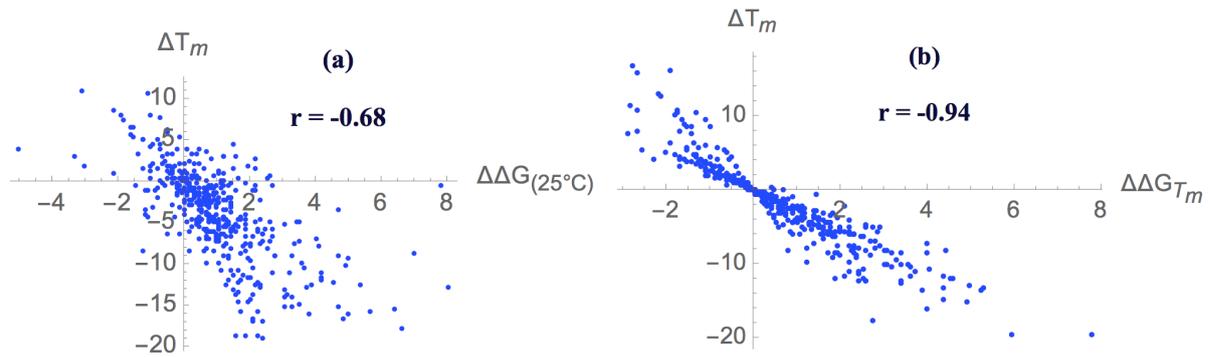
13

FIG. 6: Anticorrelation between the two descriptors of protein stability. (a) Anticorrelation between ΔT_m and $\Delta\Delta G(25°C)$ and (b) anticorrelation between ΔT_m and $\Delta\Delta G(T_m)$ where T_m is the melting temperature of the wild type protein.

Sources of experimental errors

The experimental errors on the measured thermodynamic quantities describing the folding transition have to be taken in consideration. The most noisy thermodynamic quantities are ΔC_P. Their error is generally of the order of 10-20%. Sometimes it is of the same order as $\Delta\Delta C_P$ itself, which makes the numeric evaluation of equation (5) not quite precise, even though the $\Delta\Delta C_P$ term is subleading compared to the others.

The errors on the two thermodynamic descriptors ΔH_m and T_m are in general less severe, being of the order of a few percents. These should thus not really affect the results obtained in this analysis.

Another source of error comes from the fact that the experiments are often performed in different environmental conditions in terms of pH, buffer type, ionic concentration and additives. Such errors are non negligible, even if their effect can be expected to be less important for the variation of the thermal characteristics upon mutations compared to that of the thermal characteristics themselves. Moreover, to decrease this effect, we have collected data that are as much as possible uniform in terms of environmental variables, as explained in the section "Dataset design". Note that the size of this type of error is difficult to quantify in general.

Usage notes

The dataset that we have constructed is available as a pdf file in attachment to this paper and can be downloaded as a text file at the address http://babylone.ulb.ac.be.

Acknowledgments

[1] Pucci, F., Bourgeas, R., & Rooman, M. Predicting protein thermal stability changes upon point mutations using statistical potentials: Introducing HoTMuSiC, *submitted* (2015).

[2] Haki, G.D. & Rakshit, S.K., Developments in industrially important thermostable enzymes: a review. *Bioresour Technol* **89**, 17-34 (2003).

[3] Cruins, M.E., Janssen, A.E. & Boom, R.M., Thermozymes and their applications. *Appl Biochem Biotechnol* **90**, 155-186 (2001).

[4] Becktel, W. J. & Schellman, J. A. Protein stability curves. *Biopolymers* **26**, 1859–1877 (1987).

[5] Kumar, M.D.S. *et al.*, ProTherm and ProNIT: thermodynamic databases for proteins and protein-nucleic acid interactions. *Nucl. Acids Res.* **34**, D204–D206 (2006).

[6] Di Nardo, A.A , Larson, S.M. & Davison, A.R., The relationship between conservation, thermodynamic stability, and function in the SH3 domain hydrophobic core. *J. Mol. Biol.* **333**, 641-655 (2003).

[7] Pinker, R.J. , Lin, L., Rose, G.F.& Kallenbach, N.R., Effects of alanine substitutions in alpha-helices of sperm whale myoglobin on protein stability. *Protein Science*, **2**, 1099-1105 (1993).

[8] Lin, L., Pinker, R.J. , & Kallenbach, N.R., Alpha-helix stability and the native state of myoglobin. *Protein Science*, **2**, 1099-1105 (1993).

[9] Berman, H. M. *et al.*, The Protein Data Bank. *Nucl. Acids Res.* **28**, 235–242 (2000).

[10] De Laet, M., Gilis, D. & Rooman, M. Stability strengths and weaknesses in protein structures

detected by statistical potentials. Application to bovine seminal ribonuclease. *Proteins* Epub ahead of print (2015).

[11] Bryngelson, J. D. & Wolynes, P. G., Spin glasses and the statistical mechanics of protein folding. *Proc Natl Acad Sci USA* **84**, 7524–7528 (1987).

[12] Dalkas, A. D, Teheux, F., Kwasigroch, J. M. & Rooman, M. Cation-π, amino-π, π-π, and H-bond interactions stabilize antigen-antibody interfaces. *Proteins* **82**, 1734–1746 (2014).

[13] Pucci, F., Bernaerts, K., Teheux, F., Gilis, D. & Rooman, M. Symmetry principles in optimization problems : an application to protein stability prediction. *IFAC-PapersOnLine* **48**, 458–463 (2015).

[14] Schubert, W.D. *et al.*, X-ray crystallographic and calorimetric studies of the effects of the mutation Trp59–>Tyr in ribonuclease T1. *Eur. J. Biochem.* **220**, 527–534 (1994).

Data Citations

Bibliographic information for the data records described in the manuscript.

1. Spuergin, P., Abele, U. & Schulz, G.E. *Eur. J. Biochem.*, **231**, 405-413 (1995).

2. Evrard C., Fastrez J. & Soumillion P., *FEBS Lett.*, **460**, 442-446 (1999).

3. Rose T. *et al.*, *J. Biol. Chem.*, **266**, 23654-23659 (1991).

4. Reinstein J. *et al.*, *J. Biol. Chem.*, **264**, 8107-8112 (1989).

5. D'Amico S., Gerday C. & Feller G., *J. Biol. Chem.*, **276**, 25791-25796 (2001).

6. Vogl T, Jatzke C, Hinz HJ, Benz J & Huber R, *Biochemistry*, **36**, 1657-1668 (1997).

7. Hannemann F. *et al.*, *Biochemistry*, **41**, 11008-11016 (2002).

8. Burova T.V. *et al.*, *Protein Science*, **5**, 1890-1897 (1996).

9. Bedell J.L., Edmondson S.P. & Shriver J.W., *Biochemistry*, **44**, 915-925 (2005).

10. Clark A.T. *et al.*, *Biochemistry*, **43**, 2840-2853 (2004).

11. Catanzano F. *et al.*, *Biochemistry*, **37**, 10493-10498 (1998).

12. Clark A.T. *et al.*, *Biochemistry*, **43**, 2840-2853 (2004).

13. Vu N.D., Feng H. & Bai Y. *Biochemistry*, **43**, 3346-3356 (2004).

14. Matouschek A. *et al.*, *Protein Eng.*, **7**, 1089-1095 (1994).

15. Axe D.D., Foster N.W. & Fersht A.R., *J. Mol. Biol.*, **286**, 1471-1485 (1999).

16. Kellis J.T., Nyberg K. & Fersht A.R., *Biochemistry*, **28**, 4914-4922 (1989).

17. Pinker R.J. *et al.*, *Protein Science*, **2**, 1099-1105 (1993).

18. Lin L., Pinker R.J. & Kallenbach N.R., *Biochemistry*, **32**, 12638-12643 (1993).

19. Perl D. & Schmid F.X., *J. Mol. Biol.*, **313**, 343-357 (2001).

20. Wunderlich M. & Schmid F.X., *Protein Eng.*, **19**, 355-358 (2006).

21. Wunderlich M., Martin A. & Schmid F.X., *J. Mol. Biol.*, **347**, 1063-1076 (2005).

22. Gribenko A.V. & Makhatadze G.I., *J. Mol. Biol.*, **366**, 842-856 (2007).

23. Schindler T. *et al.*, *Proteins*, **30**, 401-406 (1998).

24. Yao P. *et al.*, *Protein Eng.*, **10**, 575-581 (1997).

25. Wang Z.Q. *et al.*, *Biophys. Chem.*, **83**, 3-17 (2000).

26. Xue L.L. *et al.*, *Biochemistry*, **38**, 11961-11972 (1999).

27. Roberge M. *et al.*, *Protein Eng.*, **12**, 251-257 (1999).

28. Benito A .*et al.*, *Protein Eng.*, **15**, 887-893 (2002).

29. Tigerstrom A. *et al.*, *Biochemistry*, **43**, 12563-12574 (2004).

30. Cavallin A. *et al.*, *J. Biol. Chem.*, **275**, 1665-1672 (2000).

31. Carra J.H. & Privalov P.L., *Biochemistry*, **34**, 2034-2041 (1995).

32. Byrne M.P. & Stites W.E., *Biophys. Chem.*, **125**, 490-496 (2007).

33. Alexandrescu A.T., Hinck A.P. & Markley J.L., *Biochemistry*, **19**, 4516-4525 (1990).

34. Chen J. *et al.*, *J. Mol. Biol.*, **303**, 125-130 (2000).

35. Tanaka A., Flanagan J. & Sturtevant J.M., *Protein Science*, **2**, 567-576 (1993).

36. Byrne M.P. & Stites W.E., *Protein Science*, **4**, 2545-2558 (1995).

37. Spencer D.S. & Stites W.E.,*J. Mol. Biol.*, **257**, 497-499 (1996).

38. Carra J.H., Anderson E.A. & Privalov P.L., *Protein Science*, **3**, 952-959 (1994).

39. Leung K.W. *et al.*, *J. Biol. Chem.*, **276**, 46039-46045 (2001).

40. Eftink M.R. *et al.*, *Biochemistry*, **30**, 1193-1199 (1991).

41. Hirano S. *et al.*, *Proteins*, **58**, 271-277 (2005).

42. Koide A. *et al.*, *Biochemistry*, **40**, 10326-10333 (2001).

43. Moutiez M. *et al.*, *Protein Science*, **8**, 106-112 (1999).

44. Morii H. *et al.*, *J. Mol. Biol.*, **292**, 909-920 (1999).

45. Lee C.F. *et al.*, *J. Biol. Chem.*, **48**, 419-431 (2005).

46. Lee C.F., Makhatadze G.I. & Wong KB, *Biochemistry*, **44**, 16817-16825 (2005).

47. Sandgren M. *et al.*, *Protein Science*, **12**, 848-860 (2003).

48. Sandgren M. *et al.*, *Protein Science*, **12**, 2782-2793 (2003).

49. Greene L.H. *et al.*, *Protein Eng.*, **12**, 581-587 (1999).

50. Vanhooren A. *et al.*, *Proteins*, **60**, 118-130 (2005).

51. Oroguchi T. *et al.*,*J. Mol. Biol.*, **354**, 164-172 (2005).

52. Merz A. *et al.*, *J. Mol. Biol.*, **288**, 753-763 (1999).

53. Venkataramani R.N. *et al.*,*J. Biol. Chem.*, **277**, 48827-48833 (2002).

54. Mukaiyama A. *et al.*,*Biochemistry*, **45**, 12673-12679 (2006).

55. Dong H.A. *et al.*,*J. Mol. Biol.*, **378**, 264-272 (2008).

56. LeMaster D. & Hernandez G., *Biophysical Chemistry*, **125**, 483-489 (2007).

57. Gray R.D. & Trent J.O., *Biochemistry*, **44**, 2469-2477 (2005).

58. Nikolopoulos G. *et al.*, *Biochim et Bioph Acta*, **1774**, 772-780 (2007).

59. Beadle B.M. *et al.*, *Protein Science*, **8**, 1816-1824 (1999).

60. Mavromatis K. *et al.*, *Protein Eng.*, **16**, 497-503 (2003).

61. Gassner N.C. *et al.*, *Biochemistry*, **38**, 14451-14460 (1999).

62. Xu J. *et al.*, *Protein Science*, **7**, 158-177 (1998).

63. Lipscomb L.A. *et al.*, *Protein Science*, **7**, 765-773 (1998).

64. Eriksson A.E., Baase W.A.& Matthews B.W., *J. Mol. Biol*, **229**, 747-769 (1993).

65. Wray J.W. *et al.*, *J. Mol. Biol.*, **292**, 1111-1120 (1999).

66. Blaber M. *et al.*, *J. Mol. Biol.*, **246**, 317-330 (1999).

67. Xu J. *et al.*, *Protein Science*, **10**, 1067-1078 (2001).

68. Daopin S. *et al.*, *J. Mol. Biol.*, **221**, 647-667 (1991).

69. Mooers B.H. *et al.*, *J. Mol. Biol.*, **332**,741-756 (2003).

70. Blaber M. *et al.*, *Biochemistry*, **32**, 11363-11373 (1993).

71. Shoichet B.K. *et al.*, *Proc. Natl. Acad. Sci. USA*, **92**, 452-456 (1995).

72. Blaber M. *et al.*, *J. Mol. Biol.*, **235**, 600-624 (1994).

73. Pjura P. *et al.*, *Protein Science*, **2**, 2217-2225 (1993).

74. He M.M. *et al.*, *Protein Science*, **13**, 2716-2724 (2004).

75. Anderson D.E., Becktel W.J. & Dahlquist F.W., *Biochemistry*, **29**, 2403-2408 (1990).

76. Nicholson H. *et al.*, *Biochemistry*, **30**, 9816-9828 (1991).

77. Bell J.A. *et al.*, *Biochemistry*, **31**, 3590-2596 (1992).

78. Heinz D.W., Baase W.A. & Matthews B.W., *Proc. Natl. Acad. Sci. USA*, **89**, 3751-3755

18

(1992).

79. Lu J. *et al.*, *Biochemistry*, **31**, 7765-7772 (1992).

80. Sauer U.H., Daopin S., Matthews B.W., *J. Biol. Chem.*, **267**, 2393-2399 (1992).

81. Liu R., Baase W.A. & Matthews B.W., *J. Mol. Biol.*, **295**, 127-145 (2000).

82. Hurley J.H., Baase W.A. & Matthews B.W., *J. Mol. Biol.*, **224**, 1143-1159 (1992).

83. Anderson DE al., *Protein Science*, **2**, 1285-1290 (1993).

84. Eriksson A.E. *et al. Science*, **255**, 178-183 (1992).

85. Daopin S. *et al.*, *Biochemistry*, **30**, 11521-11529 (1991).

86. Gassner N.C., Baase W.A. & Matthews B.W., *Proc. Natl. Acad. Sci. USA*, **93**, 12155-12!58 (1996).

87. Takano K., Yamagata Y. & Yutani K, *J. Mol. Biol.*, **280**, 749-761(1998).

88. Takano K., Yamagata Y. & Yutani K, *Protein Eng.*, **16**, 5-9 (1998).

89. Takano K. *et al.*, *J. Biol. Chem.*, **278**, 31790-31795 (2003).

90. Grimsley G.R. *et al.*, *Protein Science*, **8**, 1843-1849 (1999).

91. Trevino S.R. *et al.*, *J. Mol. Biol.*, **354**, 967-978 (2005).

92. Yakolev G.I. *et al.*, *Protein Science*, **12**, 2367-2373 (2003).

93. Herbert E.J. *et al.*, *Biochemistry*, **37**, 16192-16200 (1998).

94. Takano K. *et al.*, *Eur. J. Biochem.*, **266**, 675-682 (1999).

95. Funahashi J. *et al.*, *J. Biol. Chem.*, **277**, 21792-21800 (2002).

96. Funahashi J. *et al.*, *Biochemistry*, **39**, 14448-14456 (2000).

97. Takano K. *et al.*, *Biochemistry*, **36**, 688-698 (1997).

98. Takano K. *et al.*, *Biochemistry*, **39**, 12375-12381 (2000).

99. Takano K., Yamagata Y. & Yutani K., *Biochemistry*, **40**, 4853-4858 (2001).

100. Takano K. *et al.*, *Proteins*, **45**, 274-280 (2001).

101. Yamagata Y. *et al.*, *Biochemistry*, **37**, 9355-9362 (1998).

102. Takano K. *et al.*, *Proteins*, **44**, 233-243 (2001).

103. Takano K. *et al.*, *Biochemistry*, **38**, 6623-6629 (1999).

104. Takano K., Yamagata Y., Yutani K., *Biochemistry*, **39**, 8655-8665 (2000).

105. Takano K .*et al.*, *Biochemistry*, **38**, 12698-12708 (1999).

106. Takano K.*et al.*, *J. Mol. Biol.*, **274**, 132-142 (1997).

107. Takano K. *et al.*, *Protein Eng.*, **12**, 663-672 (1999).

108. Herning T. *et al.*, *Biochemistry*, **31**, 7077-7085 (1992).

109. Funahashi J. *et al.*, *Protein Eng.*, **12**, 841-850 (1999).

110. Vu D.M. *et al.*, *Protein Science*, **10**, 2028-2036 (2001).

111. Notomista E. *et al.*, *Biochemistry*, **39**, 8711-8718 (2000).

112. Arnold U. *et al.*, *Biochemistry*, **45**, 3580-3587 (2006).

113. Choi E.J. & Mayo S.L., *Protein Eng, Des, Sel*, **19**, 285-289 (2006).

114. Merkel J.S., Sturtevant J.M. & Regan L., *Structure*, **7**, 1333-1343 (1999).

115. Nicholson E.M., Peterson R.W. & Scholtz J.M., *J. Mol. Biol.*, **321**, 355-362 (2002).

116. Cabrita L.D., Whisstock J.C. & Bottomley S.P., *Biochemistry*, **41**, 4575-4581 (2002).

117. Tew J.D. & Bottomley S.P., *J. Mol. Biol.*, **313**, 1161-1169 (2001).

118. Gilis D. *et al.*, *J. Mol. Biol.*, **325**, 581-589 (2003).

119. Greene L.H. *et al.*, *Protein Science*, **10**, 2301-2306 (2001).

120. De Vos S. *et al.*, *Biochemistry*, **40**, 10140-10149 (2001).

121. Schubert W.D. *et al.*, *Eur. J. Biochem.*, **220**, 527-534 (1994).

122. Giletto A. & Pace C.N., *Biochemistry*, **38**, 13379-13384 (1999).

123. Shirley B.A. *et al.*, *J. Biol. Chem.*, **264**, 11621-11625 (1989).

124. Herbet E.J. *et al.*, *Biochemistry*, **37**, 16192-16200 (1998).

125. Grimsley G.R. *et al.*, *Protein Science*, **8**, 1843-1849 (1998).

126. Agah S., Larson J.D. & Henzl M.T., *Biochemistry*, **42**, 10886-10895 (2003).

127. Navon A. *et al.*, *Biochemistry*, **40**, 93-104 (2001).

128. Gotte G. *et al.*, *Biochemistry*, **45**, 10795-10806 (2006).

129. Chatani E. *et al.*, *Biochemistry*, **41**, 4567-4574 (2002).

130. Kadonosono T. *et al.*, *Biochemistry*, **42**, 10651-10658 (2003).

131. Dodge R.W. & Scheraga H.A., *Biochemistry*, **35**, 1548-1559 (1996).

132. Font J. *et al.*, *Biol. Chem. Hoppe-Seyler*, **387**, 285-296 (2006).

133. Catanzano F. *et al.*, *Protein Science*, **6**, 1682-1693 (1997).

134. Johnson R.J., Lin S.R. & Raines R.T., *Protein Science*, **16**, 1609-1616 (2007).

135. Eberhardt E.S. *et al.*, *Protein Science*, **5**, 1697-1703 (1996).

136. Schultz D.A., Baldwin RL, *Protein Science*, **1**, 910-916 (1992).

137. Klink T.A. & Raines R.T., *J. Biol. Chem.*, **275**, 17463-17467 (2000).

138. Schultz L.W. *et al.*, *Protein Science*, **7**, 1620-1625 (1998).

20

139. Di Nardo A.A., Larson S.M. & Davison A.R., *J. Mol. Biol.*, **333**, 641-655 (2003).

140. Maxwell K.L. & Davison A.R., *Biochemistry*, **37**, 161721-16182 (1998).

141. Cobos E.S. *et al.*, *J. Mol. Biol.*, **328**, 221-233 (2003).

142. Varela L. *et al.*, *FEBS Letters*, **583**, 801-806 (2009).

143. Pantoliano M.W. *et al.*, *Biochemistry*, **26**, 2077-2082 (1987).

144. Pantoliano M.W. *et al.*, *Biochemistry*, **28**, 7205-7213 (1989).

145. Bryan P.N. *et al.*, *Proteins*, **1**, 326-334 (1986).

146. Zhang N. *et al.*, *Protein Eng.*, **16**, 599-605 (2003).

147. Kelley R.F., Cleary S., *Biochemistry*, **28**, 4047-4054 (1989).

148. Ahmed S.A. *et al.*, *Biochem. Biophys. Res. Comm.*, **151**, 672-678 (1988).

149. Lim W.K., Brouillette C. & Hardman J.K., *Arch. Biochem. Biophys.*, **292**, 34-41 (1992).

150. Hiraga K. & Yutani K., *Protein Eng.*, **9**, 425-431 (1996).

151. Hiraga K. & Yutani K., *Eur. J. Biochem.*, **240**, 63-70 (1996).

152. Yutani K. *et al.*, *Proteins*, **9**, 90-98 (1991).

153. Matthews C.R. *et al.*, *Biochemistry*, **19**, 1290-1293 (1980).

154. Marmorino JL, Pielak GJ, *Biochemistry*, **34**, 3140-3143 (1995).

155. Pielak G.J. *et al.*, *Biochemistry*, **34**, 3268-3276 (1995).

156. Betz S.F. & Pielak G.J., *Biochemistry*, **31**, 12337-12344 (1992).

157. Hickey D.R. *et al.*, *J. Biol. Chem.*, **266**, 11686-11694 (1991).

158. Hermann L.M. & Bowler B.E., *Protein Science*, **6**, 657-665 (1997).

159. Fetrow J.S. *et al.*, *Biochemistry*, **37**, 2477-2487 (1998).

160. Liggins J.R. *et al.*, *Protein Science*, **8**, 2645-2654 (1999).

161. Lett C.M. *et al.*, *Biochim. Biophys. Acta*, **1432**, 40-48 (1999).

162. Linske-O'Connell L.I., Sherman F. & McLendon G., *Biochemistry*, **34**, 7094-7102 (1995).

163. McGee W.A. *et al.*, *Biochemistry*, **35**, 1995-2007 (1996).

164. Tang Y. *et al.*, *J. Mol. Biol.*, **355**, 1066-1077 (2006).

165. Frank B.S. *et al.*, *Protein Science*, **11**, 680-687 (2002).

166. Kubelka J., Eaton W.A. & Hofrichter J., *J. Mol. Biol.*, **329**, 625-630 (2003).

167. Ginsburg A. *et al.*, *Protein Science*, **9**, 1085-1094 (2000).

168. Tanaka M. *et al.*, *Biochemistry*, **45**, 10351-10358 (2006).

169. Gorshkova I.N. *et al.*, *Biochemistry*, **41**, 10529-10539 (2002).

170. Ruiz-Sanz J. *et al.*, *Biochemistry*, **34**, 1695-1701 (1995).

171. Jackson S.E. *et al.*, *Biochemistry*, **32**, 11259-11269 (1993).

172. de Prat Gay G., Johnson C.M. & Fersht A.R., *Protein Eng.*, **7**, 103-108 (1994).

173. elMasry N.F. & Fersht A.R., *Protein Eng.*, **7**, 777-782 (1994).

174. Manna S.K. & Mazumdar S., *Biochemistry*, **45**, 12715-12722 (2006).

175. Dolla A. *et al.*, *Biochemistry*, **38**, 33-41 (1999).

176. Cho J.H. *et al.*, *J. Mol. Biol.*, **338**, 827-827 (2004).

177. Luisi D.L. *et al.*, *Biochemistry*, **42**, 7050-7060 (2003).

178. Iwagami S.G. *et al.*, *Protein Science*, **4**, 2562-2572 (1995).

179. Matsumara M., Becktel W.J. & Matthews B.W., *Nature*, **334**, 406-410 (1988).

180. Ladbury J.E., Hu C.Q. & Sturtevant J.M., *Biochemistry*, **31**, 10699-10702 (1992).

181. Becktel W.J. & Schellman J.A., *Biopolymers*, **26**, 1859-1877 (1987).

182. Shoichet B.KK *et al.*, *Proc. Natl. Acad. Sci. USA*, **92**, 452-456 (1995).

183. Daopin S. *et al.*, *J. Mol. Biol.*, **221**, 873-887 (1991).

184. Poteete A.R. *et al.*, *Biochemistry*, **30**, 1425-1432 (1991).

185. Chen B.L. *et al.*, *Biochemistry*, **31**, 1464-1470 (1992).

186. Zhang X., Baase W.A. & Matthews B.W., *Biophys. Chem.*, **101**, 43-56 (2002).

187. Zhang X. *et al.*, *Protein Eng.*, **8**, 1017-1022 (1995).

188. Wetzel R. *et al.*, *Proc. Natl. Acad. Sci. USA*, **85**, 401-405 (1988).

189. Nicholson H. *et al.*, *J. Mol. Biol.*, **210**, 181-193 (1989).

190. Daopin S. *et al.*, *Biochemistry*, **30**, 7142-7153 (1991).

191. Nicholson H. *et al.*, *Biopolymers*, **32**, 1431-1441 (1992).

192. Alber T. & Matthews B.W., *Methods Enzymol*, **154**, 511-533 (1992).

193. Cornish V.W. *et al.*, *Biochemistry*, **33**, 12022-12031 (1994).

194. Alber T. *et al.*, *Science*, **239**, 631-635 (1988).

195. Hawkes R., Grutter M.G. & Schellman J., *J. Mol. Biol.*, **175**, 195-212 (1984).

196. Pjura P. *et al.*, *Proteins*, **15**, 401-412 (1993).

197. Nicholson H. *et al.*, *Biochemistry*, **30**, 9816-9828 (1991).

198. Daopin S. *et al.*, *Proteins*, **15**, 401-412 (1993).

199. Zhang X., Baase W.A. & Matthews B.W., *Protein Science*, **1**, 761-776 (1992).

200. Karpusas M. *et al.*, *Proc. Natl. Acad. Sci. USA*, **86**, 8237-8241 (1989).

201. Elwell M.L. & Schellman J.A., *Biochim Biophys Acta*, **494**, 367-383 (1977).

202. Alber T. *et al.*, *Nature*, **330**, 41-46 (1987).

203. Kanaya S., Oobatake M. & Liu Y., *J. Biol. Chem.*, **27**, 32729-32736 (1996).

204. Akasako A. *et al.*, *Biochemistry*, **34**, 8115-8122 (1995).

205. Haruki M. *et al.*, *J. Biol. Chem.*, **269**, 26904-26911 (1994).

206. Akasako A. *et al.*, *J. Biol. Chem.*, **272**, 18686-18693 (1997).

207. Kimura S.A. *et al.*, *Eur. J. Biochem.*, **206**, 337-343 (1992).

208. Kimura S. *et al.*, *J. Biol. Chem.*, **267**, 21535-21542 (1992).

209. Kimura S., Kanaya S. & Nakamura H., *J. Biol. Chem.*, **267**, 22014-22017 (1992).

210. Kanaya S. *et al.*, *J Biotech*, **28**, 117-136 (1993).

211. Ishikawa K. *et al.*, *Biochemistry*, **32**, 6171-6178 (1993).

212. Haruki M. *et al.*, *Eur. J. Biochem.*, **220**, 623-631 (1994).

213. Ishikawa K. *et al.*, *Biochemistry*, **32**, 7136-7142 (1993).

214. Bolon D.N. & Mayo S.L., *Biochemistry*, **40**, 10047-10053 (2001).

215. Chakrabarti A. *et al.*, *Protein Science*, **8**, 2455-2459 (1999).

216. Ladbury J.E. *et al.*, *Biochemistry*, **34**, 2148-2152 (1995).

217. Prajapati R.S *et al.*, *Proteins*, **53**, 863-871 (2003).

218. Witarto A.B. & Sode K., *Protein Eng.*, **14**, 891-896 (2001).

219. Takahashi Y. *et al.*, *Biochemestry*, **45**, 11005-11011 (2006).

220. Hasegawa J. *et al.*, *J. Biol. Chem.*, **274**, 37533-37537 (1999).

221. Uchiyama S. *et al.*, *Protein Eng*, **15**, 455-461 (2002).

222. Santos J. *et al.*, *Biophysical Journal*, **93**, 1707-1718 (2007).

223. Ellerby L.M. *et al.*, *Biochemestry*, **29**, 5797-5806 (1990).

224. Shi P. & Kirsch J.F., *Protein Science*, **4**,2063-2072 (1995).

225. Ohmura T. *et al.*, *Protein Science*, **10**, 313-320 (1995).

226. Shi P., Holland D.R. & Kirsch J.K., *Protein Science*, **4**, 2050-2062 (1995).

227. Masumoto K. *et al.*, *Protein Eng.*, **13**, 691-695 (2000).

228. Villafranca J.E. *et al.*, *Biochemestry*, **26**, 2182-2189 (1987).

229. Gekko K. *et al.*, *J Biochem*, **116**, 34-41 (1994).

230. Castro M.J.M. & Anderson S., *Biochemistry*, **35**, 11435-11446 (1996).

231. Kim K.K. *et al.*, *Protein Science*, **2**, 588-596 (1993).

23

232. Berndt K.D. *et al.*, *Biochemistry*, **32**, 4564-4570 (1993).

233. Krowarsch D. & Otlewski J., *Protein Science*, **10**, 715-724 (2001).

234. Roy S. *et al.*, *Biochim. Biophys .Acta*, **1774**, 975-984 (2007).

235. Honda Y. *et al.*, *Biochim. Biophys. Acta*, **1429**, 365-376 (1999).

236. Xie H. *et al.*, *J. Mol. Biol.*, **360**, 157-167 (2006).

237. Heinz D.W. *et al.*, *J. Mol. Biol.*, **236**, 869-886 (1994).

238. Guo J. *et al.*, *Biochemistry*, **53**, 2855-2863 (2014).

239. Decroos C. *et al.*, *ACS Chem. Biol.* , **9**, 2157-2164 (2014).

240. Behera R.K. *et al.*, *Biochemistry*, **52**, 1373-1384 (2013).

241. Lan D. *et al.*, *Int. J. Mol. Sci.*, **16**, 7273-7288 (2015).

242. Shen T., Guo Z. & Ji C., *Acta Cryst.*, **F70**, 697-702 (2014).

243. Torpenholt S. *et al.*, *Comp. Struct. Biotech. J.*, **13**, 256-264 (2015).

244. Bian F. *et al.*, *PLoS One*, **10**, e0121108 (2015).

245. Boone C.D. *et al.*, *FEBS*, **282**, 1445-1457 (2015).

246. Porter J., Collyer C.A. & Ollis D.L., *Protein J.*, **34**, 82-89 (2015).

247. Roland B.P. *et al.*, *Biochem. Biophys. Acta*, **1852**, 61-69 (2015).

248. Verma S. *et al.*, *Biochem. Biophys. Acta*, **1844**, 1219-1230 (2014).

249. Dow B.A. *et al.*, *Arch. Biochem. Biophys.* , **550**, 20-27 (2014).

250. Tian J. *et al.*, *PLoS Comp. Biol.*, **11**, e1004207 (2015).

251. Floor R.J. *et al.*, *Chem. Bio. Chem.*, **15**, 1660-1672 (2014).

252. Daudé D. *et al.*, *Protein Science*, **22**, 1754-1765 (2013).

253. Predki P.F. *et al.*, *Nature Struc. Biol.*, **3**, 54-57 (1996).

254. Cha H.J. *et al.*, *Mol. Cells*, **36**, 39-46 (2013).

255. Narasimhan D. *et al.*, *Proteins*, **23**, 537-547 (2010).

256. Ahmad S. & Madhusudhana R., *Protein Science*, **18**, 1183-1196 (2009).

257. Grinberg A.V. & Bernhardt R., *Arch. Biochem. Biophys.*, **396**, 25-34 (2001).

258. Uhlmann H, Bernhardt R, *J. Biol. Chem.*, **270**, 29959-29966 (1995).

259. De Serrano V.S. & Castellino F.J., *Biochemistry*, **33**, 3509-3514 (1994).

260. Chang Y., Zajiceck J. & Castellino F.J., *Biochemistry*, **36**, 7652-7663 (1995).

261. De Serrano V.S. & Castellino F.J., *Biochemistry*, **31**, 3326-33335 (1992).

262. Vercillo N.C. *et al.*, *Protein Science*, **16**, 362-368 (1992).