Biochemical and Biophysical Research Communications 471 (2016) 57–62

Contents lists available at ScienceDirect

Biochemical and Biophysical Research Communications

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

# The role of stabilization centers in protein thermal stability

Csaba Magyar [a], M. Michael Gromiha [b], Zoltán Sávoly [a], István Simon [a, *]

[a] Institute of Enzymology, Research Centre for Natural Sciences, Hungarian Academy of Sciences, Magyar Tudósok krt 2, H-1117 Budapest, Hungary
[b] Department of Biotechnology, Indian Institute of Technology Madras, Chennai 600036, India

ARTICLE INFO

*Article history:*
Received 14 January 2016
Accepted 30 January 2016
Available online 2 February 2016

*Keywords:*
Protein stability
Stabilization centers
Stability changing mutations
Thermal stability
Thermophiles

ABSTRACT

The definition of stabilization centers was introduced almost two decades ago. They are centers of noncovalent long range interaction clusters, believed to have a role in maintaining the three-dimensional structure of proteins by preventing their decay due to their cooperative long range interactions. Here, this hypothesis is investigated from the viewpoint of thermal stability for the first time, using a large protein thermodynamics database. The positions of amino acids belonging to stabilization centers are correlated with available experimental thermodynamic data on protein thermal stability. Our analysis suggests that stabilization centers, especially solvent exposed ones, do contribute to the thermal stabilization of proteins.



## 1. Introduction

We introduced the concept of stabilization centers (SCs) to solve the problem raised by the fact that the three-dimensional structure of a protein cannot be calculated from amino acid sequence considering only local (short range) interactions [1,2]. The idea of stabilization centers came from the fact that the functional sites of proteins are usually composed only of a couple of residues. According to our hypothesis there might be a few additional residues which might be responsible for the stability of three-dimensional structure of proteins. The definition of stabilization centers is based on the contact map of a protein with known three-dimensional structure [1] in which residue contacts are identified with pairs of heavy atom distances. If there is at least one pair of heavy atoms with a distance less than the sum of the van der Waals radii of the two atoms plus 1.0 Å, the corresponding residues are considered to be in contact. Long-range contacts are defined as contacts between residues, which are separated by at least 10 residues or belong to different polypeptide chains. Two residues are stabilization center elements (SCEs) and form a SC together if they are involved in long-range contacts and at least one supporting residue can be found in both flanking tetra-peptides of these residues in a special way that at least seven out of the possible nine interactions are actually found between the two triplets. Stabilization centers can be identified with the SCide public server (www.enzim.hu/scide) [3]. It is important to note that a SCE can participate in more than one SC simultaneously. The role of SCs on protein thermal stability was not considered when the concept was introduced and has not yet been explored in detail. There were some efforts to reveal a possible connection between stabilization center content and proteins stability [4] and we investigated our hypothesis that some residues determine function, while others are responsible for stability and found that despite our expectations there is an overlap between residues responsible for protein function and stability [5]. However, this is the first comprehensive study until now, which deals with possible connections of stabilization centers and protein thermal stability in detail.

Long range cross-links help maintaining the native structure of proteins [6]. These cross-links can be either covalent, like disulfide bonds, or noncovalent like SCs. It is reasonable to consider that both types of crosslinks stabilize the folded structure. While covalent crosslinks do it by reducing the entropy of the unfolded state, thereby they increase the refolding rate, noncovalent crosslinks disappear during denaturation thus only slowing down this process, therefore they stabilize the folded state by reducing the unfolding rate [7]. The relationship of SCs with secondary structural elements as well as functionally important regions of proteins were already presented [1,5,8] to show a detailed picture of these clusters. Stabilization centers incorporate clusters of cooperative long-range interactions. Although the Gibbs free energy contribution of an individual noncovalent bond is generally marginal due to the

*Abbreviations:* CD, Circular Dichroism; DSC, Differential Scanning Calorimetry; PDB, Protein Data Bank; SC, Stabilization Center; SCE, Stabilization Center Element.

* Corresponding author.
*E-mail address:* simon.istvan@ttk.mta.hu (I. Simon).

http://dx.doi.org/10.1016/j.bbrc.2016.01.181

entropy loss, it can become significant when several inter-residue interactions act together [9], thus we believe that SCs might be important from the viewpoint of stability.

To check the putative role of SCs in thermal stabilization of proteins a survey was done on the ProTherm database [10]. The interplay between SCEs and residues which influence the thermal stability has been investigated with multiple approaches. First we have analyzed the relationship between melting temperature or Gibbs free energy of unfolding in water and the SCE content of wild type proteins. Subsequently, we have explored the possible correlation between protein stability change upon mutations and SCEs using thermodynamic parameters in the ProTherm database.

## 2. Material and methods

We have used our in-house computer programs for identifying the stabilization center elements. In our protocol, we have not only considered the intra-chain stabilization centers, where both SC forming residues come from the same polypeptide chain, but also inter-chain SCs. We used the transformation matrix found in the PDB files to generate the quaternary structure of the proteins.

To get an idea of the average SCE content of proteins we checked the PDBselect [11] database. We took all protein chains in the 25% list from November 2012, obtained from http://bioinfo.mni.th-mh.de/pdbselect/recent.pdb_select25.nsigma3.0 and identified intra- and interchain SCs. We investigated a total of 3096 protein chains.

The thermal stability data for proteins and mutants have been obtained from ProTherm database, which is a collection of a large number of experimental data on protein stability such as unfolding Gibbs free energy change, enthalpy change, heat capacity change and transition temperature. ProTherm is freely available at http://www.abren.net/protherm/, which has been effectively used to understand the factors influencing the stability of protein mutants and developing methods for predicting the stability upon mutations [12–14].

Two different databases were constructed from the original Protherm database. First a wild type stability database was set up to check if there is a correlation between the thermal stability and the SCE content of a protein, called the wild type database. In order to investigate broader aspects of stability two datasets were set up within this database. In the first dataset entries with data from thermal denaturation experiments, while in the second dataset entries with data from chemical denaturation experiments were included. We were using the following filtering criteria:

i) Only entries with known 3D structure of wild type proteins were taken into account.
ii) In the first wild type dataset only entries with $T_m$ melting temperature data were included. In order to get a uniform database only $T_m$ values obtained by Circular Dichroism (CD) or Differential Scanning Calorimetry (DSC) measurement methods were used. In the second wild type dataset only entries with $\Delta G\_H_2O$ (Gibbs free energy of unfolding in water) data were included.
iii) Since we are interested in stability of the native like protein structure entries with measurements under extreme conditions were discarded. Only entries with $5 \leq pH \leq 9$ experimental conditions were kept. Furthermore we used the $T_m \leq 105$ °C filtering criteria for the first dataset. In the second dataset entries with the most extreme $\Delta G\_H_2O$ values (>25.5) were discarded.

After applying these filters in many cases more than one $T_m$ value was present in the first dataset. We discarded proteins where difference between the lowest and highest $T_m$ value was greater than 10°. In the remaining cases the average of the different $T_m$ values was used. In the second dataset proteins, where differences in the Gibbs free energy of unfolding in water values were bigger than 1.5 kcal/mol, were discarded. For proteins with smaller differences average values were used.

A second database was set up to investigate if stability altering amino acid substitutions overlap with SCEs or not. This is called the mutant database. We used the following filtering criteria:

i) Only entries of single point mutants were taken into account. Wild type and multiple mutation entries were discarded. In the latter case we would not be able to distinguish, which mutation has high contribution to the cumulative effect of mutations. This is not a serious restriction since more than 80% of data in the Protherm database is for single point mutations.
ii) One of the following experimental thermodynamic data had to be available, $\Delta T_m$: melting temperature change upon mutation, $\Delta\Delta G$: change in Gibbs free energy of unfolding obtained by thermal denaturation or $\Delta\Delta G\_H_2O$: change in Gibbs free energy of unfolding in water, determined by chemical denaturation. These measures were collected in different datasets because of the different nature of these measurements. The values within one datasets can originate from different type of measurement methods. In order to get reliable and comparable data, only entries obtained by CD or DSC methods were kept for the $\Delta T_m$ and $\Delta\Delta G$ datasets and only fluorescence or CD methods in the case of $\Delta\Delta G\_H_2O$ dataset [15].
iii) Since an atomic resolution structure is needed for the identification of the stabilization centers, all remaining entries had to have an available PDB structure. Ideally this would be the structure of the mutant, but since the number of mutated protein structures in the Protherm database is very low compared to the wild type structures, the wild type structures were used in all cases.
iv) Like in the wild type case, we have considered the stability of the native like protein structures and the entries with measurements under extreme conditions were discarded.

In all three datasets negative values mean a decrease in stability, while positive ones denote increase in stability. Since we are interested in stability changes related to an amino acid position rather than the amino acid itself, we were using the maximum value for stabilizing and the minimum value for destabilizing mutations if measurements for several different mutations are available for the same position, rather than an average value. An average value would be unadvisable because of the asymmetrical nature of the database composition, there are a lot more destabilizing mutations, than stabilizing ones.

## 3. Results

We calculated the average SCE content of the protein chains from the PDBselect database. The number of stabilization center elements was 69,706, while the total number of residues was 351,385, giving an average SCE content of 19.84%, which is used as a reference value hereafter.

We were interested if there is any general correlation between the thermal stability of a protein and its stabilization center element content. First we identified the SCEs in the proteins of first dataset of the wild type database and calculated the average SCE content for all proteins. There were a total of 181 $T_m$ measurements in the range of 25 and 105 °C. We averaged the SCE content of proteins within a $T_m$ window of 20°, i.e. the data at temperature $T$ is

the average SCE value of proteins with a $T_m$ value within the $T \pm 10$ interval. Since there were only a few points in the highest temperature range, the entries with the next two highest $T_m$ value (glutamate dehydrogenase from *Thermococcus litoralis* and *Pyrococcus furiosus;* 1BVU, $T_m = 109$ °C and 1GTM, $T_m = 114$ °C) [16,17] were added to this dataset, which were originally excluded from the database. One entry (1BGL) was deleted from the dataset, because our program could not process this structure. This is not a significant problem, because this structure with $T_m$ value of 49.8 °C fall into a region, where relatively many data can be found. There were a couple of PDB structures in which no SCs could be identified. Most of these protein structures were solution NMR structures. Since the definition of SCs is very sensitive to the quality of the structure due to the hard limit of heavy atom distances, we prefer high resolution X-ray structures. We searched the PDB database for X-ray structures of the same proteins. We used the PDB clustering available at http://resources.rcsb.org/sequence/clusters/bc-100.out to identify X-ray equivalents. We managed to find an appropriate structure in the cases listed in Table 1.

SCE content was recalculated using these alternate X-ray structures. Results can be seen in Fig. 1.

It can be stated that the SCE content is increasing until 80 °C. Above this temperature, the average SCE content goes into saturation or even starts to decrease. However, standard deviations are quite large, and there is a low number of data points in the $T_m > 80$ region, this temperature range coincides well with the temperature range in which organisms are classified as hyperthermophiles. There are several reviews which indicate that thermophiles and hyperthermophiles apply different strategies (like optimizing cavities) in the optimization of their protein structures to reach increased thermal stability needed at their physiological temperature [18,19]. This finding might explain the change in the SCE content at extreme temperatures.

Gibbs free energy of unfolding in water $\Delta G\_H_2O$ values obtained by chemical denaturation measurements represent a different aspect of protein stability. The second dataset of the wild type database includes entries with $\Delta G\_H_2O$ measurements. There were 170 entries within the range of 0 and 25.5 kcal/mol. We created 8 intervals as in the previous $T_m$ wild type dataset and used a window size of 6 kcal/mol for averaging the SCE content for all proteins of the dataset within the given range. Results can be seen in Fig. 2. Since there is an overlap between entries belonging to adjacent intervals, a relatively smooth curve is expected even for an uncorrelated case. From the S-shape of the curve we can conclude that there is no correlation between the SCE content and Gibbs free energy of unfolding in water of proteins.

Thermodynamic data from the Protherm derived mutant datasets were analyzed in the following way. The $\Delta T_m$, $\Delta\Delta G$ and $\Delta\Delta G\_H_2O$ datasets were handled separately because of the different nature of the measurements. In all three cases data were represented with histograms. For every N whole number of $\Delta T_m$, $\Delta\Delta G$ and $\Delta\Delta G\_H_2O$ values we counted the number of entries within the [N−0.5,N+0.5[ intervals. For example the 0 point in Fig. 3 gives the number of mutation which have a $-0.5 \leq \Delta T_m < 0.5$ value.

We fitted the sum of two Gaussians to these data points. By means of the equation of the fitted curve, the two points of the

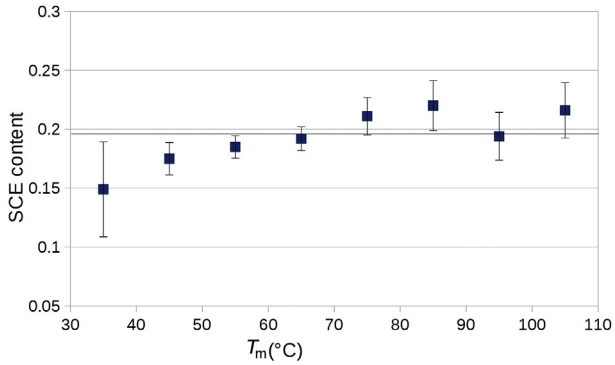


**Fig. 1.** Average stabilization center element content as the function of the melting temperature. The standard error of the mean is represented with error bars. The average SCE content obtained on the PDBselect database is shown as a horizontal line.

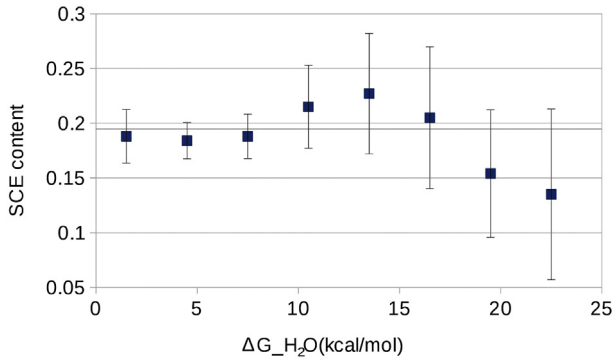


**Fig. 2.** Average stabilization center element content as the function of the Gibbs free energy of unfolding in water. The standard error of the mean is represented with error bars. The average SCE content obtained on the PDBselect database is shown as a horizontal line.

**Table 1**
PDB codes of the replaced NMR structures and their X-ray equivalents.

| PDB code of replaced NMR structures | PDB code of alternate X-ray structures |
|---|---|
| 1A23 | 1FVK |
| 1A2I | 2CTH |
| 1ARR | 1BDT |
| 1BTA | 1AY7 |
| 1BUY | 1CN4 |
| 1CEY | 1CHN |
| 1DVC | 1NB5 |
| 1EZA | 1ZYM |
| 1HUE | 1HUU |
| 1QQV | 2RJY |
| 1SAP | 1AZP |
| 1URK | 3K24 |
| 1Y90 | 4OJ1 |
| 2AIT | 1HOE |

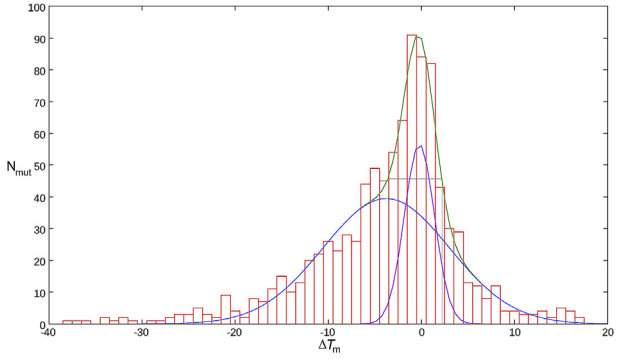


**Fig. 3.** Distribution of mutations with different $\Delta T_m$ values. The distribution is approximated with the sum (green) of two Gaussian curves (blue and purple). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

curve at the half maximum were determined and rounded to one decimal precision. These two points defined 3 regions (destabilizing, neutral and stabilizing), which were different for the three datasets.

The ratio of stabilization center elements was calculated for all the protein structures in a given dataset. Instead of averaging over all protein structures in a dataset, the average value was defined as the sum of the stabilization centers elements divided by the sum of the residues in all protein structures. The ratio of the observed ($N_{obs}$) and expected ($N_{exp}$) overlaps between SCEs and mutations can be seen in Equation (1), where $N_{sce}$ is the number of SCEs; $N_{mut}$ is the number of positions, where mutations occur; $N_{SCE+mut}$ is the overlap, i.e. the number of mutated positions, where the wild type residue is a SCE and $N_{tot}$ is the total number of residues in a protein. This equals to the quotient of the ratio of SCEs under the mutated positions and the ratio of SCEs under all residues (i.e. the average SCE content).

$$\frac{N_{obs}}{N_{\exp}} = \frac{N_{sce+mut}}{\frac{N_{sce}}{N_{tot}}\frac{N_{mut}}{N_{tot}}N_{tot}} = \frac{N_{sce+mut}/N_{mut}}{\frac{N_{sce}}{N_{tot}}} \quad (1)$$

*The ratio of the observed and expected number of SCEs at mutated positions.*

The average stabilization center element content was 20.10%, 19.50% and 20.50% for the $\Delta T_m$, $\Delta\Delta G$ and $\Delta\Delta G\_H_2O$ datasets, respectively. We identified the stabilization center elements in the wild type PDB structures assigned to the mutations. In every entry we checked if the position of the mutation is a stabilization center element or not. We calculated the number of overlaps of mutations with stabilization center elements for all the three (destabilizing, neutral and stabilizing) regions in the different datasets. Results can be seen in Table 2.

Mutations in Range2 are considered neutral in terms of thermal stability. Mutations in Range1 decrease stability and mutations in Range3 increase thermal stability. According to the results, it can be stated that for all the three measurements, in the destabilizing range (1) the ratio of stabilization centers among the mutated positions is higher than the average SCE content, thus there is an overlap between the SCEs and the positions of mutations. In the neutral range (2) we can see segregation of SCEs and mutations in all three datasets. In the stabilizing range (3) only in the $\Delta\Delta G$ dataset is there a significant overlap. Since this dataset has the lowest number of data, results for this range do not provide a strong evidence. The results for the first two ranges, that SCE overlap with destabilizing mutation and segregate from neutral mutation positions strengthen our hypothesis that SCs are important in maintaining protein stability, mutations of SCE residues usually destabilize the structure. The result for Range3 that the mutation of SCEs which already take part in the stabilization of a proteins structure do not improve stability further is not surprising either.

**Table 2**
Number of overlaps between SCEs and positions of mutations within the specified range and the ratio of observed over expected stabilization center element content within the same dataset.

| | Range1 | Range2 | Range3 |
|---|---|---|---|
| $\Delta T_m$ | $\Delta T_m < -3.6$ | $-3.6 \leq \Delta T_m \leq 2.3$ | $2.3 < \Delta T_m$ |
| | 92/391: 23.53% | 47/374: 12.57% | 31/141: 21.99% |
| | 1.171 | 0.625 | 1.094 |
| $\Delta\Delta G$ | $\Delta\Delta G < -1.8$ | $-1.8 \leq \Delta\Delta G \leq 1$ | $1 < \Delta\Delta G$ |
| | 55/176: 31.25% | 44/313: 14.06% | 23/93: 24.73% |
| | 1.603 | 0.721 | 1.268 |
| $\Delta\Delta G\_H_2O$ | $\Delta\Delta G\_H_2O < -3.2$ | $-3.2 \leq \Delta\Delta G\_H_2O \leq 1$ | $1 < \Delta\Delta G\_H_2O$ |
| | 76/310: 24.52% | 132/864: 15.28% | 21/156: 13.46% |
| | 1.196 | 0.745 | 0.657 |

Since the geometric definition of stabilization centers might prefer buried residues over solvent exposed ones, we investigated if this might play any role in our findings. First we investigated if SC elements are biased to favor buried positions. Solvent accessibility was calculated using the Naccess 2.1.1 software [20]. Before solvent accessibility calculation the quaternary structure of all proteins were created based on the BIOMT transformation matrices. Based on the relative surface area (RSA) all protein residues were divided into three categories:

i) Buried residues (RSA < 0.1)
ii) Partially buried residues (0.1 < RSA ≤ 0.5)
iii) Exposed residues (0.5 < RSA)

In the PDBselect database 31.9% of all residues were classified as buried, 40.97% were partially exposed and 27.13% were exposed. We investigated the distribution of SC elements among these three categories, as well. In the PDBselect database 52.87% of SC elements were buried, 39.2% were partially buried and only 7.93% were exposed, thus SC elements are more frequent among buried residues. Further we found that 33.24% of buried residues are SCEs, while in the partially buried and exposed categories this ratio is 19.19% and 5.86%, respectively. Compared to the average SCE content of about 19% we can see that buried residues are really overrepresented among SCEs. This is probably because of the geometric criterion in the SC definition which can be much easier fulfilled in the inner part of a protein structure than on the surface. Taking this

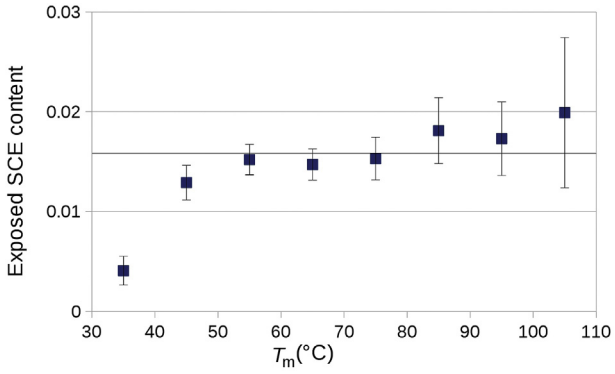


**Fig. 4.** Average exposed stabilization center element content as the function of the melting temperature. The standard error of the mean is represented with error bars. The average SCE content obtained on the PDBselect database is shown as a horizontal line.

**Table 3**
Mutations, where removal of a SCE causes decrease in stability.

| PDB code of wild type | PDB code of mutant | Mutation | Difference in $\Delta T_m$ or $\Delta\Delta G$ |
|---|---|---|---|
| 1N0J | 1VAR | I58T | −13.6 °C ($\Delta T_m$); −2.7 kcal/mol ($\Delta\Delta G$) |
| 2LZM | 1QS5 | A98L | −13.9 °C ($\Delta T_m$) |
| 1BNI | 1BNS | T26A | −2.14 kcal/mol ($\Delta\Delta G$) |

observation into account exposed SCEs might be even more important in stabilizing protein structures. We analyzed the largest subset of the Protherm database, the destabilizing mutation of Range1 of the $\Delta T_m$ dataset. We found in this dataset the strongest correlation between SCE content and stability change. We calculated how many residues of the three classes are SCEs in this dataset: 33.71% and 17.91% of buried and partially buried residues were SCEs, respectively. This ratios are almost the same as the values obtained for the PDBselect database. However for exposed residues in this dataset 14.04% of the residues were SCEs, while for the PDBselect database the corresponding value is only 5.86%. This finding underlines the role of exposed SCs in the thermal stabilization of protein structures. In virtue of this result we checked if solvent exposed SCE content shows any correlation with thermal stability using the wild type database. The result can be seen in Fig. 4, where the exposed SCE content is plotted as function of melting temperature. As previously seen in Fig. 1, we can see that protein structures with higher thermal stability contain more solvent exposed SCE residues.

Although the number of 3D structures for mutated proteins is low in the Protherm database, we managed to find a couple of examples, where the effects of a mutation at an SCE position can be verified by the structure of the mutated protein. In Table 3 [21–23] we can see a couple of destabilizing mutations, where the wild type residue is a SCE, but in the mutated structure the corresponding residue is not a SCE, any more. We even found a special case, where a stabilizing mutation resulted in the formation of a SC, which was not present in the wild type structure. It is the *Staphylococcus aureus* nuclease; 1STN, where the proline 117 residue was mutated [24–26]. 1SYE, P117T, $\Delta T_m = +0.6$ °C [24] and $\Delta\Delta G\_H_2O = +1.10$ kcal/mol [25]; 1SNP, P117G, $\Delta T_m = +5$ °C [24] and $\Delta\Delta G\_H_2O = +1.60$ kcal/mol [25]; 1SYG, P117A, $\Delta\Delta G\_H_2O = +0.8$ kcal/mol [26].

## 4. Discussion

Stabilization centers are thought to have a role in “maintenance” of the three-dimensional structure of proteins by preventing their decay due to their cooperative long range interactions. Our results coincide well into this picture that proteins with higher melting temperature have a higher average number of SCEs, but there is no correlation between Gibbs free energy of unfolding in water and SCE content. Mutations which do not influence protein stability are not overlapping with SCs, while mutations which change stability in either way are more frequently found at SC residues. The investigation of the wild type protein structures showed that structures with higher denaturation temperatures up to 85 °C have a higher number of SCEs. However, the trend is not true in the temperature range of hyperthermophilic proteins. This seems even more clearly in the case of solvent exposed SCs, where the trend is even true for hyperthermophilic proteins. These findings strengthen our hypothesis that SCs are important for the maintenance of a stable protein structure. Thus we can conclude that stabilization centers are deeply involved in thermal stabilization of protein structures.

## Acknowledgment

This work was supported by OTKA grants NK100482 and K115698. C. Magyar was supported by the János Bolyai Research Fellowship.

## Transparency document

Transparency document related to this article can be found online at http://dx.doi.org/10.1016/j.bbrc.2016.01.181.

## References

[1] Z. Dosztányi, A. Fiser, I. Simon, Stabilization centers in proteins, J. Mol. Biol. 272 (1997) 597–612.
[2] I. Simon, L. Glasser, H.A. Scheraga, Calculation of protein conformation as an assembly of stable overlapping segments, Proc. Natl. Acad. Sci. 88 (1991) 3661–3665.
[3] Z. Dosztányi, C. Magyar, G.E. Tusnady, I. Simon, SCide: identification of stabilization centers in proteins, Bioinformatics 19 (2003) 899–900.
[4] E. Tudos, A. Fiser, A. Simon, Z. Dosztanyi, M. Fuxreiter, C. Magyar, I. Simon, Noncovalent cross-links in context with other structural and functional elements of proteins, J. Chem. Inf. Comput. Sci. 44 (2004) 347–351.
[5] C. Magyar, E. Tudos, I. Simon, Functionally and structurally relevant residues of enzymes, FEBS Lett. 567 (2004) 239–242.
[6] D. Sengupta, S. Kundu, Role of long- and short-range hydrophobic, hydrophilic and charged residues contact network in protein's structural organization, BMC Bioinforma. 13 (2012) 142, http://dx.doi.org/10.1186/1471-2105-13-142.
[7] Z. Gugolya, Z. Dosztanyi, I. Simon, Interresidue interactions in protein classes, Proteins 27 (1997) 360–366.
[8] D. Kihara, The effect of long-range interactions on the secondary structure formation of proteins, Protein Sci. 14 (2005) 1955–1963.
[9] P.K. Ponnuswamy, M.M. Gromiha, On the conformational stability of folded proteins, J. Theor. Biol. 166 (1994) 63–74.
[10] M.M. Gromiha, J. An, H. Kono, M. Oobatake, H. Uedaira, A. Sarai, Protherm: thermodynamic database for proteins and mutants, Nucl. Acids Res. 27 (1999) 286–288.
[11] S. Griep, U. Hobohm, PDBselect 1992-2009 and PDBfilter-select, Nucl. Acids Res. 38 (2010) D318–D319.
[12] M.D. Kumar, K.A. Bava, M.M. Gromiha, P. Prabakaran, K. Kitajima, H. Uedaira, A. Sarai, ProTherm and ProNIT: thermodynamic databases for proteins and protein-nucleic acid interactions, Nucl. Acids Res. 34 (2006) D204–D206.
[13] Y. Dehouck, J.M. Kwasigroch, D. Gilis, M. Rooman, PoPMuSiC 2.1: a web server for the estimation of protein stability changes upon mutation and sequence optimality, BMC Bioinforma. 12 (2011) 151, http://dx.doi.org/10.1186/1471-2105-12-151.
[14] L.T. Huang, M.M. Gromiha, S.Y. Ho, iPTREE-STAB: interpretable decision tree based method for predicting protein stability changes upon mutations, Bioinformatics 23 (2007) 1292–1293.
[15] M. Masso, I.I. Vaisman, Accurate prediction of stability changes in protein mutants by combining machine learning with structure based computational mutagenesis, Bioinformatics 24 (2008) 2002–2009.
[16] C. Vetriani, D.L. Maeder, N. Tolliday, K.S. Yip, T.J. Stillman, K.L. Britton, D.W. Rice, H.H. Klump, F.T. Robb, Protein thermostability above 100 degreesC: a key role for ionic interactions, Proc. Natl. Acad. Sci. 21 (1998) 12300–12305.
[17] A.C. Sen, M.T. Walsh, B. Chakrabarti, An insight into domain structures and thermal stability of gamma-crystallins, J. Biol. Chem. 267 (1992) 11898–11907.
[18] A. Szilagyi, P. Zavodszky, Structural differences between mesophilic, moderately thermophilic and extremely thermophilic protein subunits: results of a comprehensive survey, Structure 8 (2000) 493–504.
[19] M.M. Gromiha, M.C. Pathak, K. Saraboji, E.A. Ortlund, E.A. Gaucher, Hydrophobic environment is a key factor for the stability of thermophilic proteins, Proteins 81 (2013) 715–721.
[20] S.J. Hubbard, S.F. Campbell, J.M. Thornton, Molecular recognition. Conformational analysis of limited proteolytic sites and serine proteinase protein inhibitors, J. Mol. Biol. 220 (1991) 507–530.
[21] K. Takano, Y. Yamagata, K. Yutani, A general rule for the relationship between hydrophobic effect and conformational stability of a protein: stability and

structure of a series of hydrophobic mutants of human lysozyme, J. Mol. Biol. 280 (1998) 749–761.
[22] R. Liu, W.A. Baase, B.W. Matthews, The introduction of strain and its effects on the structure and stability of T4 lysozyme, J. Mol. Biol. 295 (2000) 127–145.
[23] A. Matouschek, J.T. Kellis, L. Serrano, A.R. Fersht, Mapping the transition state and pathway of protein folding by protein engineering, Nature 340 (1989) 122–126.
[24] M.R. Eftink, C.A. Ghiron, R.A. Kautz, R.O. Fox, Fluorescence and conformational stability studies of Staphylococcus nuclease and its mutants, including the less stable nuclease-concanavalin A hybrids, Biochemistry 30 (1991) 1193–1199.
[25] T. Nakano, L.C. Antonino, R.O. Fox, A.L. Fink, Effect of proline mutations on the stability and kinetics of folding of staphylococcal nuclease, Biochemistry 32 (1993) 2534–2541.
[26] S.M. Green, A.K. Meeker, D. Shortle, Contributions of the polar, uncharged amino acids to the stability of staphylococcal nuclease: evidence for mutational effects on the free energy of the denatured state, Biochemistry 31 (1992) 5717–5728.