Noncovalent Cross-links in Context with Other Structural and Functional Elements of Proteins†

Éva Tüdős,‡ András Fiser,§ Ágnes Simon,|| Zsuzsanna Dosztányi,‡ Mónika Fuxreiter,‡ Csaba Magyar,‡ and István Simon*,‡

Institute of Enzymology, Biological Research Center, Hungarian Academy of Sciences, P.O. Box 7, H-1518 Budapest, Hungary, Department of Biochemistry and Seaver Foundation Center for Bioinformatics, Albert Einstein College of Medicine, 1300 Morris Park Avenue, Bronx, New York 10461, and Department of Neurochemistry, Chemical Institute, Chemical Research Center, Hungarian Academy of Sciences, Pusztaszeri út 59-67, H-1025 Budapest, Hungary

Received July 11, 2003

Proteins are heteropolymers with evolutionary selected native sequences of residues. These native sequences code for unique and stable 3D structures indispensable for biochemical activity and for proteolysis resistance, the latter which guarantees an appropriate lifetime for the protein in the protease rich cellular environment. Cross-links between residues close in space but far in the primary structure are required to maintain the folded structure of proteins. Some of these cross-links are covalent, most frequently disulfide bonds, but the majority of the cross-links are sets of cooperative noncovalent long-range interactions. In this paper we focus on special clusters of noncovalent long-range interactions: the Stabilization Centers (SCs). The relation between the SCs and secondary structural elements as well as the relation between SCs and functionally important regions of proteins are presented to show a detailed picture of these clusters, which are believed to be primarily responsible for major aspects of protein stability.

INTRODUCTION

Proteins have unique three-dimensional structures under physiological conditions. In contrast, small oligopeptides that correspond to short segments of the polypeptide chain of a protein exhibit a large variety of conformations in solution.[1−3] Furthermore, oligopeptides up to 10 residues with the same amino acid sequence can be found in different conformations in different proteins.[4] Thus, interactions within a short polypeptide segment are not sufficient to determine its unique conformation when it is part of a protein chain. In principle, the number of possible conformations of each segment that need to be considered in the formation of a unique protein structure could be decreased dramatically by taking into account that the common parts of the overlapping segments must appear in the same conformation.[5] However, it has been shown that a systematic build up of the bovine pancreatic trypsin inhibitor from low-energy conformations of overlapping segments still does not result in a unique protein conformation.[3] Baker et al. also found that the energetic preference for a certain local conformation is generally weak, and the assembly of short fragments consistent with the local preferences must be completed with nonlocal constraints.[6]

In globular proteins the amino acid residues form a large number of contacts with each other, which makes protein structures as dense as some liquid crystals.[7] However, these interresidue interactions are not distributed equally within a structure. It has been suggested that a few key residues are enough to ensure the fold specificity or to act as a folding nucleation site. In a recent analysis "small-world-like" features were found in proteins, where a limited number of residues formed a large network of interactions, while most other residues participated in only a smaller number of interactions.[8] Clusters of cooperative interactions may be important from the viewpoint of stability. Although in most cases, the free energy contribution of an individual noncovalent bond is marginal due to the entropy loss of loop closure in long-range interactions, it can become significant when several interresidue interactions act together.[9]

Cooperative long-range interactions inhibit local unfolding and therefore stabilize the native state by decreasing the rate of spontaneous unfolding due to thermal fluctuation under physiological conditions.[10] We have demonstrated that the number of long-range interactions in proteins, which vary in a wide range depending on the secondary structure composition, is independent from the number and distribution of disulfide bonds.[11] Due to the high glutathione concentration inside the cell, there is no disulfide bond in intracellular proteins.[12] Cys is one of the less abundant residues, the number of disulfide bonds is rather small even in most of the extracellular proteins although they have a critical role in contributing to the fold stability of small proteins lacking a true hydrophobic core. Thus, many of the cross-links stabilizing the 3D structure of proteins are clusters of noncovalent bonds. This is the reason, why clusters of interacting residues were studied intensively in the past decade.[13,14] We focused on the cooperative feature of these interactions when we introduced the concept of Stabilization Center (SC) with the following definition:

* Corresponding author phone: (36-1) 466-9276; fax: (36-1) 466-5465; e-mail: simon@enzim.hu.

† Dedicated to Prof. George W. A. Milne in gratitude for his long-term service in publishing the papers of the Math-Chem-Comp Conferences.

‡ Biological Research Center, Hungarian Academy of Sciences.

§ Albert Einstein College of Medicine.

|| Chemical Research Center, Hungarian Academy of Sciences.

10.1021/ci030409i CCC: $27.50
Published on Web 02/12/2004

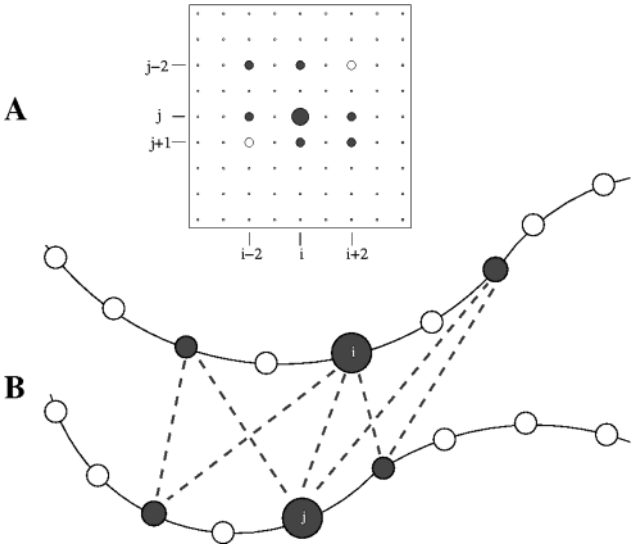

Figure 1. Definition of stabilization centers (SCs). A: The contact map around the $i = 285$, $j = 129$ residue pair of the *Leptoxyphium fumago* chloroperoxidase (PDB-id: 1cpo) is pictured. The larger circle, [*i*,*j*] marks the SC pair. In this example $i-2$, $i+2$, $j-2$, and $j+1$ are the supporting residues in the flanking tetrapeptides. Shaded circles indicate long-range residue contacts. B: Schematic representation of the interresidue interactions participating in the formation of the stabilization center.

Two residues are in long-range interaction (i) if the distance between two non-hydrogen atoms of the two residues is less than the sum of the two van der Waals radii plus 1 Å and (ii) the two interacting residues are separated by at least 10 other residues in the amino acid sequence. Two residues are SC elements if they are involved in long-range interactions and at least one supporting residue can be found in each of the flanking tetrapeptides of these residues in such a way that at least 7 out of the possible 9 interactions are formed between the two triplets [Figure 1].[15]

The properties of SC forming residues were analyzed on a data set containing 80 unrelated protein structures.[15] It was found that hydrophobic residues are dominant in SCs, while Gly, Pro, charged, and polar residues are rather rare. Interestingly, among the 210 types of interacting pairs in SCs the four kinds of salt bridges DK, DR, EK, and ER appear more often than any particular hydrophobic residue pairs. This could be a consequence of the specificity of interactions, because all hydrophobic residues can form interactions with any hydrophobic residues without any special pairing among them. In general, SC residues have small surface accessibility and low flexibility as reflected by an average 20% lower B factor in the crystal structures. A comparison of the interatomic interactions showed that SC residues form 1.77 times more long-range atomic contacts than other residues involved in long-range interactions. SC residues have been found to have higher sequential and structural conservation than an average residue. The average sequence environments of residues involved and not involved in SC formation were found to be different. This observation served as a basis to develop a neural network based algorithm, which predicts SC residues from the sequence of a protein with 65% accuracy, or by using multiple sequence profiles of homologous proteins with up to 68% accuracy. Public servers are available to identify SC residues from the 3D structure, http://www.enzim.hu/scide, or to predict SC residues from amino acid sequences, http://www.enzim.hu/scpred.[15,16]

Table 1. Accuracy of Secondary Structure Prediction Methods on Stabilization Center Elements, 3271 Residues, and on the Same Size of Subsets of Randomly Selected Residues as References, in 80 Unrelated Proteins Studied

	ALB	DPM	GOR	CFA	
helices: SC	74.19%	67.62%	64.47%	64.45%	
helices: reference	70.28%	67.18%	65.41%	65.70%	
sheets: SC	57.69%	64.18%	62.97%	60.01%	
sheets: reference	58.96%	65.50%	65.24%	63.39%	
turns: SC	58.83%	53.03%	60.62%		
turns: reference	59.33%	53.33%	57.93%	turns+coils	60.66%
coils: SC	49.35%	61.32%	54.85%		
coils: reference	53.92%	60.53%	54.69%		

In this paper we focus on the special clusters of noncovalent long-range interactions: stabilization centers. The relation between the SCs and secondary structural elements as well as the relation between SCs and functionally important regions of proteins are presented. Combining our earlier data with new ones, we provide a view of a special cluster, which is believed to be responsible for major aspects of protein stability.

RELATION BETWEEN STABILIZATION CENTER ELEMENTS AND SECONDARY STRUCTURES

SCs are clusters of interactions connecting residues close in the tertiary structure but far in the sequence i.e., connecting different secondary structure elements. The relations between stabilization centers and secondary structure elements have been analyzed in two directions: (i) the influence of SCs on the formation of secondary structure elements and (ii) the secondary structure preference of SCs. The first question was studied by comparing the efficiencies of various secondary structure prediction methods. Prediction methods were applied to predict the secondary structural state of residues from the local sequence, without considering any interaction between the residues and the rest of the polypeptide chain. If long-range interactions were significantly influencing the formation of the various secondary structure elements, more accurate predictions would be expected for those residues, which are involved in less long-range interactions. A study showed that this is not the case.[17] Four methods, ALB,[18] CFA,[19] DPM,[20] and GOR,[21] worked with similar accuracy on SCs than on any other residues [Table 1]. Apparently, SCs cross-link secondary structure elements to each other, but in most of the cases they do not seem to override the local preferences of a certain secondary structure. Considering the other side of the coin, the frequency of occurrences of SC residues in various secondary structure elements and in proteins from different fold classes (all-α, all-β, α/β, α+β and others) were investigated using a data set of 527 unrelated polypeptide chains containing 140 151 residues. The distribution of SCs in the various secondary structure elements and in different fold classes strongly depends on the dominant secondary structure type of proteins[22] (Tables 2 and 3). The coil structure is a favored place for the formation of SCs because regions in the nonregular structure are more flexible, hence they can more easily accommodate a large number of interresidue interactions.

SCs were the least frequent in all-α proteins suggesting that this type of protein might use other means of stabilization

Table 2. Secondary Structure Propensity of All Residues in 527 Unrelated Proteins and the Subset of Stabilization Center Residues

	whole data set	SC elements
helices	30.70%	14.06%
extended	21.52%	51.91%
turns	11.97%	5.24%
coils	35.81%	28.80%

Table 3. Percentage of Stabilization Center Residues in Various Secondary Structural Classes of Proteins

secondary structural class	number of proteins	number of residues	number and percentage of SC elements
all-α	77	16 078	2127 (13.23%)
all-β	98	20 761	6923 (33.35%)
α/β	141	44 299	10646 (24.03%)
α+β	91	18 739	4605 (24.57%)
other	120	40 274	10021 (24.88%)
total:	527	140 151	34322 (24.49%)

as well. Dipole−dipole interactions between α-helices are considered to provide a major contribution to the structural stability of all-α proteins.[23] In the case of four-helix bundle proteins, the analysis of SCs connecting α-helices with different relative angles helped to estimate the energetic contribution of an average SC to the stabilization of the protein. Assuming that the relative angular position of the helices in four helix bundles follows the Boltzmann distribution, it was shown that the presence of SCs decrease the energy difference between the parallel and antiparallel states from 1.1 kcal/mol to 0.4 kcal/mol. It illustrates that SCs not only influence the unfolding rate but, although only in a small extent, also SCs can contribute to the thermodynamic stability of proteins as well.[24]

RELATIONS BETWEEN STABILIZATION CENTERS AND FUNCTIONALLY IMPORTANT RESIDUES

Kinetic stability is a crucial issue for proteins that perform their function in a different cellular location than where their folding takes place. Protein must maintain their native conformation in their acting place to accomplish their biological function and to avoid enzymatic degradation. Meanwhile the lifetime of proteins must be limited, since constitutive proteolysis is the main source of amino acids for de novo protein synthesis. A simple means of regulation, which makes a useful form of a protein stable and the useless form unstable, was uncovered when SCs of the major histocompatibility complex (MHC) proteins were analyzed. MHC proteins are single-use peptide receptors and transporters that carry proteolytic degradation products, peptides of various sizes, to the cell surface for presenting them to T cells. It is known that MHC−peptide complexes are stable for several hours or even days[25] and that empty MHC class II molecules are found only in association with chaperones on the surface of professional antigen presenting cells.[26] These proteins fulfill their function as long as they bind their dissociable ligand, the peptide. The noncovalently bonded peptide ligand dissociates from the protein with a rate that depends on the activation free energy of dissociation. The ligand-free MHC molecules on the cell surface are practically useless for their primary biological function, but may become an important source of amino acids as soon as they lose their compact stable structure resisting proteolytic attacks. By analyzing the SCs of 31 unrelated MHC proteins with various peptide ligands, we recognized a function-related regulation of the stability of MHC proteins. Stabilization centers composed from residues of the ligand peptides and from the MHC proteins were found in crucial positions. The ligand-binding domain of MHC protein is built up from an eight-stranded β-sheet with two mostly α-helical regions on the top of the β-sheet. The extended chains of β-sheets of this protein are cross-linked to each other with several SCs. However, the two α-helices are not cross-linked to each other, neither to the β-sheet, but they are cross-linked to the peptide ligand by the SC-s composed of residues from both the ligand and the protein (Figure 2). Hence, these combined SCs disappear when the peptide dissociates from the protein, so the helical parts of the protein are not protected by SCs any longer. The formation of these combined stabilization centers provides a simple tool for regulating protein turnover. The protein is stable in its ligand-bound form, while it becomes unstable (degradable) after losing its ligand.[27]

Figure 2. Stabilization centers in the peptide-binding domain of MHC-peptide complex, PDB code: 1a1m. Stabilization centers that involve residues from the bound peptide are marked in black. The rest of the stabilization centers are marked in dark gray. The picture was prepared using PyMOL (W. L. DeLano. The PyMOL Molecular Graphics System. (2002) on World Wide Web http://www.pymol.org).

A recent study indicated that MHC is not the only protein family where residues responsible for stability are not segregated from those that are involved in the biological function. Type II restriction endonucleases protect bacterial cells from viral infection by catalyzing phosphodiester bond hydrolysis in specific DNA sequences. These enzymes are sequentially diverse, the homology among them is less than 25%. On the other hand, type II restriction endonucleases exhibit similar structural features. Each structure consists of a core motif, with 5−6 β-strands surrounded by several α-helices. The active sites also show a weak consensus sequence. We investigated the evolutionary relationship between these enzymes by analyzing the common stabilization centers in the available crystal structures of PD-(D/E)-XK endonucleases [Figure 3]. In general, the importance of SCs in structural stability is reflected in a higher evolutionary conservation. The amino acid conservation of residues involved in SCs was higher compared to the rest of the protein. SCs also showed higher conservation at a structural

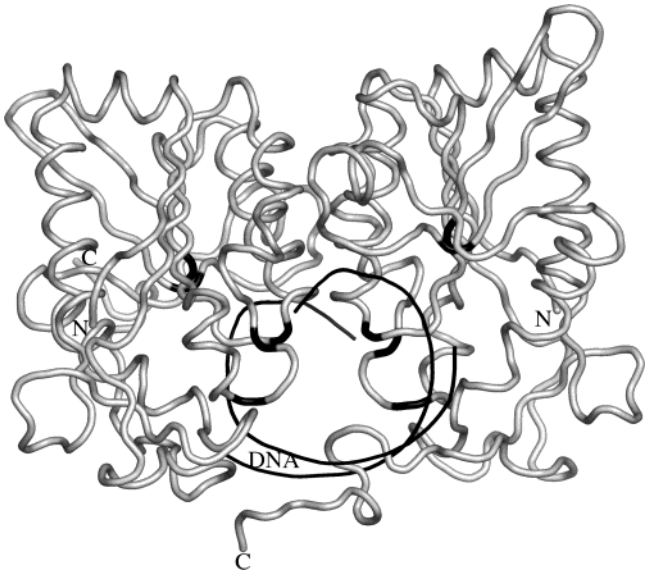

Figure 3. Stabilization center residues of the recognition and the active site of type II restriction endonuclease, PDB code: 2bam, marked in black. The bound DNA is indicated in a continuous black line. The picture was prepared using PyMOL (W. L. DeLano. The PyMOL Molecular Graphics System. (2002) on World Wide Web http://www.pymol.org).

level, as the interactions formed between the SCs were more conserved than other long-range interactions.[15] This observation suggested a way to identify the evolutionary most conserved parts of the structure through localization of SCs. We focused on three parts of PD-(D/E)XK endonuclease: core motif, active site residues, and residues involved in DNA recognition. The structures of 14 enzymes in free form, in complexes with substrate DNA, and in a catalytically active form with DNA and a metal ion were analyzed. In general, the common α/β core motif was not found to provide a dominant contribution to structure stabilization. However, the active site residues and some of the residues of the DNA binding site are involved in SC formation. Thus, we can conclude that these functionally important parts represent the evolutionary most conservative residues in PD-(D/E)XK endonucleases. This result confirms the divergent evolutionary origin of these enzymes. Despite the low level of sequence similarity and the different strategies for interaction with their substrate, they have evolved from a common ancestor.[28]

What is more, in the case of AmpC β-lactamase it was shown that active site residues destabilize the protein structure. Mutations of the active site residues not only resulted in a 10^3-10^5-fold decrease of the enzymatic activity, but also resulted in an up to 4.7 kcal/mol increase in structural stability.[29]

The cases discussed above raise the general question about the overlap of structurally and functionally relevant residues. Many enzymes have the same folds, e.g. most dehydrogenases have the Rossman fold, i.e., the same overall structure, but a different substrate and reaction specificity. These enzymes evolved from a common ancestor by mutating, deleting, and inserting residues, which are involved in enzyme specificity, and by keeping unchanged those residues which are involved in the formation and stability of the original overall structure. At the same time, the same enzymatic function of a dehydrogenase has to be conducted in a mammalian cell and in a thermophilic microorganism, too. The latter enzymes obviously have much higher thermostability. In both cases, it would look much simpler, if residues involved in function and in stability were separated. This is probably the reason sometimes conservative residues are classified separately as residues conserved for structural or functional reasons.[30] Our survey on hundreds of unrelated proteins shows that nature does not follow this anthropomorphic logic.

Starting from the FSSP database[31] 772 unrelated polypeptide chains were selected with functionally important residues. Functionally important residues were identified from the SITE record in the PDB database.[32] From each FSSP family the representative or its closest homologue with SITE record was chosen. The resulting data set was checked for redundancy using the BLASTCLUST program.[33]

From the 196 248 residues of the 772 polypeptide chains 5775 residues (2.94%) had SITE record in the PDB, and 48 538 residues (24.73%) were SC elements. It was found that there is no stronger separation between residues in functional sites and in SCs than for any other groups of residues of the same sizes. On the contrary, the overlap between the structurally and functionally relevant residue sets, 1524 residues, is even marginally higher than expected from the sizes of these sets, i.e., 1427 residues. Considering binomial distribution for the random case, the difference is about 2.5 standard deviation units, which is a significant difference from a statistical point of view. It is worth noting that more than half of the proteins are enzymes, where the difference between the observed and expected values of the overlapping residues is more than 5 standard deviation units.

CONCLUDING REMARKS

Polypeptide chains with a genetically determined amino acid residue sequence are resistant to proteolytic degradation and exhibit biochemical function only in certain three-dimensional structures. Since the energy of individual long-range interresidue interactions are much too small to resist thermal fluctuation driven spontaneous temporally unfolding, which could lead to irreversible proteolysis, "stable crosslinks" are necessary to maintain the native structure of proteins. The overwhelming majority of these cross-links are clusters of noncovalent interresidue contacts. We proposed a definition to locate representatives of these clusters, SCs. Many of the properties of residues involved in the formation of these clusters were discussed earlier. Here we focused on the relation between these special clusters and other structurally or functionally important elements of proteins. We found that secondary structure elements significantly influence the formation of SCs, but this later one has little, if any, effect on secondary structure formation. Analysis of SCs in MHC proteins showed the significant effect of these centers for unfolding kinetics, while the analysis of SCs in the four-helix bundle proteins indicated that SC also has a small, but detectable contribution to the conformational free energy. We got the surprising results that residues involved in protein stabilization and in biochemical functions are not segregated. What is more, the overlap between these two sets of residues is even higher than between any randomly selected residue sets of the same sizes. We conclude that the discussed examples suggest a strong interplay between noncovalent

cross-links and other structural-functional elements in proteins, instead of the expected "specialization" among residues.

ACKNOWLEDGMENT

This work was sponsored by OTKA Grants: D34572, D38487, F43569, F43609, and T34131. Z.D and M.F. also acknowledge support from Bolyai János Scholarship.

REFERENCES AND NOTES

(1) Vasquez, M.; Scheraga, H. A. Calculation of protein conformation by the build-up procedure. Application to bovine pancreatic trypsin inhibitor using limited simulated nuclear magnetic resonance data. *J. Biomol. Struct. Dyn.* **1988**, *5*, 705−755.
(2) Vasquez, M.; Scheraga, H. A. Variable-target-function and build-up procedures for the calculation of protein conformation. Application to bovine pancreatic trypsin inhibitor using limited simulated nuclear magnetic resonance data. *J. Biomol. Struct. Dyn.* **1988**, *5*, 757−784.
(3) Simon, I.; Glasser, L.; Scheraga, H. A. Calculation of protein conformation as an assembly of stable overlapping segments: application to bovine pancreatic trypsin inhibitor. *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 3661−3665.
(4) Kabsch, W.; Sander, C. On the use of sequence homologies to predict protein structure: identical pentapeptides can have completely different conformations. *Proc. Natl. Acad. Sci. U.S.A.* **1984**, *81*, 1075−1078.
(5) Simon, I. Investigation of protein refolding: a special feature of native structure responsible for refolding ability. *J. Theor. Biol.* **1985**, *113*, 703−710.
(6) Bonneau, R.; Tsai, J.; Ruczinski, I.; Chivian, D.; Rohl, C.; Strauss, C. E.; Baker, D. Rosetta in CASP4: progress in ab initio protein structure prediction. *Proteins Suppl.* **2001**, *5*, 119−126.
(7) Richards, F. M. The interpretation of protein structures: total volume, group volume distributions and packing density. *J. Mol. Biol.* **1974**, *82*, 1−14.
(8) Vendruscolo, M.; Dokholyan, N. V.; Paci, E.; Karplus, M. Small-world view of the amino acids that play a key role in protein folding. *Phys. Rev. E Stat. Nonlin. Soft. Matter Phys.* **2002**, *65*, 061910.
(9) Creighton, T. *Proteins: Structures and Molecular Properties*; W. H. Freeman: New York, 1993; pp 162−164.
(10) Baker, D. A surprising simplicity to protein folding. *Nature* **2000**, *405*, 39−42.
(11) Gugolya, Z.; Dosztányi, Z.; Simon, I. Interresidue interactions in protein classes. *Proteins* **1997**, *27*, 360−366.
(12) Thornton, J. M. Disulphide bridges in globular proteins. *J. Mol. Biol.* **1981**, *151*, 261−287.
(13) Zehfus, M. H. Automatic recognition of hydrophobic clusters and their correlation with protein folding units. *Protein Sci.* **1995**, *4*, 1188−1202.
(14) Kannan, N.; Vishveshwara, S. Identification of side-chain clusters in protein structures by a graph spectral method. *J. Mol. Biol.* **1999**, *292*, 441−464.
(15) Dosztányi, Z.; Fiser, A.; Simon, I. Stabilization centers in proteins: identification, characterization and predictions. *J. Mol. Biol.* **1997**, *272*, 597−612.
(16) Dosztányi, Z.; Magyar, Cs.; Tusnády, G. E.; Simon, I. SCide: identification of stabilization centers in proteins. *Bioinformatics* **2003**, *19*, 899−900.
(17) Fiser, A.; Dosztányi, Z.; Simon, I. The role of long-range interactions in defining the secondary structure of proteins is overestimated. *Comput. Appl. Biosci.* **1997**, *13*, 297−301.
(18) Ptitsyn, O. B.; Finkelstein, A. V. Theory of protein secondary structure and algorithm of its prediction. *Biopolymers* **1983**, *22*, 15−25.
(19) Chou, P. Y.; Fasman, G. D. Prediction of the secondary structure of proteins from their amino acid sequence. *Adv. Enzymol. Relat. Areas Mol. Biol.* **1978**, *47*, 45−148.
(20) Deleage, G.; Clerc, F. F.; Roux, B.; Gautheron, D. C. ANTHEPROT: a package for protein sequence analysis using a microcomputer. *Comput. Appl. Biosci.* **1988**, *4*, 351−356.
(21) Garnier, J.; Osguthorpe, D. J.; Robson, B. Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. *J. Mol. Biol.* **1978**, *120*, 97−120.
(22) Dosztányi, Z.; Simon, I. Stabilization centers in various proteins. *Theor. Chem. Acc.* **1999**, *101*, 27−32.
(23) Rogers, N. K.; Sternberg, M. J. Electrostatic interactions in globular proteins. Analysis of contributions from induced protein dipoles. *J. Mol. Biol.* **1984**, *174*, 527−542.
(24) Fuxreiter, M.; Simon, I. Role of stabilization centers in 4 helix bundle proteins. *Proteins* **2002**, *48*, 320−326.
(25) Lanzavecchia, A.; Reid, P. A.; Watts, C. Irreversible association of peptides with class II MHC molecules in living cells. *Nature* **1992**, *357*, 249−252.
(26) Arndt, S. O.; Vogt, A. B.; Markovic-Plese, S.; Martin, R.; Moldenhauer, G.; Wolpl, A.; Sun, Y.; Schadendorf, D.; Hammerling, G. J.; Kropshofer, H. Functional HLA-DM on the surface of B cells and immature dendritic cells. *EMBO J.* **2000**, *19*, 1241−1251.
(27) Simon, Á.; Dosztányi, Z.; Rajnavölgyi, É.; Simon, I. Function-related regulation of the stability of MHC proteins. *Biophys. J.* **2000**, *79*, 2305−2313.
(28) Fuxreiter, M.; Simon, I. Protein stability indicates divergent evolution of PD-(D/E)XK type II restriction endonucleases. *Protein Sci.* **2002**, *11*, 1978−1983.
(29) Beadle, B. M.; Shoichet, B. K. Structural bases of stability-function tradeoffs in enzymes. *J. Mol. Biol.* **2002**, *321*, 285−296.
(30) Golovanov, A. P.; Efremov, R. G.; Jaravine, V. A.; Vergoten, G.; Arseniev, A. S. Amino acid residue: is it structural or functional? *FEBS Lett.* **1995**, *375*, 162−166.
(31) Holm, L.; Sander, C. Mapping the protein universe. *Science* **1996**, *273*, 595−602.
(32) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucl. Acids Res.* **2000**, *28*, 235−242.
(33) Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **1990**, *215*, 403−410.

CI030409I