Theor Chem Acc (2001) 106:121–127
DOI 10.1007/s002140000207

Theoretical
Chemistry Accounts

Regular article

Stabilization centers and protein stability

Á. Simon[2,3], Z. Dosztányi[2], C. Magyar[2], G. Szirtes[2], É. Rajnavölgyi[3,4], I. Simon[1,2]

[1] e-mail: simon@enzim.hu; Tel.: +36-1-4669276; Fax: +36-1-4665465
[2] Institute of Enzymology, BRC, Hungarian Academy of Sciences, P.O. Box 7, 1518 Budapest, Hungary
[3] Department of Immunology, Eötvös Loránd University, Jávorka S. u. 14, 2131 Göd, Hungary
[4] Institute of Immunology, Faculty of Medicine, Medical and Health Science Center, University of Debrecen, Nagyerdei krt. 98, 4012 Debrecen, Hungary

Received: 25 August 2000 / Accepted: 6 September 2000 / Published online: 21 December 2000

Abstract. The well-balanced stability of protein structures allows large-scale fluctuations, which are indispensable in many biochemical functions, ensures the long-term persistence of the equilibrium structure and it regulates the degradation of proteins to provide amino acids for biosynthesis. This balance is studied in the present work with two sets of proteins by analyzing stabilization centers, defined as certain clusters of residues involved in cooperative long-range interactions. One data set contains 56 proteins, which belong to 16 families of homologous proteins, derived from organisms of various physiological temperatures. The other set is composed of 31 major histocompatibility complex (MHC)–peptide complexes, which represent peptide transporters complexed with peptide ligands that apparently contribute to the stabilization of the MHC proteins themselves. We show here that stabilization centers, which had been identified as special clusters of residues that protect the protein structure, evolved to serve also as regulators of function – related degradation of useless protein as part of protein housekeeping.

Key words: Protein thermostability – Major histocompatibility complex protein – Protein stability

1 Introduction

To perform their function most protein have to be rather flexible to allow large-scale structural fluctuation on the 10^{-7}–10^{-3} s timescale, which is indispensable for many biochemical functions, like enzyme activity. Therefore, proteins should have a well-balanced stability allowing structural fluctuations and concomitantly ensuring a long-lasting equilibrium structure as indicated by many authors [1, 2, 3]. Long lasting does not mean lasting forever, not even lasting for the lifetime of the cell. Protein housekeeping involves protein turnover, controlled mainly by the slow, spontaneous degradation of proteins, which typically takes place on the 10^3–10^7 s timescale. This slow process also has to be well controlled and the protein structure should be intact as long as it is required to fulfill biological functions. On the other hand, for many functions, like protease-controlled processes of blood clotting or transporting intracellular peptides to the cell surface by "single use" transporters, the major histocompatibility complex encoded proteins (MHC proteins) etc. need newly synthesized proteins to accomplish their biological function and for the degradation of inactive proteins, which provide amino acids for the biosynthesis of new proteins.

In this article, we suggest that stabilization centers (SCs), composed of certain clusters of residues, involved in cooperative long-range interactions in proteins, that regulate flexibility and rigidity (stability) of protein structures are also important in the regulation of the turnover of certain proteins.

A SC is a cluster of two or more residues, which is defined as follows (Fig. 1) [4]

1. Two residues are in contact if there is at least one atom–atom distance between the two residues which is shorter than the sum of their van der Waals radii plus 1 Å.
2. A contact is recognized as "long-range interaction" if the two residues involved are separated by at least 10 residues in the primary structure.
3. Two residues are SC elements if they are in long-range interaction, and if it is possible to select one–one residues from both flanking tetrapeptides of these two residues that make at least seven contacts between these two triplets.

According to this definition any residues of a polypeptide chain can be a SC element, except the N- and C-terminal ones, which do not have "supporting residues" from flanking regions at both sides of the chain. Finding a SC element in the proximate vicinity (fewer than four amino acids) of the termini is possible, but less probable

Correspondence to: I. Simon
Contribution to the Symposium Proceedings of Computational Biophysics 2000

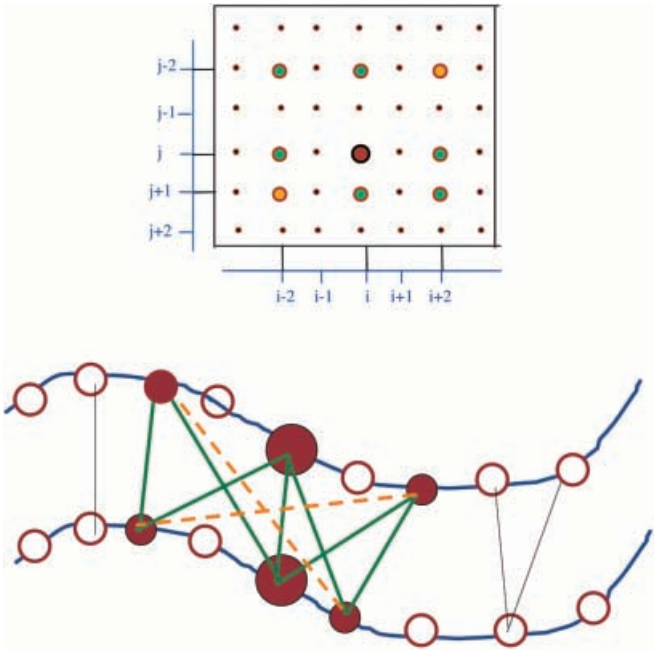

Fig. 1. Two residues, *i* and *j*, form stabilization centers, since they are in long-range contact and they have one–one supporting residues from their flanking tetrapeptides, so the triplets $i-2$, i, $i+2$ and $j-2$, j, $j+1$ form seven out of the nine possible contacts. The two missing contacts are indicated in *yellow* in the contact map and in the ribbon scheme

than inside the chain, since the "supporting residues" can only be selected from one, two or three residues instead of four as inside the chain.

Note, that a SC element can, and generally does, form SCs with more than one other SC element. In fact, one stretch of SC usually contains several SC elements. The properties of SCs were described in more detail in Ref. [4] and the relation of these tertiary structure elements to the formation of secondary structure elements are discussed in Ref. [5]. The appearance of SCs in various secondary structure subclasses were presented at the Computational Chemistry and the Living World meeting in Chanberry, France, 1998, and were published in Ref. [6]. A public server is available at http://www.enzim.hu/scpred/scpred.html to identify SC elements of certain proteins of known structure and to predict such elements when only the amino acid sequence of the protein is available.

According to our earlier survey on a 600-membered representative set of unrelated protein in the Protein Data Bank (PDB) [7] SCs were found in all proteins and almost 25% of the residues were identified as SC elements. This percentage strongly depended on the secondary structure composition of the protein, ranging from 13% for the all-α protein subclass members to 33% for those of the all-β protein subclass [6]. Owing to the quadratic nature of the contact maps, SCs occupy only a very small part of the residue–residue contact maps. The free energy of protein stabilization is a small difference between a large contribution of attractive and repulsive forces [8]; therefore, it is expected that even a few additional SC elements can be significant in terms of stability. The effect of SCs in protein stability is analyzed in two datasets in this article. One contains 56 proteins, which belong to 16 families of homologous proteins of organisms living at various temperatures, collected by Vogt and Argos [9]. SCs of themophillic–mesophillic counterparts are compared in this dataset. The other set is composed of 31 MHC–peptide complexes, including MHC class I and class II proteins complexed with various natural peptide ligands. Classical MHC–proteins are considered "for single use" peptide transporters, which carry selected peptide products of limited proteolytic degradation from inside the cell to the cell surface to present the peptides to T cells. These proteins fulfill their functions as long as they bind their dissociable ligand, the peptide. Peptide-free MHC-molecules on the cell surface are practically useless for their primary biological function, which is contacting the appropriate T cell receptor, and are less stable than the complex. In this article, we show that in the stabilization of these complexes special SCs are involved; namely, SCs that are composed of residues derived from both the protein and the peptide ligand. These SCs disappear if the peptide dissociates from the protein, and thus these peptide–protein SCs provide an effective means of function-related regulation of protein stability.

2 Datasets

Homologous proteins with different thermostability were selected by Vogt and Argos [9]. After checking their quality 49 protein structures remained in 15 families. Seven structures (3gpd,1llc,1cer,2ptk,3pfk,1lpf, 3lad) were excluded from the original database because of their poor resolution or missing atomic coordinates.

Peptide binding domains of eight MHC class I and seven MHC class II proteins complexed with one or more peptides, altogether 21 MHC I–peptide complexes and ten MHC II–peptide complexes, were selected by us [10].

The PDB file code and the characteristics of the proteins studied are summarized in Tables 1, 2, and 3.

3 Results and discussion

The comparison of the number of SC elements per residue in the 15 families (49 proteins) studied showed a slight, but significant, increase in the "density" of the SC elements with the elevation of the physiological temperature. One pair of homologous proteins, rubredoxin from *Desulfovibro gigas* and from *Pyrococcus furiosus,* with different thermostability is compared in Fig. 2. There is only one more SC present in the thermophillic protein compared to the one from mesophillic origin. The physiological temperature of *D. gigas* is 35.5 °C, while for *P. furiosus* it is 110 °C. The extra pair of SC elements forming an SC is marked green in the contact map and at the upper part of the ribbon representation of the thermostable protein (Fig. 2), while the lower parts of both proteins contain the same SC pattern. Similar differences in the SC pattern involving only a few

Table 1. Mesophillic and thermophillic proteins and their Protein Data Bank (PDB) codes: (the average physiological temperature of the source organism and the ratio of the number of stabilization center elements and the total number of residues are shown in *parentheses*)

Protein	PDB identification
Malate dehydrogenase	4mdh(37;0.237),1bmd(72.5;0.174)
Glycosyltransferase	1cdg(35;0.337),1cgt(35;0.327),1cyg(52.5;0.325),1ciu(60;0.362)
Glyceraldehyde-3 phosphate dehydrogenase	4gpd(20;0.18),1gad(37;0.3),1gd1(52.5;0.272),1hdg(82.5;0.316)
Lactate dehydrogenase	6ldh(20;0.2),5ldh(37;0.051),9ldb(37;0.199),1lld(39;0.227), 1ldn(52.5;0.234)
Thermolysin	1npc(30;0.202),1lnf(52.5;0.222)
Ribonuclease H	2rn2(37;0.148),1ril(72.5;0.17)
Subtilisin	1st3(30;0.346),1sup(35;0.345),1sca(42.5;0.365),1thm(60;0.355)
Ferredoxin	1fca(28;0.164),1fdx(37;0.204),2fxb(52.5;0.198)
Superoxide dismutase	3sdp(27.5;0.097),1abm(37;0.121),1isa(37;0.156),1ids(37;0.1), 3mds(72.5;0.172)
Phosphoglycerate kinase	3pgk (27.5;0.154), 1php (52.5;0.258)
Triose phosphate isomerase	1ypi(27.5;0.206),1hti(37;0.238),1tim(37;0.142),1tpe(41;0.233), 1btm(52.5;0.246)
Rubredoxin	1rdg(35.5;0.25),6rxn(35.5;0.311),8rxn(35.5;0.25), 5rxn(37;0.278),1caa(110;0.283)
Hydrolase	1ino(37;0.326),2prd(72.5;0.345)
Glycosylhydrolase	2exo(30;0.253),1xyz(60;0.256)
Reductase	1lvl(27.5;0.225),1edb(52.5;0.258)

Table 2. Major histocompatibility complex (*MHC*) I–peptide complexes

PDB code	MHC I allotype	Origin of peptide
1hhg	HLA-A*0201	HIV-1 glycoprotein 120 (197–205)
1hhh	HLA-A*0201	Hepatitis B virus capsid (18–27)
1hhi	HLA-A*0201	Influenza A matrix (58–66)
1hhj	HLA-A*0201	HIV-1 reverse transcriptase (476–484)
1hhk	HLA-A*0201	HTLV-Tax
2clr	HLA-A*0201	Calreticulin
1b0g	HLA-A*0201	HLA-A2 specific human peptide "P1049"
1agd	HLA-B*0801	HIV-1 gag (24–31)
1a1n	HLA-B*3501	HIV-1 Nef (74–81)
1a9e	HLA-B*3501	Epstein-Barr virus nuclear antigen (EBNA)–3C derivative
1a1o	HLA-B*5301	Plasmodium-falciparum liver antigen (1786–1794)
1alm	HLA-B*5301	HIV-2gag(182–190)
1hoc	H-2D^b	Influenza A virus nucleoprotein (366–374)
1ce6	H-2D^b	Sendai virus nucleoprotein (324–332)
1bz9	H-2D^b	Synthetic peptide "P1027"
2vaa	H-2K^b	Vesicular stomatitis virus nucleoprotein (52–59)
2vab	H-2K^b	Sendai virus nucleoprotein (324–332)
1vac	H-2K^b	Chicken ovalbumin (258–265)
1vad	H-2K^b	α-Glucosidase(438–446)
1ld9	H-2L^d	Endogenous peptide p29
1ddh	H-2D^d	HIV-1 glycoprotein160 V3 loop (P18-I10)

Table 3. MHC II–peptide complexes

PDB code	MHC II allotype	Origin of peptide
1dlh	HLA-DR1 (DRA*0101, DRB1*0101)	Influenza virus hemagglutinin (306–318)
1aqd	HLA-DR1 (DRA*0101, DRB1*0101)	Endogenous peptide A2(104–117)
2seb	HLA-DR4 (DRA*0101, DRB1*0401)	Human collagen II (1168–1179)
1bx2	HLA-DR2 (DRA*0101, DRB1*1501)	Human myelin basic protein (85–98)
1a6a	HLA-DR3 (DRA*0101, DRB1*0301)	CLIP (87–101)
1iea	H-2E^k	Murine hemoglobin (64–76)
1ieb	H-2E^k	Murine heat shock protein 70 (236–248)
1iao	H-2A^d	Ovalbumin (323–334) + signal residues
2iad	H-2A^d	Influenza hemagglutinin (126–138) + signal residue
1iak	H-2A^k	Hen egg lysozyme (50–62)

SC elements were found in the other families. The ratios of the number of SC elements and the number of protein residues are also listed in Table 1. On average, almost one (0.92) extra SC element per 100 residues was found with an increase of 10 °C in the physiological temperature. This is only valid for the whole database studied;

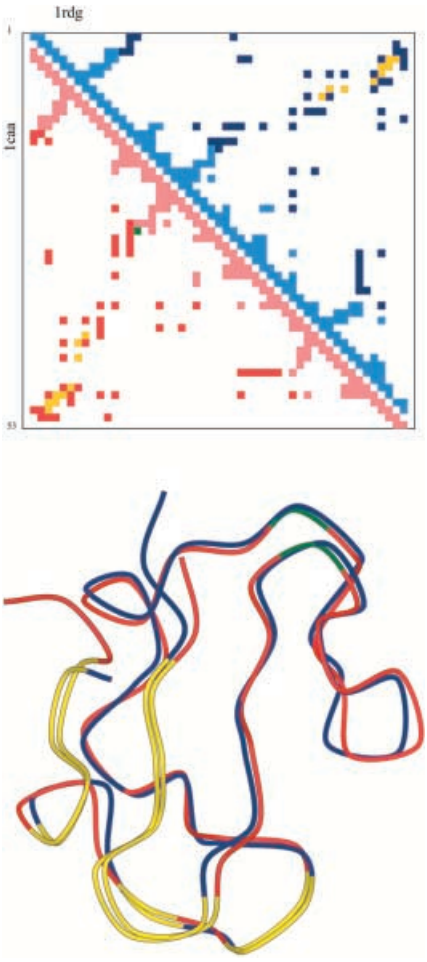

Fig. 2. *Upper part*: contact map of interactions in rubredoxin. Upper right: *Desulfovibro gigas* (1rdg, mesophillic). Short-range interactions are shown in *light blue*, long-range interactions other than stabilization centers are *blue*, stabilization centers are marked *yellow*. Lower left: *Pyrococcus furiosus* (1caa, thermophillic). Short-range interactions are shown in *pink*, long-range interactions other than stabilization centers are *red*, stabilization centers are marked *yellow*. The extra stabilization center as compared to the mesophillic analogue is shown in *green*. *Lower part*: ribbon diagram of rubredoxin isolated from *D. gigas* (*blue*), mesophillic, and *P. furiosus* (*red*), thermophillic organisms. Stabilization centers are shown in *yellow*. The extra stabilization center on *P. furiosus* is highlighted in *green*

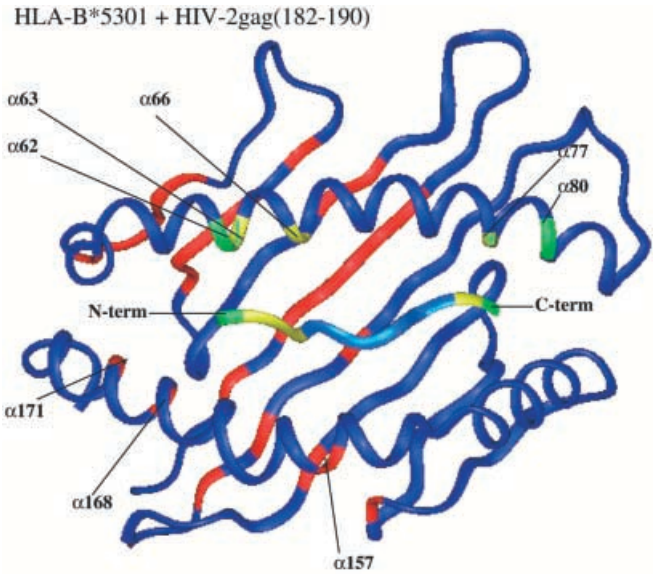

Fig. 3. Human major histocompatibility complex (*MHC*) I molecule HLA-B*5301 complexed with a peptide from HIV-2gag (*light blue*). The *numbers* indicate MHC residues of helical origin that form stabilization centers in this complex. Stabilization centers formed within the MHC molecule are marked *red*. Stabilization centers which involve residues of the bound peptide are shown in *yellow*. Quasi stabilization centers are marked in *green*

individual families or pairs of homologous proteins from sources of different physiological temperature exhibit significant deviation from this pattern. It is also known that the thermostability of the individual proteins does not correlate perfectly with the physiological temperature of the source organism.

There are several surveys in the literature on the comparison of various kinds of interactions: hydrogen bonds, salt bridges, van der Waals interactions and others, for thermostable and nonthermostable protein counterparts [11, 12, 13]. None of these surveys pinpoints any single kind of interactions responsible for the different thermostability of homologous proteins, as a general rule. We do not think either that we have found the philosopher's stone in SCs, but the results of our comparison suggests that the level of cooperativity among the interresidue interactions might be more significant from the viewpoint of stability than the number, the energy or some other feature of the individual interactions. Apparently, the delicate balance between rigidity and flexibility may be regulated by a few SC residues, at least in some proteins.

Maintaining the equilibrium structure in the course of function-related fluctuation is only one part of the story. The other part concerns protein housekeeping, since degradation of nonfunctional proteins is the main source of amino acids for the biosynthesis of novel proteins. It is especially important for proteins which have to be newly made to accomplish their biological function, to evolve mechanisms for the regulated degradation of useless or dispensable proteins. MHC–proteins are examples of such proteins. The SC pattern of the ligand-binding domains of MHC–peptide complexes was studied to uncover how evolution solved the problem of function – related regulation of stability.

SCs of the ligand–binding domains of the MHC class I protein HLA-B*5301 complexed with the HIV2-gag 182–190 peptide [14] are shown in Fig. 3. Similar pictures are given for the other 20 MHC class I and ten MHC class II complexes depicted in Figs. 4 and 5, respectively. Figure 3 clearly demonstrates that the eight-stranded β sheets are stabilized by a large number of SCs. Considering the 31 complexes studied, an average of 27 residues were involved in these interactions. The average number of SC residues connecting the helices to the β-sheet plateau is only 3. There are only eight out of the 31 cases where both helices are linked to the β sheet by SCs. This finding is in good agreement with our previous results obtained on a large dataset – that is for the α/β

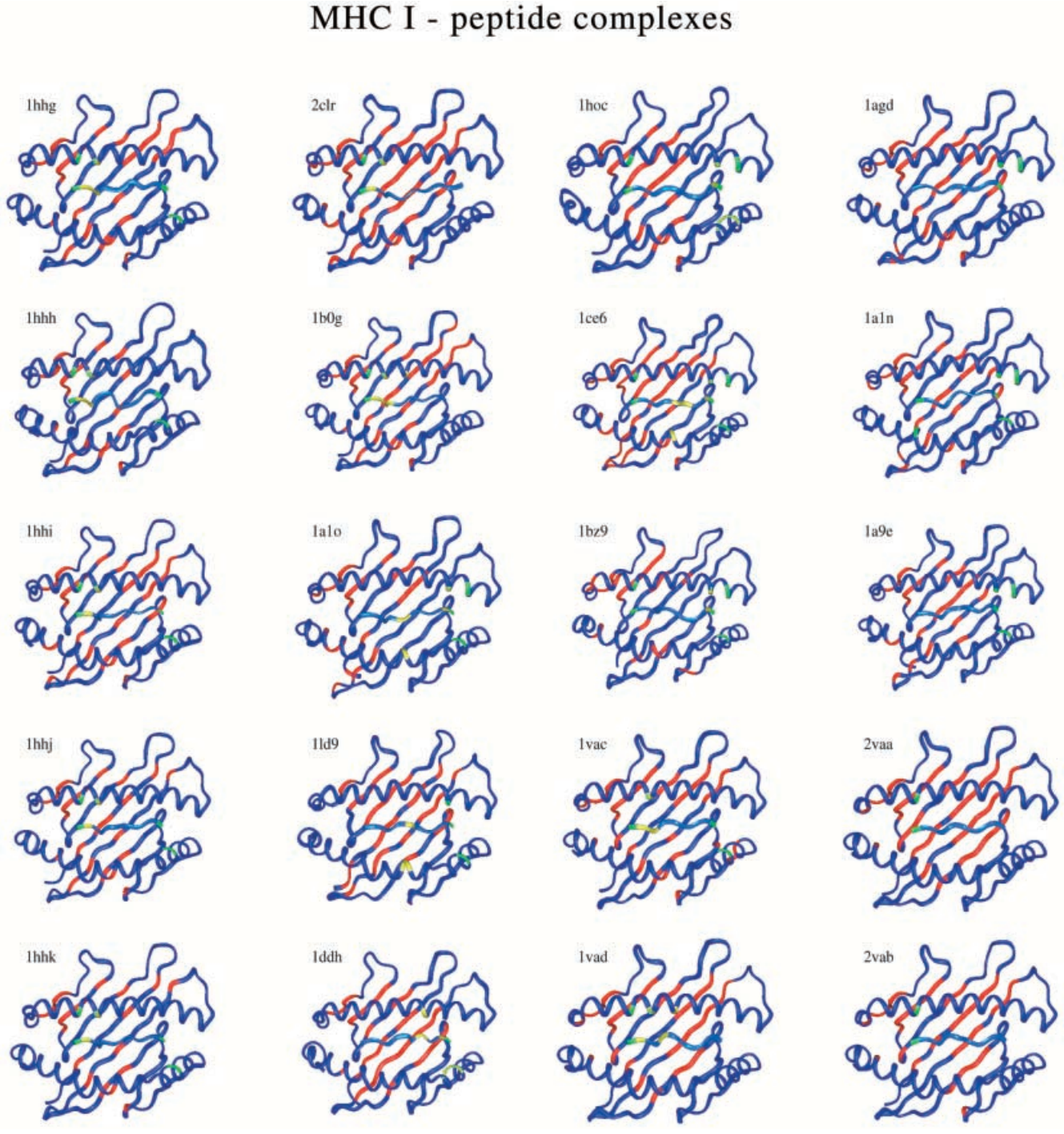

Fig. 4. MHC I proteins complexed with peptides selected from the Protein Data Bank (*PDB*) (Table 2). The color codes are the same as in Fig. 3. Note, that the last column contains complexes with no peptidic stabilization center elements (no yellow spots)

subclass of proteins, which MHC–complexes belong to, the average number of SCs connecting extended chains were found to be much more abundant than those connecting the extended chain to an α helix [6].

At least one residue of the bound peptide is involved in SCs in all the class II protein–peptide complexes studied and in 16 of the 21 class I–peptide complexes analyzed. In all the cases investigated, the complementary residues of the SC elements identified in the bound peptide were localized in the helices and never in the β sheet of the MHC–molecules. The remaining five complexes are shown in the last column of Fig. 4. It is interesting that while the stability of MHC I molecules is known to be more dependent on the bound peptide than that of MHC II, all the SC-free peptides were found in MHC I–peptide complexes. It comes from the special shape of the ligand-binding groove of MHC I proteins and from the definition of the SC. The main difference

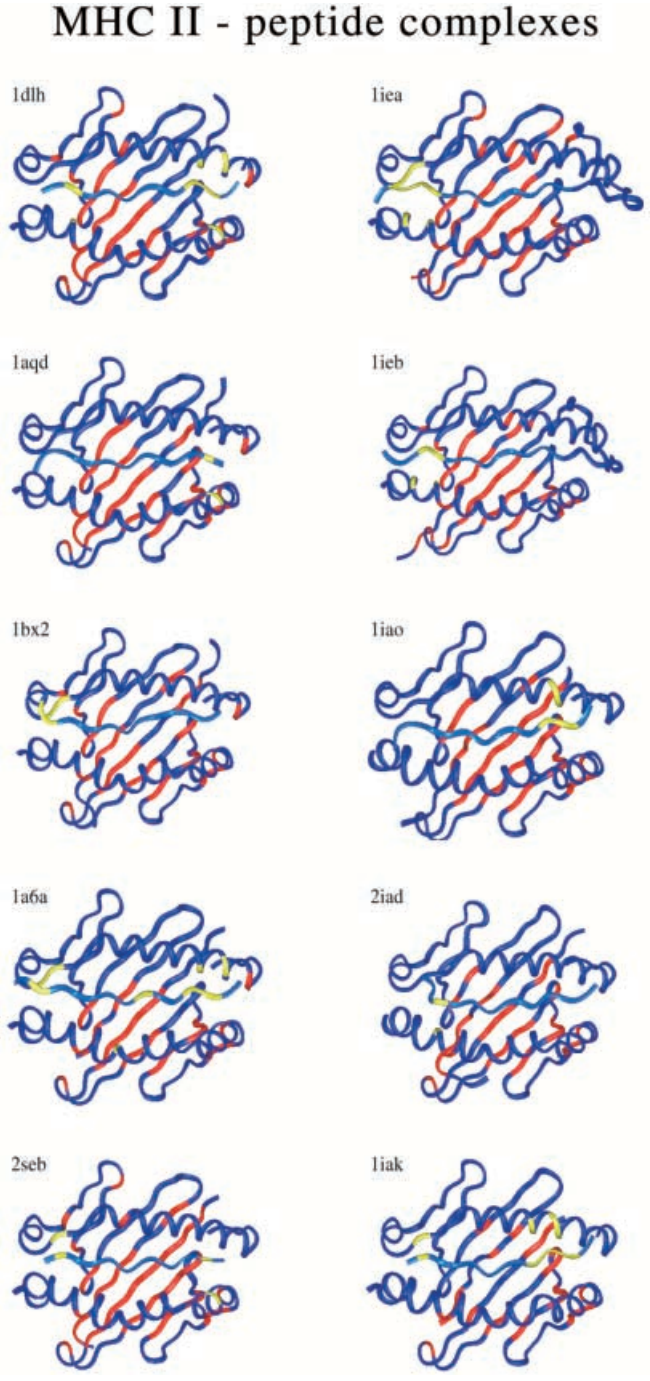

Fig. 5. MHC II proteins complexed with peptides, selected from the PDB database. For color codes refer to Fig. 3

between the ligand-binding grooves of MHC I and MHC II molecules is that the peptide binding groove of MHC II molecules is open at both ends, while it is closed in MHC I molecules. Therefore, MHC I molecules can accommodate ligands only of 8–10 residues in length and the strongest interacting residues of the ligand are the terminal ones, which not only interact with the side of the groove, but also with the conservative closing residues. However, these strongly interacting terminal residues of the bound peptide cannot be recognized as SC elements, since according to our definition terminal residues can never be SC elements, as defined before [4]. According to our earlier study on a large protein database it was found that SC elements are not only centers of cooperative interactions, but they themselves also made 1.77 times more long-range interactions than those residues that were involved in long-range interactions but were not SC elements; therefore, the extremely large number of interatomic interactions characterizes SC elements. The comparison of the number of long-range atomic contacts shows that the average number is larger for residues which serve as peptide termini in MHC class I complexes than the average number of such contacts of SC residues in the MHC I–peptide complex dataset or of SC residues in the larger datasets of globular proteins, published earlier [4].

The observed number of interatomic interactions of the terminal residues in question clearly shows that the interactions between these residues and the MHC class I proteins are very pronounced. Thus, they play the same role as SC elements do in structure stabilization. Therefore, although they are not formally SC elements, as "quasi SC" elements they can have the same function as SC elements. A terminal residue of the bound peptide is considered as a "quasi SC" element if it interacts with another residue in the protein and it is possible to select one residue of the flanking tetrapeptide of the terminal residue and one–one residues from the flanking tetrapeptides of the interacting protein residue, so five out of the six possible interactions between the residue pair and the triplet are made. These terminal "quasi SC" residues of the bound peptide and their interacting partner, i.e. the central residue of the triplet mentioned previously are marked in green in Figs. 3 and 4. In all 21 MHC class I protein–peptide complexes studied there is at least one terminal residue of the peptide which is recognized as a "quasi SC" element and in 11 cases both termini are "quasi SC" elements.

Our survey on the 31 MHC–peptide complexes of known three-dimensional structure suggests that SCs and "quasi SCs" which are composed of residues derived from both the protein and the complexed peptide can be vital for ensuring the compact structure of a MHC protein by fixing the α helices. The control of MHC–protein housekeeping is apparently based on the formation of these special SCs. This arrangement of SCs provides a simple means of regulation, which makes the useful form of MHC molecules stable, while it makes the useless form of the same proteins unstable and therefore degradable.

We conclude that SCs, which in general protect the protein structure against spontaneous degradation due to thermal fluctuation, have also been used in evolution to develop a simple means for regulating the housekeeping of an immunologically important protein family, the classical MHC molecules. Up to now, it is unclear if this is a more generally used way of regulating protein housekeeping or if MHC–peptide complexes represent a special case.

Acknowledgements. This work was supported by research grants OTKA T017157, T030566, FKFP0186/1999 and AKP 98–13 3,3, by the Hungarian–Israeli Intergovernmental S and T Cooperation

Programs and by the scientific collaboration project between the Hungarian Academy of Sciences and the NSF (USA). We thank Gábor E. Tusnády for his help in programming and Nicholas E. Dixon for his critical comments on the manuscript.

References

1. Anfinsen CB, Scheraga HA (1975) Adv Protein Chem 29: 205–300
2. Jaenicke R (1999) Prog Biophys Mol Biol 71: 155–241
3. Závodszky P, Kardos J, Svingor Á, Petsko GA (1998) Proc Natl Acad Sci USA 95: 7406–7411
4. Dosztányi Z, Fiser A, Simon I (1997) J Mol Biol 272: 597–612
5. Fiser A, Dosztányi Z, Simon I (1997) Comput Appl Biosci 13: 297–302
6. Dosztányi Z, Simon I (1999) Theor Chem Acc 101: 27–32
7. Bernstein F, Koetzle T, Williams G, Mayer JE, Brice M, Rodgers J, Kennard O, Shomanouchi T, Tasumi M (1977) J Mol Biol 112: 535–542
8. Jaenicke R (1991) Eur J Biochem 202: 715–728
9. Vogt G, Argos P (1997) Folding Design 2: S40–46
10. Simon A, Dosztányi Z, Rajnavölgyi É, Simon I (2000) Biophys J 79: 2305–2313
11. Karshikoff A, Ladenstein R (1998) Protein Eng 11: 867–872
12. Gromiha MM, Oobatake M, Sarai A (1999) Biophys Chem 82: 51–67
13. Szilágyi A, Závodszky P (2000) Struct Folding Design 8: 493–504
14. Smith KJ, Pyrdol J, Gauthier L, Wiley DC, Wucherpfennig KW (1996) Immunity 4: 215–228