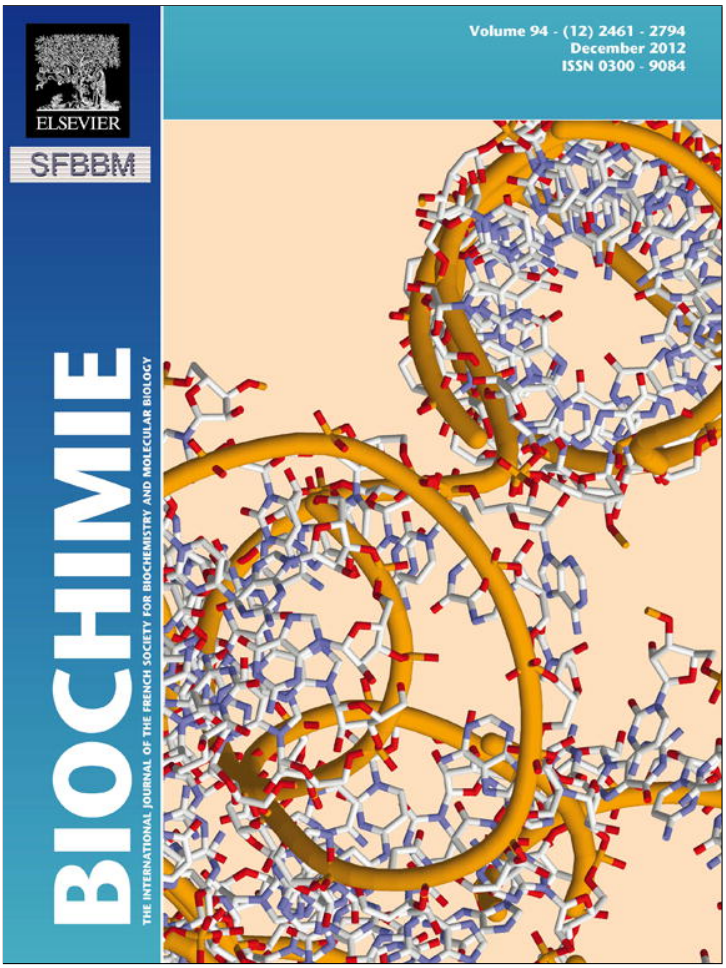

Biochimie 94 (2012) 2706–2715

Contents lists available at SciVerse ScienceDirect

Biochimie

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

Research paper

Stabilization of secondary structure elements by specific combinations of hydrophilic and hydrophobic amino acid residues is more important for proteins encoded by GC-poor genes

Vladislav Victorovich Khrustalev*, Eugene Victorovich Barkovsky

Department of General Chemistry, Belarussian State Medical University, Minsk 220029, Dzerzinskogo 83, Belarus

ARTICLE INFO

Article history:
Received 13 June 2012
Accepted 9 August 2012
Available online 21 August 2012

Keywords:
Secondary structure
Propensity scale
Coil-sheet transition
Helix-sheet transition
Mutational pressure

ABSTRACT

Stabilization of secondary structure elements by specific combinations of hydrophobic and hydrophilic amino acids has been studied by the way of analysis of pentapeptide fragments from twelve partial bacterial proteomes. PDB files describing structures of proteins from species with extremely high and low genomic GC-content, as well as with average G + C were included in the study. Amino acid residues in 78,009 pentapeptides from alpha helices, beta strands and coil regions were classified into hydrophobic and hydrophilic ones. The common propensity scale for 32 possible combinations of hydrophobic and hydrophilic amino acid residues in pentapeptide has been created: specific pentapeptides for helix, sheet and coil were described. The usage of pentapeptides preferably forming alpha helices is decreasing in alpha helices of partial bacterial proteomes with the increase of the average genomic GC-content in first and second codon positions. The usage of pentapeptides preferably forming beta strands is increasing in coil regions and in helices of partial bacterial proteomes with the growth of the average genomic GC-content in first and second codon positions. Due to these circumstances the probability of coil-sheet and helix-sheet transitions should be increased in proteins encoded by GC-rich genes making them prone to form amyloid in certain conditions. Possible causes of the described fact that importance of alpha helix and coil stabilization by specific combinations of hydrophobic and hydrophilic amino acids is growing with the decrease of genomic GC-content have been discussed.

1. Introduction

Secondary structure formation is one of the most extensively studied processes in the field of protein science [1–8]. Amino acid content of alpha helices, beta strands and unstructured regions (coil) have been studied in proteins with determined 3D structures by dozens of researchers. Numerous propensity scales have been proposed for amino acid residues [1,3], for combinations of two amino acid residues (for dipeptides and pairs of amino acids) [4–6] and for tripeptides [7]. In general, they have been confirmed in *in vitro* studies on model peptides [8]. Those propensity scales have brought new information on the theoretical issues of secondary structure formation and also have been used for secondary structure prediction in numerous computer algorithms [1,3].

It is known that amino acid content of proteins highly depends on GC-content of genes coding for them [9]. Symmetric mutational GC-pressure causes frequent AT to GC mutations in genes which result in the growth of the usage of amino acid residues encoded by GC-rich codons [9]. Those amino acid residues are glycine, alanine, arginine and proline (GARP). As we have found out, negative selection controls GARP growth in proteins producing the following asymmetry: the usage of alanine increases due to GC-pressure to much higher levels than usages of glycine, proline and arginine [10,11]. Symmetric mutational AT-pressure causes frequent GC to AT mutations. This process leads to the growth of the usage of amino acid residues encoded by AT-rich codons [9]. Those amino acid residues are phenylalanine, tyrosine, methionine, isoleucine, asparagine and lysine (FYMINK). Due to the negative selection, usages of lysine, isoleucine and asparagine increase under the pressure of GC to AT mutations to much higher levels than usages of tyrosine, phenylalanine and methionine [10,11].

Mutational pressure brings different consequences [12]. For example, mutational GC-pressure increases total hydrophobicity of proteins [13]. This happens due to the decrease of the usage of hydrophilic amino acid residues (Asn and Lys) and the growth of the usage of hydrophobic Ala, Gly and Pro residues [12,13]. However, the usage of strongly hydrophobic amino acid residues (Ile, Tyr and Phe) in proteins also decreases with the growth of G + C in genes coding for them [12,13].

* Corresponding author. Tel.: +375 80172845957.
E-mail address: vvkhrustalev@mail.ru (V.V. Khrustalev).

0300-9084/$ – see front matter © 2012 Elsevier Masson SAS. All rights reserved.
http://dx.doi.org/10.1016/j.biochi.2012.08.008

It is widely accepted that short-distance hydrophobic interactions play important role in secondary structure formation and stabilization [14–16]. Hydrophobic interactions between first and fourth ($i, i+3$), as well as between first and fifth ($i, i+4$) amino acid residues in a polypeptide chain are frequently found in alpha helices [6,14]. Hydrophobic interactions in beta strands are usually found between first and third amino acid residues (i, $i+2$). As a result, hydrophobic amino acid residues can frequently be found in first, third and fifth positions of beta strands ($i, i+2, i+4$) [6,17]. As one can admit, there is at least one kind of periodicity in hydrophobic amino acid residues appearance (i, $i+4$) which is characteristic for both helices and beta strands. In our opinion, the best way to separate ($i, i+4$) periodicity existing in helices from the pattern of hydrophobic amino acid residues appearance often existing in beta strands is to study pentapeptides and not pairs of amino acids.

The aim of the current study was to create propensity scale for pentapeptides composed of hydrophobic and hydrophilic amino acid residues and to find out whether stabilization of secondary structure elements by those specific combinations of hydrophobic and hydrophilic amino acid residues is equally important for proteins encoded by genes with low, average and high GC-content.

To analyze specific periodicities in hydrophobic and hydrophilic amino acid appearance in helices, sheet and coil we developed original methodology. Alpha helices, beta strands and coil regions have been cut down into pentapeptides. Propensity scales have been created for each of the twelve partial proteomes of bacterial species. We used four species with GC-rich genomes, four species with AT-rich genomes and four species with average GC-content in their genomes. We used the term "partial proteome" to highlight that all 3D structures of proteins from each bacterial specie available via Protein Data Bank (www.pdb.org) have been analyzed. Even though tertiary and secondary structures have not been determined for all the proteins forming any complete bacterial proteome yet, the number of proteins with known 3D structures for certain bacteria is enough to create specie-specific propensity scales.

Common propensity scale has been created based on analysis of twelve scales for those partial proteomes. According to this scale, 22 from 32 types of pentapeptides have significant preference to be found in one of the two elements of secondary structure or in coil; 6 from 32 types of pentapeptides have significant preference to be found in two from three possible conformations. There are just 4 absolutely indifferent types of pentapeptides which have no preference to be included in helix, sheet or coil.

Thirty two types of pentapeptides have been classified into five main groups: helix-like pentapeptides, sheet-like pentapeptides, coil-like pentapeptides, coil/helix pentapeptides and indifferent pentapeptides. We calculated total levels of usage for each group of pentapeptides in helices, beta strands and regions of coil for partial proteomes studied. Analysis of their usages led us to the conclusion that specific stabilization by combinations of hydrophilic and hydrophobic amino acid residues is more important for helices and coil regions from proteins encoded by AT-rich genes.

2. Material and methods

2.1. Material

Since the main aim of the present study was to estimate the influence of symmetric mutational pressure on secondary structure stabilization by specific combinations of hydrophobic and hydrophilic amino acid residues, we have collected files from Protein Data Bank (www.pdb.org) describing 3D and secondary structures of proteins encoded by AT-rich and GC-rich bacterial species, as well as from those with average G + C. Bacterial species with extremely AT-rich genomes included in this study are: *Borrelia burgdorferi*; *Clostridium perfringens*; *Francisella tularensis* and *Staphylococcus aureus*. Studied GC-rich bacterial species are: *Mycobacterium tuberculosis*; *Rhodococcus jostii*; *Xanthomonas campestris* and *Streptomyces coelicolor*. Four bacterial species the genomes of which have average levels of G + C are: *Yersinia pestis*; *Porphyromonas gingivalis*; *Shigella flexneri*; *Synechococcus elongatus*. Lists of PDB files for all the proteins studied can be found in Supplementary material in form of MS Excel table entitled: "Partial proteomes". We used all the available PDB files for each of the 12 bacterial species studied. There are no different PDB files describing the same protein in our data base. However, several exceptions were made for PDB files with information on different regions of the same protein. Total number of PDB files studied is equal to 615. The volume of data analyzed in this study is represented in the Table 1.

Average GC-content in first and second codon positions "(1GC + 2GC)/2" [9] have been calculated in completely sequenced genomes of reference strains for each of the bacterial species studied (names of reference strains can be found in Table 1). To perform the abovementioned calculation we used information on codon usage in each open reading frame from those reference genomes deposited in the Codon Usage Database (www.kazusa.or.jp/codon) [18].

2.2. Methods

Information on boarders of alpha helices and beta strands was extracted directly from PDB files. All the regions which are not included in alpha helices and in beta strands go in this study under the name "coil".

Table 1
Information on the material used in the present work.

Specie	Strain	G + C	(1GC + 2GC)/2	Number of PDB files	Distribution of 78,009 pentapeptides studied		
					Helix	Sheet	Coil
Borrelia burgdorferi	B31	0.284	0.322	18	1200	228	556
Clostridium perfringens	str. 13	0.291	0.353	33	2060	1075	1570
Francisella tularensis	FSC198	0.326	0.381	33	3176	494	1421
Staphylococcus aureus	Mu50	0.329	0.380	80	6350	1621	3428
Yersinia pestis	KIM	0.479	0.475	76	5666	1274	2435
Porphyromonas gingivalis	W83	0.487	0.462	46	3202	848	1300
Shigella flexneri	2a str. 301	0.513	0.496	62	4226	691	1508
Synechococcus elongatus	PCC 7942	0.558	0.537	36	2726	406	1705
Xanthomonas campestris	ATCC 33913	0.652	0.574	36	2291	293	600
Mycobacterium tuberculosis	H37Rv	0.650	0.584	80	6285	1095	3046
Rhodococcus jostii	RHA1	0.673	0.589	33	2983	293	600
Streptomyces coelicolor	A3(2)	0.721	0.620	82	7661	886	2810
Total				615	47,826	9204	20,979

Each alpha helix, beta strand and coil region has been cut down into pentapeptides. Sliding window methodology with a step equal to one amino acid residue has been used for this operation. It means that "$X - 4$" pentapeptides have been collected from each element of secondary structure or from coil region X amino acids in length ($X \geq 5$). Short alpha helices, beta strands and coil regions have not been included in the present study.

According to the Eisenberg hydrophobicity scale [14] amino acids in pentapeptides were roughly divided into two groups. Hydrophilic amino acids (arginine, lysine, aspartic and glutamic acids, asparagine, glutamine, serine, threonine and histidine) are designated by the letter "W" (water). Hydrophobic amino acids (glycine, proline, alanine, valine, isoleucine, leucine, methionine, tyrosine, phenylalanine, cysteine and tryptophan) are designated by the letter "O" (oil). Since there can be just two types of amino acids (W and O) in each pentapeptide, then there are $2^5 = 32$ possible types of pentapeptides. Probabilities to be included in alpha helix, in beta sheet and in coil regions for each type of those 32 pentapeptides have been calculated. The probability to be included in alpha helix for the given type of pentapeptide is equal to the usage of this pentapeptide in alpha helix divided by the sum of its usages in alpha helix, in beta sheet and in coil. Probabilities to be included in beta sheet and in coil were calculated in a similar way. Those calculations have been performed separately for proteins from each partial bacterial proteome. Significance of the preference to be found in certain type of secondary structure has been checked for each type of pentapeptide within twelve groups of proteins by two-tailed *t*-test.

There are five main groups of pentapeptides: seven pentapeptides have significant preference to be included in alpha-helix (helix-like pentapeptides); nine pentapeptides have significant preference to be included in beta-sheet (sheet-like pentapeptides); six pentapeptides have significant preference to be included in coil (coil-like pentapeptides); four pentapeptides have significant preference to be included in either coil or alpha-helix and not in beta-sheet (coil/helix pentapeptides); four pentapeptides have no significant preference to be included in certain type of secondary structure (indifferent pentapeptides). There is also a single pentapeptide with significant preference to be included in coil or in sheet and not in alpha-helix. Remaining single pentapeptide has a preference to be included in either sheet or alpha-helix and not in coil.

Total usages of helix-like, sheet-like, coil-like, coil/helix and indifferent pentapeptides have been calculated in a given type of secondary structure (or in coil) for each of the twelve partial bacterial proteomes. Then coefficients of correlation of those total usages of specific pentapeptides in certain types of secondary structure on average genomic GC-content in first and second codon positions "(1GC + 2GC)/2" have been calculated.

GC-content in first (1GC) and second (2GC) codon positions has been calculated for each open reading frame identified in completely sequenced reference genomes of twelve bacterial species by the original "Coding Genome Scanner" algorithm (www.barkovsky.hotmail.ru) [11]. Then sums of those indexes have been divided by two. The resulting index "(1GC + 2GC)/2" is the measure of the influence of the symmetric mutational pressure on amino acid usage [9]. This index, unlike the total GC-content (G + C), ignores the influence of mutational pressure on nucleotide usage in third codon positions (3GC). Generally speaking, the level of "(1GC + 2GC)/2" demonstrates the direction and strength of symmetric mutational pressure which had existed in evolutionary predecessors of a given genome during the long period of time [11]. In contrast, the level of 3GC demonstrates the direction and strength of symmetric mutational pressure existed in more recent period of time [11,19]. It is known that during the evolutionary history of some prokaryotic genomes, such as *Haloquadratum walsbyi*, GC-pressure has been changed to AT-pressure. As a result, 3GC and G + C levels in its genes decreased much more than levels of "(1GC + 2GC)/2" [11].

To find out what amino acids are usually stabilized by specific combinations of hydrophobic and hydrophilic ones in alpha-helices, beta-strands and coil regions we calculated levels of helix-like, sheet-like and coil-like pentapeptides containing each amino acid in helix, sheet and coil. Then levels of helix-like, sheet-like and coil-like pentapeptides containing a given amino acid have been compared with each other by two-tailed *t*-test in helix, sheet and coil. Moreover, coefficients of correlation between those levels of specific pentapeptides on "(1GC + 2GC)/2" have been calculated.

As a result, we found out amino acids which usually exist in alpha-helices, beta-strands and coil regions in "stabilized" and "destabilized" states. Those amino acids which are usually "stabilized" in helices and coil regions of proteins encoded by AT-rich genes and "destabilized" in helices and coil regions of proteins encoded by GC-rich genes have also been found out.

3. Results

3.1. Propensity scale with probabilities to be included in helix, sheet and coil for 32 types of pentapeptides composed of hydrophobic (O) and hydrophilic (W) amino acids

There are 32 types of pentapeptides containing different combinations of hydrophilic (W) and hydrophobic (O) amino acids. As one can see in Fig. 1, the combination of hydrophilic and hydrophobic amino acid residues inside a given pentapeptide has a great impact on its secondary structure.

As one can see in Fig. 1 and Table 2, pentapeptides composed of hydrophilic amino acid residues, in general, have a preference to be found in unstructured regions of a protein (in random coil). This statement is correct for completely hydrophilic pentapeptides (WWWWW) and for three from five almost hydrophilic pentapeptides (OWWWW; WWWOW; WWWWO). It is important to highlight that one of the almost hydrophilic pentapeptides (WWOWW) has significantly higher probability to be included in alpha helix than in coil or in beta sheet (*P*-values can be found in Table 2). As to another almost hydrophilic pentapeptide (WOWWW), the probability to be included in helix is significantly higher for it than that to be included in sheet, probability to be included in coil is also significantly higher than the probability to be included in sheet, while the difference between probabilities to be included in coil and helix is not significant (see Table 2). That is why WOWWW pentapeptide has been classified as "coil/helix pentapeptide".

Pentapeptides composed of five hydrophobic amino acid residues (OOOOO) have a significant preference to be found in beta sheet (see Fig. 1 and Table 2), as well as four from five pentapeptides composed of four hydrophobic amino acids and single hydrophilic residue (WOOOO, OWOOO, OOOWO, OOOOW). This finding is in consistence with previous suggestions that hydrophobicity plays important role in beta sheet formation [20–22]. However, our results showed that one from those five almost hydrophobic pentapeptides (OOWOO) has significantly higher probability to be found in alpha helix than in beta sheet or in coil.

Four specific pentapeptides with average hydrophobicity have probabilities to be found in alpha helix which are higher than 50% (see Fig. 1). Those pentapeptides contain either two hydrophobic residues situated near each other and surrounded by hydrophilic ones (WOOWW and WWOOW), or two hydrophilic residues situated near each other and surrounded by hydrophobic ones (OWWOO and OOWWO). In two last pentapeptides "$i - i + 3$" and "$i - i + 4$" hydrophobic interactions stabilizing helical

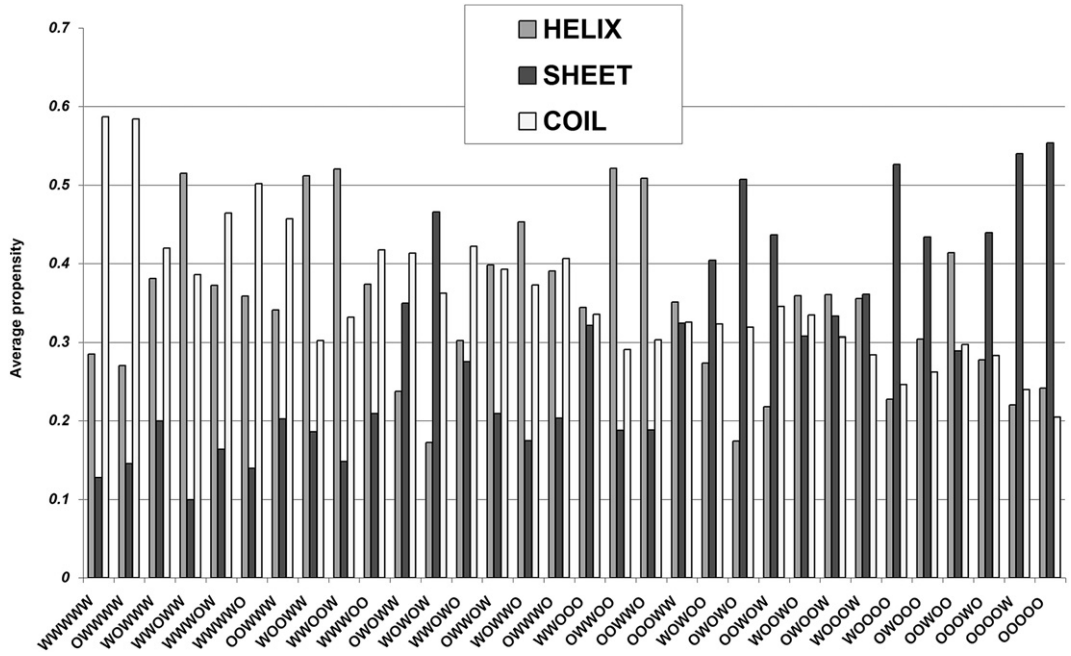

Fig. 1. Propensity scale with average probabilities to be included in alpha helix, beta sheet and coil for pentapeptides composed of hydrophilic (W) and hydrophobic (O) amino acid residues for proteins from twelve partial bacterial proteomes. *P*-values for differences between probabilities can be found in Tables 2 and 3.

conformation are allowed. In first two pentapeptides "$i - i + 3$" and "$i - i + 4$" interactions between hydrophilic amino acid residues should stabilize alpha helix.

Interestingly, WOWWO pentapeptide demonstrates significant preference to be included in helix, while OWWOW (its reversed repeat) is a coil/helix pentapeptide (see Table 3).

In general, pentapeptides preferably forming alpha helix usually have composition featuring "OO" and/or "WW" motifs. It means that in alpha helices two "OO" motifs are often separated either by "WW" motif or by a single hydrophilic amino acid residue; two "WW" motifs are often separated by "OO" motif or by a single hydrophobic amino acid residue.

Four amphiphilic pentapeptides (WOWOW, OWOWO, WOWOO and OOWOW) have significant preferences to be found in beta-sheet than in helix or coil (see Fig. 1 and Table 3). The presence of "OWO" motif makes those pentapeptides prone to form beta-strand due to the increased probability of hydrophobic interactions between the first and the third amino acid residues in that "OWO" motif. On the other hand, there are several pentapeptides containing that "OWO" motif which have no significant preference to be found in beta sheet (see Discussion section).

These data confirm that beta strands can be roughly divided into two types. The first type of beta strands is formed by hydrophobic regions of a protein which lack any regular periodicity of hydrophilic and hydrophobic amino acids appearance (see Table 1). The second type of beta sheet is formed by amphiphilic regions which have very specific regular periodicity (see Table 2). The difference between helix-like and amphiphilic sheet-like peptides is in the type of periodical changes in amino acid hydrophobicity along the polypeptide chain (see Fig. 1).

Pentapeptides with WWOWO and OOWWW compositions are coil pentapeptides, while WWWOO, OWWWO and OWWOW are coil/helix pentapeptides.

Pentapeptides with WWOOO, OOOWW, WOOWO and OWOOW compositions show no significant preferences at all (see Table 3), so they are classified as indifferent ones.

Probabilities to be found in alpha helix and in beta sheet are higher than probability to be found in coil for pentapeptide with WOOOW composition (see Table 3). Probabilities to be found in coil and beta strand are higher than probability to be found in helix for OWOWW pentapeptide.

In this section we described the propensity scale for pentapeptides composed of hydrophobic and hydrophilic amino acid residues which is common for proteins encoded by both AT-rich and GC-rich genes. We confirmed the fact that secondary structure of proteins is usually stabilized by short-distance hydrophobic

Table 2
Preferable states of secondary structure for hydrophilic and hydrophobic pentapeptides. *P*-values for differences in propensities to be found in helix, sheet and coil are given.

Pentapeptide	WWWWW	OWWWW	WOWWW	WWOWW	WWWOW	WWWWO
Preferable state	Coil	Coil	Coil/helix	Helix	Coil	Coil
Helix vs. sheet	0.00174	0.00120	0.00130	2.9×10^{-7}	9.2×10^{-6}	4.6×10^{-5}
Helix vs. coil	0.00026	1.2×10^{-6}	0.21169	0.00123	0.00604	0.00329
Sheet vs. coil	4.8×10^{-6}	8.2×10^{-8}	0.00040	5.4×10^{-7}	5.0×10^{-7}	3.0×10^{-6}
Pentapeptide	WOOOO	OWOOO	OOWOO	OOOWO	OOOOW	OOOOO
Preferable state	Sheet	Sheet	Helix	Sheet	Sheet	Sheet
Helix vs. sheet	6.0×10^{-7}	8.2×10^{-5}	3.8×10^{-5}	7.8×10^{-7}	1.7×10^{-7}	1.7×10^{-6}
Helix vs. coil	0.05462	0.04782	0.00062	0.68608	0.11396	0.01834
Sheet vs. coil	1.2×10^{-6}	1.3×10^{-5}	0.71997	5.0×10^{-5}	3.9×10^{-7}	7.7×10^{-7}

Table 3
Preferable states of secondary structure for pentapeptides with average hydrophobicity. *P*-values for differences in propensities to be found in helix, sheet and coil are given.

Pentapeptide	OOWWW	WOOWW	WWOOW	WWWOO	OWWWO
Preferable state	Coil	Helix	Helix	Coil / helix	Coil / helix
Helix vs. sheet	1.0×10^{-5}	1.1×10^{-7}	9.8×10^{-9}	5.0×10^{-5}	8.8×10^{-5}
Helix vs. coil	0.00039	4.4×10^{-7}	2.7×10^{-5}	0.05452	0.48200
Sheet vs. coil	4.0×10^{-6}	0.00076	8.3×10^{-6}	6.4×10^{-5}	9.0×10^{-5}
Pentapeptide	OWOWW	WOWOW	WWOWO	OWWOW	WOWWO
Preferable state	Coil / sheet	Sheet	Coil	Coil / helix	Helix
Helix vs. sheet	0.03123	1.2×10^{-6}	0.41958	2.1×10^{-5}	1.8×10^{-7}
Helix vs. coil	5.6×10^{-5}	2.2×10^{-7}	0.00154	0.79071	0.00414
Sheet vs. coil	0.18462	0.00449	0.00601	3.0×10^{-5}	5.0×10^{-6}
Pentapeptide	OOWWO	OWWOO	WOWOO	OWOWO	OOWOW
Preferable state	Helix	Helix	Sheet	Sheet	Sheet
Helix vs. sheet	9.6×10^{-10}	8.4×10^{-8}	0.00012	9.9×10^{-8}	1.3×10^{-6}
Helix vs. coil	7.7×10^{-9}	3.7×10^{-7}	0.00178	2.0×10^{-6}	7.6×10^{-5}
Sheet vs. coil	6.2×10^{-5}	0.00013	0.02518	4.7×10^{-5}	0.00311
Pentapeptide	WWOOO	OOOWW	WOOWO	OWOOW	WOOOW
Preferable state	Indifferent	Indifferent	Indifferent	Indifferent	Sheet / helix
Helix vs. sheet	0.45056	0.29399	0.11730	0.36528	0.88791
Helix vs. coil	0.72089	0.29392	0.29831	0.07500	0.00364
Sheet vs. coil	0.67902	0.96234	0.40957	0.48786	0.01774

interactions, as well as by interactions between hydrophilic amino acids (i.e. by polar or ionic interactions). We also came to the conclusion that those specific interactions are visualized well by the way of the analysis of pentapeptide propensities. The next step of our analysis led us to the conclusion that stabilization of alpha helices and coil regions by specific combinations of hydrophobic and hydrophilic amino acid residues is especially important for proteins encoded by AT-rich genes.

3.2. The usage of helix-like pentapeptides in helices shows strong inversed correlation on average GC-content in first and second codon positions

According to our results, there are seven helix-like pentapeptides: WWOWW, OOWOO, WOOWW, WWOOW, OWWOO, OOWWO and WOWWO. We calculated their total level of usage in helices, beta-strands and coil regions in partial proteomes of each of the twelve bacterial species. In Fig. 2 one can see that their total level of usage in helices is higher than in coil and much higher than in beta strands. Modules of coefficients of correlation between average GC-content in first and second codon positions "(1GC + 2GC)/2" and levels of helix-like pentapeptides in sheet and coil are low, unlike that for their level in alpha helices (see Fig. 2). It means that the usage of pentapeptides stabilized by specific combinations of hydrophilic and hydrophobic amino acid residues is growing in helices under the influence of mutational AT-pressure and decreasing under the influence of mutational GC-pressure.

3.3. The usage of sheet-like pentapeptides in helices and coil regions shows strong correlation on average GC-content in first and second codon positions

Total usage of nine pentapeptides with significant preference to be found in beta-strands (WOWOW, OWOWO, WOWOO, OOWOW, WOOOO, OWOOO, OOOWO, OOOOW and OOOOO) is growing in alpha helices and coil regions with the increase of the average GC-content in first and second codon positions (see Fig. 3). The level of "stabilized" pentapeptides in beta strands shows week dependence on "(1GC + 2GC)/2": it remains high in proteins encoded by both AT-rich and GC-rich genes.

The usage of pentapeptides which are not protected from coil to sheet and helix to sheet transitions by specific combinations of hydrophobic and hydrophilic amino acid residues is high in proteins encoded by GC-rich genes. Does it mean that proteins encoded by genes under the influence of mutational GC-pressure are at a higher risk of those transitions?

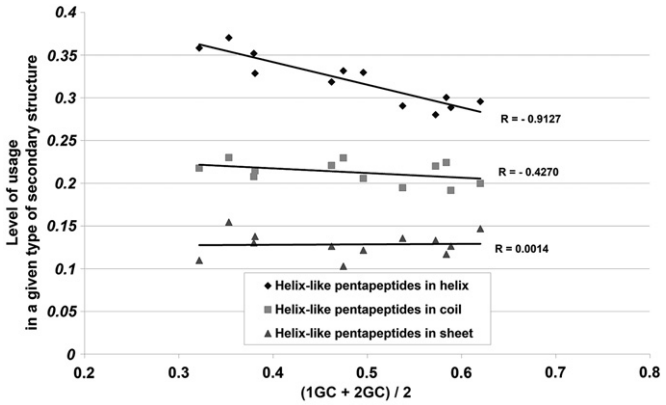

Fig. 2. Dependences between average genomic GC-content in first and second codon positions "(1GC + 2GC)/2" and total level of helix-like pentapeptides in helices, beta strands and coil regions of partial bacterial proteomes.

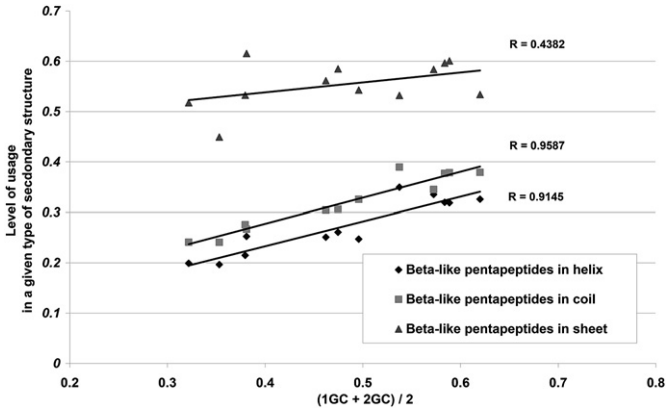

Fig. 3. Dependences between average genomic GC-content in first and second codon positions "(1GC + 2GC)/2" and total level of sheet-like pentapeptides in helices, beta strands and coil regions of partial bacterial proteomes.

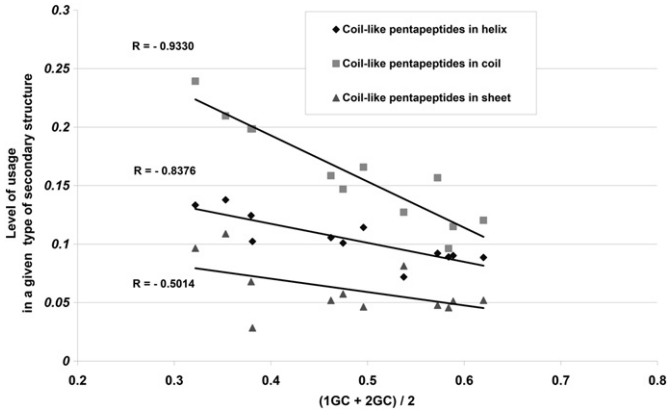

Fig. 4. Dependences between average genomic GC-content in first and second codon positions "(1GC + 2GC)/2" and total level of coil-like pentapeptides in helices, beta strands and coil regions of partial bacterial proteomes.

3.4. The usage of coil-like pentapeptides in coil regions shows strong inversed correlation on average GC-content in first and second codon positions

In Fig. 4 one can see that the total usage of six pentapeptides with significant preference to be included in coil (WWWWW, OWWWW, WWWOW, WWWWO, OOWWW and WWOWO) is growing in coil regions under the influence of mutational AT-pressure. It means that regions of coil in proteins encoded by GC-rich genes contain less "stabilized" pentapeptides than regions of coil in proteins encoded by GC-poor genes.

Total usage of coil-like pentapeptides under the influence of mutational AT-pressure is growing in alpha helices as well, however, under the lower slope than that in coil regions (see Fig. 4). This fact allows us to suggest that the usage of pentapeptides in alpha helices which are not protected from helix to coil transitions increases under the influence of AT-pressure.

3.5. The usage of coil/helix pentapeptides in coil regions and helices shows strong inversed correlation on average GC-content in first and second codon positions

Total usage of four coil/helix pentapeptides (WOWWW, WWWOO, OWWWO and OWWOW) is growing under the influence of mutational AT-pressure in coil regions under the higher slope than in alpha helices (see Fig. 5). This fact is also supporting our suggestion that the risk of helix to coil transitions is growing with the increase of mutational AT-pressure.

3.6. The usage of indifferent pentapeptides in coil regions and helices shows strong correlation on average GC-content in first and second codon positions

On the other hand, the risk of helix-coil, as well as other kinds of transitions, is growing under the influence of mutational GC-pressure too, since total level of indifferent pentapeptides (WWOOO, OOOWW, WOOWO and OWOOW) in coil and helix shows direct linear dependence on average level of GC-content in first and second codon positions (see Fig. 6).

In general, total usage of "stabilized" pentapeptides (total usage of helix-like pentapeptides in helix, sheet-like pentapeptides in sheet and coil-like pentapeptides in coil) shows inverse correlation (coefficient of correlation is equal to −0.888) on "(1GC + 2GC)/2". Mutational GC-pressure leads to the decrease of the importance of secondary structure stabilization by specific combinations of hydrophobic and hydrophilic amino acids. To understand why it happens we calculated levels of usage for pentapeptides containing each of the 20 amino acids in stabilized and destabilized states in helices, beta-strands and coil regions of the twelve partial bacterial proteomes.

3.7. Alanine is the unique helix-former which preferably exists in alpha-helices in sheet-like pentapeptides

We calculated levels of usage of helix-like and sheet-like pentapeptides containing each amino acid residue in alpha helices. Then we applied paired two-tailed *t*-test to find out what amino acid residues are found in alpha helices in helix-like pentapeptides more frequently than in sheet-like pentapeptides. As one can see in Table 4, all the nine hydrophilic amino acid residues (Asp, Glu, Arg, Lys, His, Asn, Gln, Thr and Ser) preferably enter alpha helices in specific helix-like pentapeptides. In other words, hydrophilic amino acids are usually stabilized in alpha helices. On the other hand, levels of usage in those stabilized pentapeptides for Glu, Asp, Lys, Asn and Ser show inverse correlation on "(1GC + 2GC)/2" (see Table 4). Levels of their usage in destabilized sheet-like pentapeptides show direct dependence on "(1GC + 2GC)/2" (see Table 4).

Good example of the influence of mutational pressure on levels of stabilized pentapeptides in helices is shown in Figure S1 from Supplementary material. Aspartic acid, being a known helix-

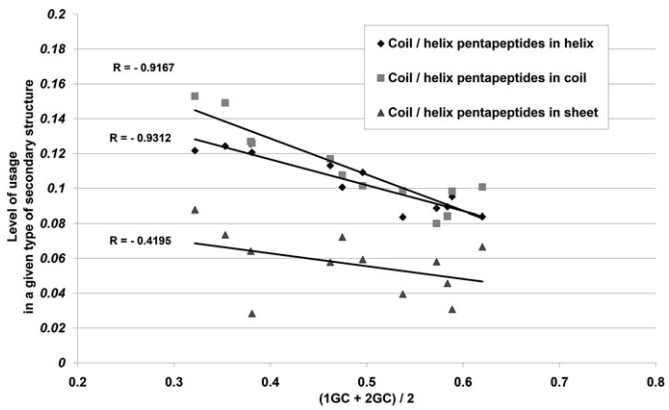

Fig. 5. Dependences between average genomic GC-content in first and second codon positions "(1GC + 2GC)/2" and total level of coil/helix pentapeptides in helices, beta strands and coil regions of partial bacterial proteomes.

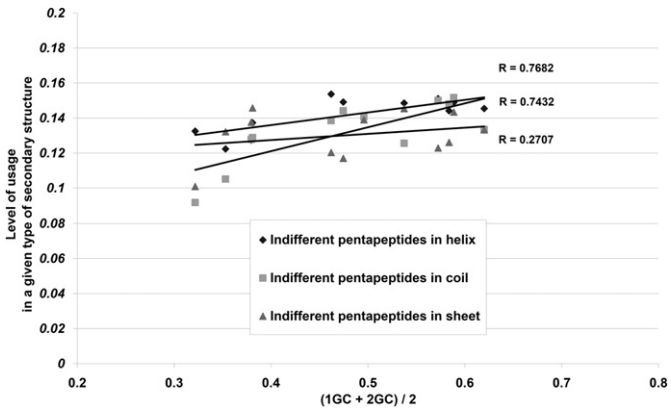

Fig. 6. Dependences between average genomic GC-content in first and second codon positions "(1GC + 2GC)/2" and total level of indifferent pentapeptides in helices, beta strands and coil regions of partial bacterial proteomes.

Table 4
Average differences between levels of helix-like and sheet-like pentapeptides containing each of the 20 amino acid residues in alpha helices.

AA	Average difference between levels of helix-like and sheet-like pentapeptides	*P*-value	Coefficient of correlation between "(1GC + 2GC)/2" and level of helix-like pentapeptides in helix	Coefficient of correlation between "(1GC + 2GC)/2" and level of sheet-like pentapeptides in helix	Preferred state in alpha helix
Thr	0.155 ± 0.027	2.26×10^{-7}	−0.3981	0.8151	Stabilized
Glu	0.231 ± 0.038	1.12×10^{-7}	−0.8943	0.9100	Stabilized
Gln	0.194 ± 0.028	3.23×10^{-8}	−0.2051	0.7525	Stabilized
Lys	0.218 ± 0.029	1.56×10^{-8}	−0.7281	0.8566	Stabilized
Arg	0.190 ± 0.030	9.00×10^{-8}	−0.3767	0.8284	Stabilized
Asn	0.193 ± 0.042	2.15×10^{-6}	−0.6703	0.7284	Stabilized
Asp	0.222 ± 0.033	4.31×10^{-8}	−0.8310	0.9034	Stabilized
Ser	0.186 ± 0.029	8.63×10^{-8}	−0.6249	0.7736	Stabilized
His	0.125 ± 0.052	0.0006	−0.0694	0.6925	Stabilized
Ala	−0.102 ± 0.043	0.0007	−0.6495	0.7835	Destabilized
Gly	−0.225 ± 0.045	8.91×10^{-7}	−0.6596	0.7022	Destabilized
Pro	−0.277 ± 0.036	9.53×10^{-9}	0.0708	0.0343	Destabilized
Val	−0.012 ± 0.057	0.6958	−0.8253	0.8470	G + C-dependent
Ile	0.037 ± 0.058	0.2385	−0.7676	0.8512	G + C-dependent
Phe	−0.017 ± 0.071	0.6561	−0.7381	0.8049	G + C-dependent
Leu	0.005 ± 0.054	0.8586	−0.9133	0.9191	G + C-dependent
Met	−0.048 ± 0.051	0.0955	−0.8010	0.8398	G + C-dependent
Tyr	−0.033 ± 0.036	0.1017	−0.5443	0.8341	G + C-dependent
Trp	−0.043 ± 0.065	0.2189	0.2426	0.8286	Indifferent
Cys	−0.061 ± 0.074	0.1339	0.1510	0.3961	Indifferent

breaker [1], may enter alpha-helices preferably in specific helix-like pentapeptides. Aspartic acid is encoded by codons with average GC-content in two first positions and so its level of usage is approximately the same in both proteins encoded by AT-rich and GC-rich genes [10]. Despite this, the level of Asp in destabilized sheet-like pentapeptides is growing under the influence of GC-pressure in alpha helices. This effect can be explained only by the influence of GC-pressure on levels of other amino acid residues usages.

Six hydrophobic amino acid residues (Val, Ile, Phe, Tyr, Leu and Met) are preferably stabilized in helices by helix-like compositions of pentapeptides only in proteins encoded by AT-rich genes (see Table 4). Those amino acid residues enter helices of proteins encoded by GC-rich mostly in sheet-like pentapeptides. Levels of helix-like and sheet-like pentapeptides containing Val (a known strong sheet-former [1]) in helices are shown in Figure S2 from Supplementary material. The usage of valine itself does not depend on GC-content [10], so all the changes represented in Figure S2 from Supplementary material are due to the influence of mutational pressure on other amino acid residues usages.

Glycine and proline are preferably used in sheet-like pentapeptides in helices (see Table 4). This fact can be interpreted as the consequence of the known position-specific propensities of these two strong helix breakers [23]. Proline is used mostly in first four positions of alpha helices [23]. Glycine is used mostly in first and last positions [23]. That is why some specificity of N-terminal and C-terminal pentapeptides containing Pro and Gly may be responsible for their odd behaviour.

Alanine is a strong helix former [1] with no known position preferences – it is frequently used in both terminals and in the middle of helices. However, this amino acid residue is usually found in sheet-like and not helix-like pentapeptides in alpha helices. The level of alanine usage reaches high levels in helices of proteins encoded by GC-rich genes (up to 17%), so its features should make the greatest contribution into the increase of sheet-like pentapeptides in helices of those proteins. In our opinion, alanine, unlike other helix-formers, can successfully promote formation of alpha-helices being a part of both helix-like and (especially) sheet-like pentapeptides. Due to the growth of alanine usage in helices of proteins encoded by GC-rich genes specific pattern of hydrophobic and hydrophilic amino acids appearance periodicity begins to lose its importance. According to the data from Table 4, hydrophobic amino acids may also be used in those sheet-like pentapeptides which contain alanine. Total level of that kind of pentapeptides (featuring alanine and hydrophobic amino acids) increases with the growth of GC-content in alpha helices.

3.8. Hydrophilic amino acids preferably exist in coil-like pentapeptides only in proteins encoded by AT-rich genes

Levels of usage of coil-like and sheet-like pentapeptides containing each amino acid residue have been calculated in regions of coil. Paired two-tailed *t*-test has been applied to find out what amino acid residues are found in coil in "stabilized" coil-like pentapeptides more frequently than in sheet-like pentapeptides. Interestingly, there is only one amino acid residue (His) which is preferably stabilized in coil of both proteins encoded by AT-rich and GC-rich genes (see Table 5). However, total level of histidine usage is usually low [10,11]. All the other hydrophilic amino acid residues are stabilized in coil by specific (in general, hydrophilic) pentapeptides of proteins encoded by AT-rich genes. For example, glutamic acid (a well-known strong helix former [1]) exists in coil of proteins encoded by AT-rich genes mostly in coil-like pentapeptides, while in coil of proteins encoded by GC-rich genes this amino acid exists mostly in sheet-like pentapeptides (see Supplementary material, Figure S3).

Hydrophobic amino acids are not stabilized in coil (see Table 5), while levels of coil-like pentapeptides containing them increase in regions of coil under the influence of AT-pressure. Indeed, such helix former, as leucine [1], is found in coil-like pentapeptides in coil of proteins encoded by AT-rich genes more frequently than in coil-like pentapeptides in coil of proteins encoded by GC-rich genes (see Supplementary material, Figure S4). In our opinion, hydrophobic amino acids enter coil mostly in case if they are situated near such strongest coil formers as glycine and proline. Those coil formers stimulate formation of coil even in case if they are surrounded by hydrophobic amino acid residues in sheet-like pentapeptides. Levels of glycine and proline are growing under the influence of GC-pressure [9–11]. This process causes the increase of sheet-like pentapeptides usage in coil of proteins encoded by GC-

Table 5
Average differences between levels of coil-like and sheet-like pentapeptides containing each of the 20 amino acid residues in coil regions.

AA	Average difference between levels of coil-like and sheet-like pentapeptides	*P*-value	Coefficient of correlation between "(1GC + 2GC)/2" and level of coil-like pentapeptides in coil	Coefficient of correlation between "(1GC + 2GC)/2" and level of sheet-like pentapeptides in coil	Preferred state in coil regions
His	0.0584 ± 0.0494	0.0407	−0.7955	0.5927	Stabilized
Val	−0.3316 ± 0.0440	1.34×10^{-8}	−0.8746	0.8466	Destabilized
Ile	−0.2693 ± 0.0575	1.73×10^{-6}	−0.9033	0.6688	Destabilized
Phe	−0.3124 ± 0.0509	1.14×10^{-7}	−0.5451	0.9089	Destabilized
Tyr	−0.3520 ± 0.0476	1.62×10^{-8}	−0.7450	0.7306	Destabilized
Trp	−0.2854 ± 0.0795	2.17×10^{-5}	−0.7870	0.8699	Destabilized
Cys	−0.3534 ± 0.0903	9.75×10^{-6}	−0.4262	0.4655	Destabilized
Leu	−0.2935 ± 0.0515	2.44×10^{-7}	−0.9194	0.9514	Destabilized
Met	−0.3501 ± 0.0441	7.78×10^{-9}	−0.0889	0.7490	Destabilized
Ala	−0.3488 ± 0.0495	2.70×10^{-8}	−0.8981	0.8440	Destabilized
Gly	−0.3462 ± 0.0456	1.23×10^{-8}	−0.7827	0.8382	Destabilized
Pro	−0.3570 ± 0.0462	1.03×10^{-8}	−0.8276	0.8462	Destabilized
Glu	0.0196 ± 0.0642	0.5613	−0.9102	0.8909	G + C-dependent
Gln	0.0210 ± 0.0447	0.3769	−0.7423	0.4052	G + C-dependent
Lys	0.0240 ± 0.0499	0.3654	−0.8423	0.4935	G + C-dependent
Arg	0.0388 ± 0.0489	0.1482	−0.7868	0.7258	G + C-dependent
Asn	0.0354 ± 0.0579	0.2561	−0.7046	0.7942	G + C-dependent
Asp	0.0001 ± 0.0546	0.9988	−0.8210	0.9211	G + C-dependent
Ser	0.0342 ± 0.0477	0.1872	−0.8880	0.6767	G + C-dependent
Thr	0.0147 ± 0.0548	0.6102	−0.8056	0.8730	G + C-dependent

rich genes. When usages of proline and glycine decrease under the influence of AT-pressure, hydrophilicity becomes the main signature of coil regions.

4. Discussion

In the present work the common propensity scale for pentapeptides composed of hydrophilic and hydrophobic amino acid residues has been created. The scale based on 32 types of pentapeptides has many benefits in comparison with scales based on pairs of amino acid residues. For example, it is known that the pair of hydrophobic amino acids separated from each other by hydrophilic residue ("OWO" motif) has a preference to be included in beta-strand [6]. According to our results, six pentapeptides (OWOOO, OOOWO, WOWOW, WOWOO, OOWOW, OWOWO) containing this motif really have significant preferences to be included in beta-sheet. However, WWOWO pentapeptide also containing "OWO" motif has a significant preference to be included in coil, while OOWOO pentapeptide has a significant preference to be found in alpha helix and OWOWW pentapeptide has equal propensities to be found in coil and sheet. Moreover, two other pentapeptides with the same motif (WOOWO and OWOOW) are totally indifferent.

According to our propensity scale, there are seven specific pentapeptides with significant preferences to be included in alpha helices. Under the influence of mutational AT-pressure the usage of those helix-like pentapeptides increases in helices. It means that alpha helices of proteins encoded by AT-rich genes are stabilized by specific combinations of hydrophobic and hydrophilic amino acids better than proteins encoded by GC-rich genes. Specific pattern of hydrophobic and hydrophilic amino acids occurrence in helices should be the product of mutational AT-pressure controlled by natural selection eliminating those amino acid substitutions which destroy alpha helices.

Mutational AT-pressure leads to the increase of FYMINK amino acids usage [9–11]. Those six amino acids are encoded by GC-poor codons. Usages of isoleucine, asparagine and lysine reach especially high levels in proteins encoded by AT-rich genes [10,11]. As one can see in Fig. 7, isoleucine usage is increasing in alpha helices with the decrease of average GC-content in first and second codon positions. This happens even though isoleucine is a strong sheet former [1]. The usage of such strong coil former as asparagine [1] is also growing in helices under the influence of AT-pressure (see Fig. 7). In our opinion, negative selection usually eliminates mutations leading to the appearance of isoleucine and asparagine in helices in case if they appear in "wrong" positions and so destabilize helical conformation. Since mutations leading to the appearance of Ile and Asn in helices are frequent under the influence of AT-pressure sooner or later they take place in "right" positions in which they do not destroy helical conformation and have a chance to be fixed. As a result, the pattern of hydrophilic and hydrophobic amino acids periodicity in helices becomes much more important for stabilization of their conformation.

According to our results, alanine is found in alpha helices in "sheet-like" pentapeptides significantly more frequently than in "helix-like" pentapeptides. This fact is the evidence of outstanding ability to promote formation of alpha helix characteristic to alanine residues. The usage of GARP amino acid residues encoded by GC-rich codons is growing under the influence of GC-pressure [9–11]. Especially high level of usage in proteins encoded by GC-rich genes is characteristic to alanine (see Fig. 7). It is not surprising

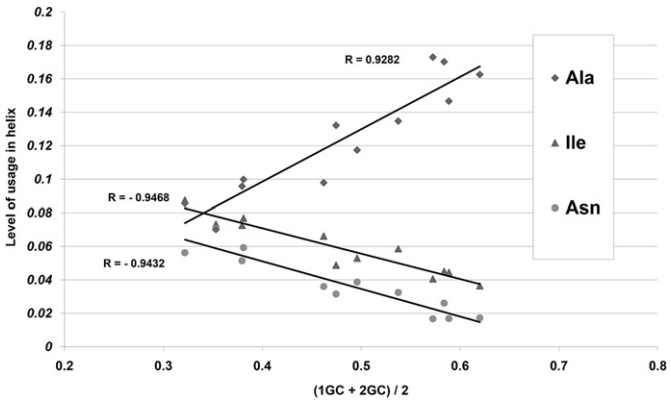

Fig. 7. Dependences between average genomic GC-content in first and second codon positions "(1GC + 2GC)/2" and levels of alanine, isoleucine and asparagine in helices from partial bacterial proteomes.

that the usage of alanine is growing in alpha helices to much higher levels than in sheet and coil [13]. Since alanine is able to enter helices in both "write" and (especially) "wrong" positions, the pattern of hydrophilic and hydrophobic amino acids periodicity in helices becomes much less important in proteins encoded by GC-rich genes. Negative selection allows substitutions of both hydrophobic and hydrophilic amino acid residues to alanine in alpha helices. This fact allows us to state that alpha helices containing a lot of sheet-like pentapeptides featuring alanine exist in helical conformation, at least, in "normal" conditions. However, in our opinion, the risk of helix to sheet transition promoted by certain triggers (increase of concentration, posttranslational modifications, changes of pH, etc.) is higher for proteins encoded by GC-rich genes than for those encoded by AT-rich genes.

As to pentapeptides which demonstrate significant preference to be included in beta sheet, their total level of usage in sheet is high in proteins encoded by both AT-rich and GC-rich genes. Moreover, the level of usage in sheet-like pentapeptides in beta strands for each of the 20 amino acid residues is significantly higher than the level of usage in helix-like or coil-like pentapeptides ($P < 0.003$). From this point of view, the growth of sheet-like pentapeptides in helices and coil regions under the influence of GC-pressure should not be ignored as a factor increasing probability of helix to sheet and coil to sheet transitions.

There are two specific coil formers (glycine and proline) among amino acids encoded by GC-rich codons. Those amino acids are hydrophobic (according to Eisenberg scale [14]) but acrophilic (according to the scale of Hopp and Woods [24]). Acrophilicity is the probability to be found on the surface of protein [24]. These amino acids are frequently found in coil in sheet-like pentapeptides in both proteins encoded by AT-rich and GC-rich genes. However, their usage is much higher in proteins encoded by GC-rich genes. According to our results, proline and glycine are able to form coil being situated in "wrong" positions. Moreover, proline and glycine were shown in this study to promote incorporation of hydrophobic amino acids in coil. This process is responsible for the increase of antigenic properties of proteins encoded by GC-rich genes [25,26]. Due to appearance of proline, glycine and hydrophobic amino acids on surface of proteins (in coil regions) percent of linear B-cell epitopes and antigenicity of conformational epitopes is increasing in lineages of homologous proteins with the growth of GC-content in genes coding for them [12]. So, natural selection allows hydrophobization of coil under the influence of GC-pressure. Appearance of hydrophobic coil regions on surfaces of proteins may theoretically make them prone not only to be better bound by antibodies [12,25,26] but also, in certain conditions, to form aggregates connected by intermolecular beta sheet.

Processes of helix-sheet and coil-sheet transitions attract a lot of attention because due to these processes amyloids are formed [27]. Pathogenesis of many diseases includes formation and deposition of amyloid [27]. Different molecular mechanisms of amyloid formation should exist. In this work we came closer to the understanding of processes making one of those mechanisms more probable. Formation of intermolecular beta-sheet from superficial hydrophobic regions of coil and alpha helices may be actual for those amyloidogenic proteins and peptides which are encoded by GC-rich genes. Moreover, it was shown that large proteins encoded by GC-rich genes usually demonstrate slower *in vitro* refolding rates than those which are encoded by AT-rich genes [28]. Slow folding of proteins encoded by GC-rich genes which may be explained by the increase of total hydrophobic amino acids usage [28] and the decrease of the usage of pentapeptides with specific combinations of hydrophobic and hydrophilic amino acids (those combinations should "guide" the process of folding) may increase probability of coil to sheet and helix to sheet transitions.

Regions of coil from proteins encoded by AT-poor genes should be protected from coil to sheet transitions by their increased hydrophilicity. This increased hydrophilicity itself should occur due to the decrease of Gly and Pro usages under the influence of AT-pressure and increase of Asn and Lys usages. Hydrophilic Asn and Lys may form helix or beta sheet being situated in specific positions relatively to other hydrophilic and hydrophobic amino acid residues. Negative selection stabilizing secondary structure of proteins should prevent their appearance in those "wrong" positions and allow their appearance in "right" ones.

The usage of coil-like pentapeptides grows under the influence of mutational AT-pressure both in coil and in helices. Theoretically, it should increase the risk of helix to coil transitions in proteins encoded by AT-rich genes. However, the usage of helix-like pentapeptides in helices is always higher than the usage of coil-like ones. Moreover, the usage of coil-like pentapeptides containing each of the 20 amino acids is also significantly lower in helix than the usage of helix-like pentapeptides ($P < 0.0001$). It is likely that some of those hydrophilic coil-like pentapeptides in helices of proteins encoded by AT-rich genes are stabilized by ionic or polar interactions between hydrophilic amino acid residues included in them.

5. Conclusions

Finally we came to the conclusion that symmetric mutational pressure has yet another previously unknown consequence. Mutational AT-pressure stimulates negative selection which prevents changes in secondary structure of proteins. Together they produce characteristic pattern of hydrophobic clusters distribution in helices and hydrophilization of coil regions. Mutational GC-pressure leads to the growth of alanine, glycine and proline usages. Alanine does not require specific stabilization in helices, while glycine and proline do not require specific stabilization in coil. That is why stabilization of helix and coil by combinations of hydrophobic and hydrophilic amino acids described in the present work becomes less important under the influence of mutational GC-pressure. Helices and coil regions of proteins encoded by GC-rich genes should keep their conformation in normal conditions even being destabilized. However, high levels of sheet-like pentapeptides should make them prone to undergo helix to sheet and coil to sheet transitions and so to form amyloid in special conditions.

Appendix A. Supplementary material

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.biochi.2012.08.008.

References

[1] P.Y. Chou, G.D. Fasman, Prediction of the secondary structure of proteins from their amino acid sequence, Adv. Enzymol. Relat. Areas. Mol. Biol. 47 (1978) 45–48.
[2] W. Pirovano, J. Heringa, Protein secondary structure prediction, Methods Mol. Biol. 609 (2010) 327–348.
[3] S. Costantini, G. Colonna, A.M. Facchiano, PreSSAPro: a software for the prediction of secondary structure by amino acid properties, Comput. Biol. Chem. 31 (2007) 389–392.
[4] E.V. Barkovsky, Prediction of the secondary structure of globular proteins by amino acid sequence, Acta Biol. Med. Germ 41 (1982) 751–758.
[5] E.V. Barkovsky, D.V. Kirilenko, Maps of preferential conformation of dipeptides in the structural regions of globular proteins, Biofizika 30 (1985) 786–790.
[6] V.I. Lim, Stereochemical theory of the secondary structure of globular proteins. II. Methods of localization of alpha-helical and beta-structural regions, Biofizika 19 (1974) 562–575.
[7] S. Anishetty, G. Pennathur, R. Anishetty, Tripeptide analysis of protein structures, BMC Struct. Biol. 2 (2002) 9.

[8] C.N. Pace, J.M. Scholtz, A helix propensity scale based on experimental studies of peptides and proteins, Biophys. J. 75 (1998) 422–427.
[9] N. Sueoka, Directional mutation pressure and neutral molecular evolution, Proc. Natl. Acad. Sci. U. S. A. 85 (1988) 2653–2657.
[10] V.V. Khrustalev, E.V. Barkovsky, Main pathways of proteome simplification in alphaherpesviruses under the influence of the strong mutational GC-pressure, J. Proteomics Bioinform. 2 (2009) 88–96.
[11] V.V. Khrustalev, E.V. Barkovsky, Study of completed archaeal genomes and proteomes: hypothesis of strong mutational AT pressure existed in their common predecessor, Genomic. Proteomics Bioinform. 8 (2010) 22–32.
[12] V.V. Khrustalev, E.V. Barkovsky, Percent of highly immunogenic amino acid residues forming B-cell epitopes is higher in homologous proteins encoded by GC-rich genes, J. Theor. Biol. 282 (2011) 71–79.
[13] T. Banerjee, S.K. Gupta, T.C. Ghosh, Role of mutational bias and natural selection on genome-wide nucleotide bias in prokaryotic organisms, Biosystems 81 (2005) P11–P18.
[14] D. Eisenberg, E. Schwarz, M. Komaromy, R. Wall, Analysis of membrane and surface protein sequences with the hydrophobic moment plot, J. Mol. Biol. 179 (1984) 125–142.
[15] G.A. Wu, E.A. Coutsias, K.A. Dill, Iterative assembly of helical proteins by optimal hydrophobic packing, Structure 16 (2008) 1257–1266.
[16] R. Rajgaria, S.R. McAllister, C.A. Floudas, Towards accurate residue-residue hydrophobic contact prediction for alpha helical proteins via integer linear optimization, Proteins 74 (2009) 929–947.
[17] N. Bhattacharjee, P. Biswas, Position-specific propensities of amino acids in the beta-strand, BMC Struct. Biol. 10 (2010) 29.
[18] Y. Nakamura, T. Gojobori, T. Ikemura, Codon usage tabulated from the international DNA sequence databases: status for the year 2000, Nucl. Acids Res. 28 (2000) 92.
[19] N. Sueoka, Y. Kawanishi, DNA G+C content of the third codon position and codon usage biases of human genes, Gene 261 (2000) 53–62.
[20] Y. Bai, S.W. Englander, Hydrogen bond strength and β-sheet propensities: the role of a side chain blocking effect, Proteins 18 (1994) 262–266.
[21] E. Gazit, Self assembly of short aromatic peptides into amyloid fibrils and related nanostructures, Prion 1 (2007) 32–35.
[22] T.M. Doran, A.J. Kamens, N.K. Byrnes, B.L. Nilsson, Role of amino acid hydrophobicity, aromaticity, and molecular volume on IAPP(20-29) amyloid self-assembly, Proteins 80 (2012) 1053–1065.
[23] R. Aurora, G.D. Rose, Helix capping, Protein Sci. 7 (1998) 21–38.
[24] T.P. Hopp, Protein antigen conformation: folding patterns and predictive algorithms; selection of antigenic and immunogenic peptides, Ann. Sclavo 2 (1984) 47–60.
[25] V.V. Khrustalev, Can mutational GC-pressure create new linear B-cell epitopes in herpes simplex virus type 1 glycoprotein B? Immunol. Invest. 38 (2009) 613–623.
[26] V.V. Khrustalev, Mutational pressure makes HIV1 gp120 linear B-cell epitopes shorter and may lead to their disappearance, Mol. Immunol. 47 (2010) 1635–1639.
[27] J.D. Sipe, M.D. Benson, J.L. Buxbaum, S.I. Ikeda, G. Merlini, M.J.M. Saraiva, P. Westermark, Amyloid fibril protein nomenclature: 2010 recommendations from the nomenclature committee of the international society of amyloidosis, Amyloid 17 (2010) 101–104.
[28] J.T. Huang, D.J. Xing, W. Huang, Choice of synonymous codons associated with protein folding, Proteins 80 (2012) 2056–2062.