

Main Pathways of Proteome Simplification in Alphaherpesviruses Under the Influence of the Strong Mutational GC-pressure

Vladislav V. Khrustalev* and Eugene V. Barkovsky

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

*Corresponding author: Vladislav V. Khrustalev, Belarus, Minsk, 220029, Communisticheskaya 7-24, Tel: 80292845957; E-mail: vvkhrustalev@mail.ru

Received January 14, 2009; Accepted February 09, 2009; Published February 20, 2009

Citation: Khrustalev VV, Barkovsky EV (2009) Main Pathways of Proteome Simplification in Alphaherpesviruses Under the Influence of the Strong Mutational GC-pressure. *J Proteomics Bioinform* 2: 088-096. doi:10.4172/jpb.1000065

Copyright: © 2009 Khrustalev VV, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Abstract

The simplification of amino acid content distribution under the influence of the strong mutational pressure takes place in simplex and varicello viruses proteins coded by genes with GC-content higher than 60%. We proved this statement by the way of in-silico calculation of Shannon's entropy of amino acid content distribution in all proteins from ten completely sequenced simplex and varicello viruses. Entropy of amino acid content distribution decreases because of the growth of GARP (glycine, alanine, arginine and proline) usage due to the decrease not only in FYMINK (phenylalanine, tyrosine, methionine, isoleucine, asparagine and lysine) but also in other amino acids (coded by codons with average GC-content) usages. Threonine, serine, glutamine and cysteine are frequently substituted to GARP in proteins coded by genes with G+C higher than 60% (threonine and serine are substituted mostly to alanine while glutamine and histidine are substituted mostly to arginine). Cysteine, valine and leucine are frequently substituted to GARP only in proteins coded by genes with G+C higher than 80%, probably because of the higher radicalism of these substitutions. Levels of tryptophan, glutamic and aspartic acids do not decrease under the influence of GC-pressure even in proteins coded by genes with G+C higher than 80%.

Key words: GC-content; Amino acid substitutions; Mutational pressure; Alphaherpesviruses; Entropy; Capsid proteins; Glycoproteins; HSV1; PaHV2

Abbreviations

GARP: total level of glycine, alanine, arginine and proline usage (amino acids coded by GC-rich codons).

FYMINK: total level of phenylalanine, tyrosine, methionine, isoleucine, asparagine and lysine usage (amino acids coded by GC-poor codons).

10AA: total level of glutamine, serine, threonine, histidine, leucine, valine, cysteine, tryptophan, aspartic and

glutamic acids usage (amino acids coded by codons average in GC-content).

G+C (GC-content): total level of guanine and cytosine in gene.

GC-pressure: mutational pressure causing the following imbalance in nucleotide substitutions rates: AT to GC substitutions occur more frequently than GC to AT substitutions.

Introduction

In this work we showed that mutational GC-pressure leads to the simplification of amino acid content of proteins coded by genes enriched with guanine and cytosine. There are only four amino acids (GARP: glycine, alanine, arginine and proline) coded by codons containing guanine or cytosine in both first and second codon positions. There is usually a strong linear correlation between GC-content of genes and the level of GARP usage in proteins coded by those genes (Singer and Hickey, 2000). There are six amino acids (FYMINK: phenylalanine, tyrosine, methionine, isoleucine, asparagine and lysine) coded by codons with no cytosine or guanine in first and second codon positions. The level of FYMINK usage in proteins usually demonstrates negative correlation with GC-content of subsequent genes (Singer and Hickey, 2000). The mutational pressure theory predicts that the level of GARP usage should be high and the level of FYMINK usage should be low in proteins coded by GC-rich genes (Sueoka, 1988; Sueoka, 2002). Can this situation be interpreted as the simplification of amino acid content of proteins coded by GC-rich genes?

To find out the answer we applied Claude Shannon's information theory (Zeeberg, 2002) to ten completely sequenced genomes of alphaherpesviruses. Claude Shannon's information theory is the best one that can be used to characterize the level of diversity of any biological system. The quantity of information (entropy) is the integral index of this theory. Entropy can be interpreted not only as the level of uncertainty, but also as a level of diversity of the system. The entropy increase is interpreted as a diversification of the system; decrease in entropy should be caused by the process leading to the increase of uniformity. Under the term "simplification" in this work we mean the situation when protein (or proteome) mostly consists of four types of amino acid residues and the levels of other residues are reduced. This interpretation of proteome simplification is equal to the loss of diversity and rise of uniformity of its amino acid content.

Previously we have shown that five genomes of simplexviruses (Human herpesvirus 1 and 2, Cercopithecine herpesvirus 2, Papiine herpesvirus 2 and Macacine herpesvirus 1) and the genome of Bovine herpesvirus 5 which belongs to varicellovirus genus are under the influence of strong mutational GC-pressure (Khrustalev and Barkovsky, 2008a). In this study we calculated the entropy of amino acid residues distribution in every protein of ten simplex and

varicello viruses. We showed that the entropy of amino acid residues distribution in proteins (the level of amino acid content uncertainty and diversity) demonstrates significant negative correlation with GC-content of subsequent genes only in genes with $G+C > 0.6$. The level of GARP usage in proteins demonstrates significant positive and the level of FYMINK demonstrates significant negative correlation with $G+C$ of either GC-rich or GC-poor genes.

The cause of the decrease of entropy in proteins coded by genes with $G+C > 0.6$ has been determined by us. We discovered previously unknown phenomenon: the total level of usage of ten amino acids coded by codons containing guanine or cytosine only in one of their two first codon positions (10AA: glutamine, serine, threonine, histidine, leucine, valine, cysteine, tryptophan, aspartic and glutamic acids) demonstrates negative correlation with GC-content, but only in genes with $G+C$ higher than 0.6.

Next step of our in-silico experiments gave us more concrete answer. Total level of 10AA decreases in proteins coded by GC-rich genes mostly due to decrease in threonine, serine, glutamine and histidine residues usage.

To find out the mechanism of GC-pressure associated decrease in Thr, Ser, Gln and His usage, we analyzed directions of these amino acid substitutions in two groups of proteins from human herpesvirus 1 (HSV1) and papiine herpesvirus 2 (PaHV2) that involved in immune answer. We made this kind of analysis in six capsid proteins and in twelve glycoproteins. Indeed, during the remission of herpes virus infection, blood plasma contains high titers of antibodies against HSV glycoproteins; at primary infection and during each relapse, the antibodies against capsid proteins are detected (Kuhn et al., 1987).

It turned out that the main mechanism of GC-pressure associated decrease in Thr and Ser usage is their substitutions to alanine. Interestingly, both quartet and duplet of serine (Ser4 and Ser2) were substituted mostly to Ala. In our opinion this is the evidence of the relative neutrality of serine to alanine and threonine to alanine substitutions: they have been fixed more frequently than serine and threonine substitutions to proline. Glutamine and histidine have been substituted mostly to arginine. On the other hand, levels of proline and glycine in PaHV2 capsid proteins and glycoproteins are also higher than those levels in HSV1. But amino acid substitutions leading to proline and glycine occurrence are not so frequent, probably, because of their radicalism.

In this paper we proved that GC-pressure simplify amino acid content distribution in proteins and showed the main cause and pathways of this simplification. We determined that Ser to Ala, Thr to Ala, Gln to Arg and His to Arg amino acid substitutions are relatively neutral, they are fixed extensively in genes with $G+C > 0.6$. This neutrality should be associated with biochemical features of these amino acid residues.

In proteins coded by genes with extremely high GC-content ($G+C > 0.8$) the level of 10AA decreases also because of Val, Leu and Cys substitutions to GARP. Interestingly, levels of Glu, Asp and Trp stay the same in proteins coded by genes with $G+C > 0.8$ as in proteins coded by genes with lower GC-content. This may be interpreted as the evidence of the extreme radicalism of Glu, Asp and Trp substitutions to amino acids from GARP group.

Materials and Methods

In this work we analyzed ten completely sequenced genomes of simplex and varicello viruses. The first part of our in-silico experiments has been performed on “lists of codon usage for each CDS (coding district)”. This kind of data can be found in Codon Usage Database (Nakamura et al., 2000) (www.kazusa.or.jp/codon). To calculate total level of guanine and cytosine in each coding district ($G+C$), levels of amino acid usage and the entropy of amino acid content distribution (H) we used our original MS Excel tool called “CGS” (Coding Genome Scanner) that can be downloaded for free from our web site www.barkovsky.hotmail.ru. All these indexes are calculated automatically after copying the data from the “list of codon usage for each CDS” into the special list of “CGS” called “All CDSs”. In this work we focused on dependences between $G+C$ and amino acid usage in all genes and subsequent proteins from ten completely sequenced genomes of alphaherpesviruses.

Completely sequenced genomes of simplex viruses used in this work are listed below. Macacine herpesvirus 1 (*MaHV1*) [NC_004812], Cercopithecine herpesvirus 2 (*CeHV2*) [NC_006560], Papiine herpesvirus 2 (*PaHV2*) [NC_007653], Human herpesvirus 1 (*HSV1*) [NC_001806], Human herpesvirus 2 (*HSV2*) [NC_001798]. Completely sequenced genomes of varicello viruses: Human herpesvirus 3 (*VZV*) [NC_001348], Bovine herpesvirus 5 (*BoHV5*) [NC_005261], Equid herpesvirus 1 (*EqHV1*) [NC_001491], Equid herpesvirus 4 (*EqHV4*) [NC_001844], Cercopithecine

herpesvirus 9 (*CeHV9*) [NC_002686].

Entropy of amino acid content distribution (H) was calculated according to Claude Shannon’s information theory (Zeeberg, 2002).

$$H = -\sum faa \cdot \log_2 faa \quad (1)$$

In this equation faa is the frequency of amino acid residue usage. So, according to equation 1, entropy is the negative sum of products of frequencies of each amino acid residue usage and logarithms of these frequencies. The maximum level of uncertainty (maximal entropy) for amino acid content of protein is 4,322 bit. The lower is the level of entropy, the lower is the uncertainty of amino acid residues distribution.

We calculated common level of usage for four amino acid residues (GARP) coded by codons with guanine or cytosine in both first and second codon positions (glycine, alanine, arginine and proline); as well as common level of usage for six amino acid residues (FYMINK) coded by codons with no guanine and cytosine in their first and second codon positions (phenylalanine, tyrosine, methionine, isoleucine, asparagine and lysine) in each protein.

The common level of usage for ten amino acid residues (10AA) coded by codons with “average” GC-content (glutamine, serine, threonine, histidine, leucine, valine, cysteine, tryptophan, aspartic and glutamic acids) has also been calculated in each protein. Lately we analyzed levels of these amino acids usage separately. We compared levels of each of these amino acids in proteins coded by genes with different GC-content. There are six groups of genes arranged according to their GC-content ($0.3 < G+C < 0.4$; $0.4 < G+C < 0.5$; $0.5 < G+C < 0.6$; $0.6 < G+C < 0.7$; $0.7 < G+C < 0.8$; $0.8 < G+C < 0.85$); $n > 30$ in each of these groups.

For the second part of our experiments we used nucleotide sequences coding for six capsid proteins (capsid portal protein, capsid triplex subunit 1, capsid triplex subunit 2, capsid scaffold protein, small capsid protein and major capsid protein) and twelve glycoproteins (envelope glycoproteins: L, M, H, B, C, N, K, G, J, D, I, E). The total length of capsid proteins (including gaps) is 3297 amino acid residues; the total length of glycoproteins is 5563 amino acid residues. Glycoproteins and capsid proteins are the main targets for protective antibodies (Kuhn, 1987), so their amino acid content (and composition) is the subject of interest for immunology.

We calculated preferable directions of amino acid substitutions from capsid proteins and glycoproteins of HSV1 to the homologous proteins of PaHV2. GC-content of genes coding for analyzed proteins of HSV1 is between 0.6 and 0.7, while GC-content of PaHV2 homologues is between 0.7 and 0.8. So, working with this material we can find out the directions of substitutions causing the significant decrease in levels of four amino acids (Gln, Thr, Ser and His) observed between proteins coding by genes with $0.6 < G+C < 0.7$ and proteins coded by genes with $0.7 < G+C < 0.8$. Levels of quartet and duplet of serine (Ser4 and Ser2) have been calculated separately.

For finding out the direction of amino acid substitutions we used our new MS Excel tool "CodonChanges" made on the basis of previously existed algorithm "VVK 3.4." (Khrustalev and Barkovsky, 2008b). This algorithm is working with previously aligned sequences in which "N" is used as a symbol for gap. In the "Codon" list one should write the codon. In the "Changes" list the numbers of codons

situated in the same sites of sequence 2 will appear. This algorithm is also available via www.barkovsky.hotmail.ru.

All the alignments in this work have been performed with the help of MEGA 4 program (Tamura et al., 2007), PAM matrix has been used.

Results

In Figure 1 you can see the dependence between the entropy of amino acid content distribution in all proteins of ten simplex and varicello viruses and the level of GC-content (G+C) in genes coding for them. The coefficient of correlation between G+C and entropy is -0.73, however, if we look at Figure 1, we can see that this dependence is not just linear. The greatest decrease in entropy occurs in proteins coded by genes with higher G+C. The coefficient of correlation (R) between entropy and G+C calculated for genes with $G+C < 0.5$ is -0.18. For genes with $G+C < 0.6$ the level of R (-0.36) shows that the negative linear dependence is

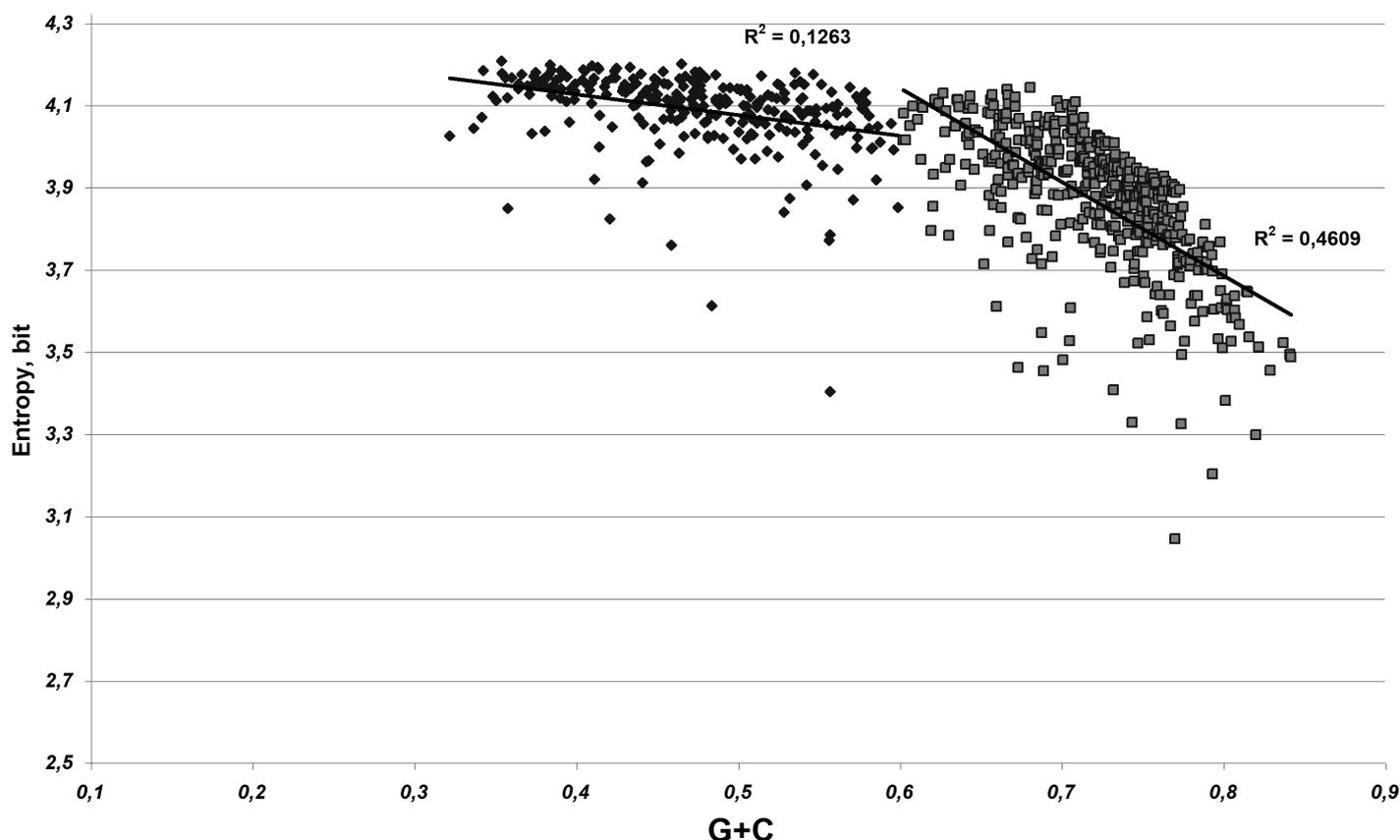


Figure 1: Dependence between entropy of amino acid content distribution in all proteins from ten simplex and varicello viruses and GC-content (G+C) of genes coding for them.

still weak. This dependence becomes stronger for genes with $G+C < 0.7$ ($R = -0.54$).

The conclusion from the analysis of graph in Figure 1 is the following. The entropy of amino acid content distribution in proteins significantly decreases with the growth of $G+C$ in subsequent genes only in genes with $G+C > 0.6$. For genes with $0.3 < G+C < 0.6$ the dependence between entropy and $G+C$ is weak.

To find out the cause of the decrease in entropy in proteins coded by genes with $G+C > 0.6$ we built the graph that is shown in Figure 2. As you can see in Figure 2, the level of GARP demonstrates the linear dependence on $G+C$ and the level of FYMINK demonstrates the negative linear dependence on $G+C$ in all genes. The level of 10AA (amino acids coded by codons with guanine or cytosine in first or in second codon positions, but never in both first and second codon positions) shows two phase dependence on $G+C$, just like the previously described dependence between entropy and $G+C$. Indeed, the level of 10AA decreases only in proteins coded by genes with $G+C > 0.6$. This fact makes us

hypothesize that the decrease in entropy in proteins coded by genes with $G+C > 0.6$ is due to the decrease in 10AA in these proteins.

In the Figure 3 one can see that the total level of 10AA decreases under the influence of GC-pressure mostly due to decrease in Gln, Thr, Ser and His usage. Levels of Cys, Leu and Val decrease only in proteins coded by genes with $G+C > 0.8$. Interestingly, levels of Glu, Trp and Asp do not decrease significantly in alphaherpesviruses proteins with the growth of GC-content in genes coding for them.

To explain these differences we created original hypothesis. Nucleotide substitutions caused by GC-pressure occur in all codons with the same rates, yet only some of them are fixed by the random genetic drift (Sueoka, 1988; Sueoka, 2002). Most of the amino acid substitutions are eliminated from population by the negative selection or random genetic drift. Negative selection should eliminate those amino acid substitutions that are “negative” for the function of proteins. It means that “radical” amino acid substitutions should be fixed much rarely than “neutral” ones.

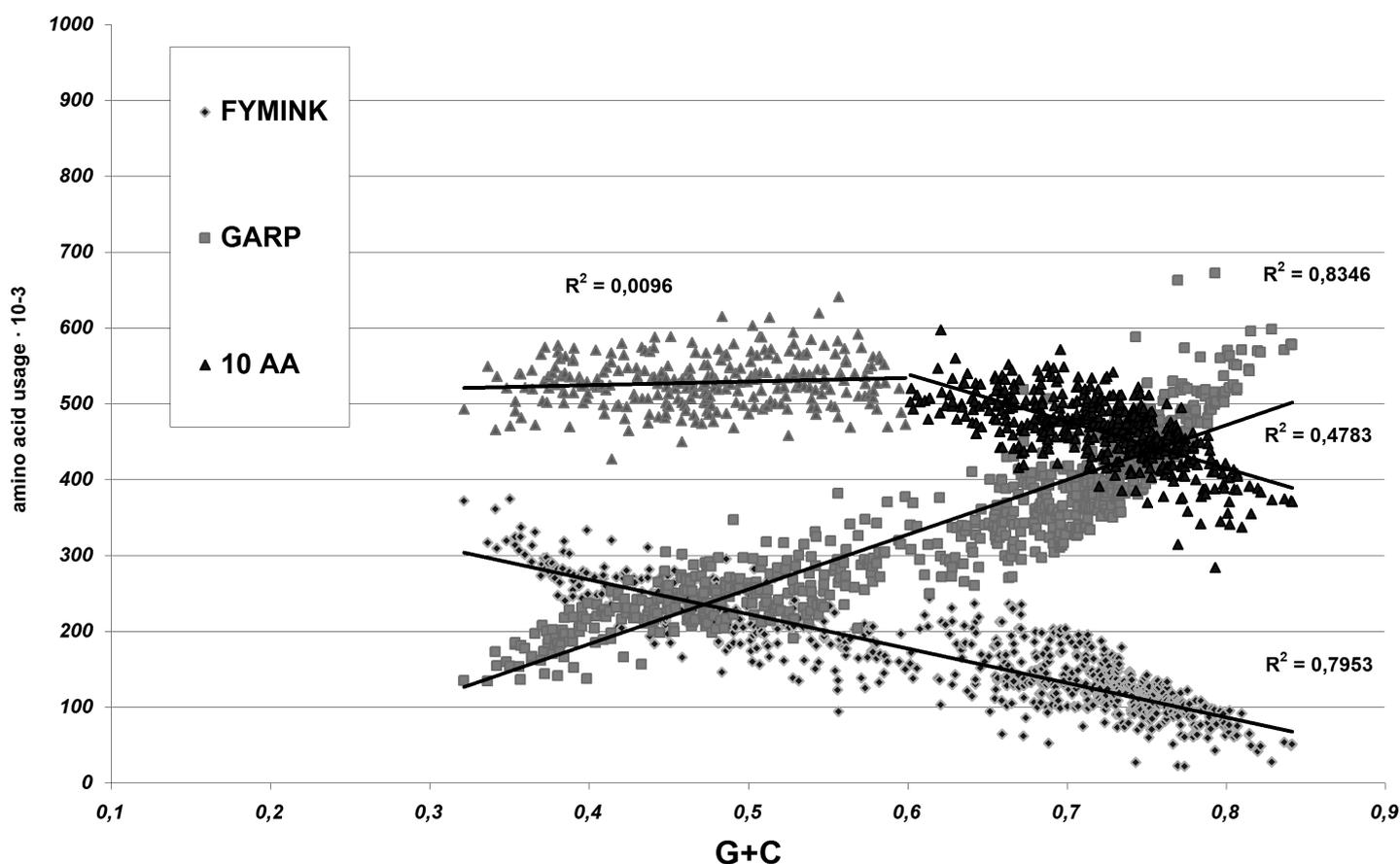


Figure 2: Dependence of amino acid content (GARP, FYMINK and 10AA) in all proteins from ten simplex and varicello viruses on GC-content ($G+C$) of genes coding for them.

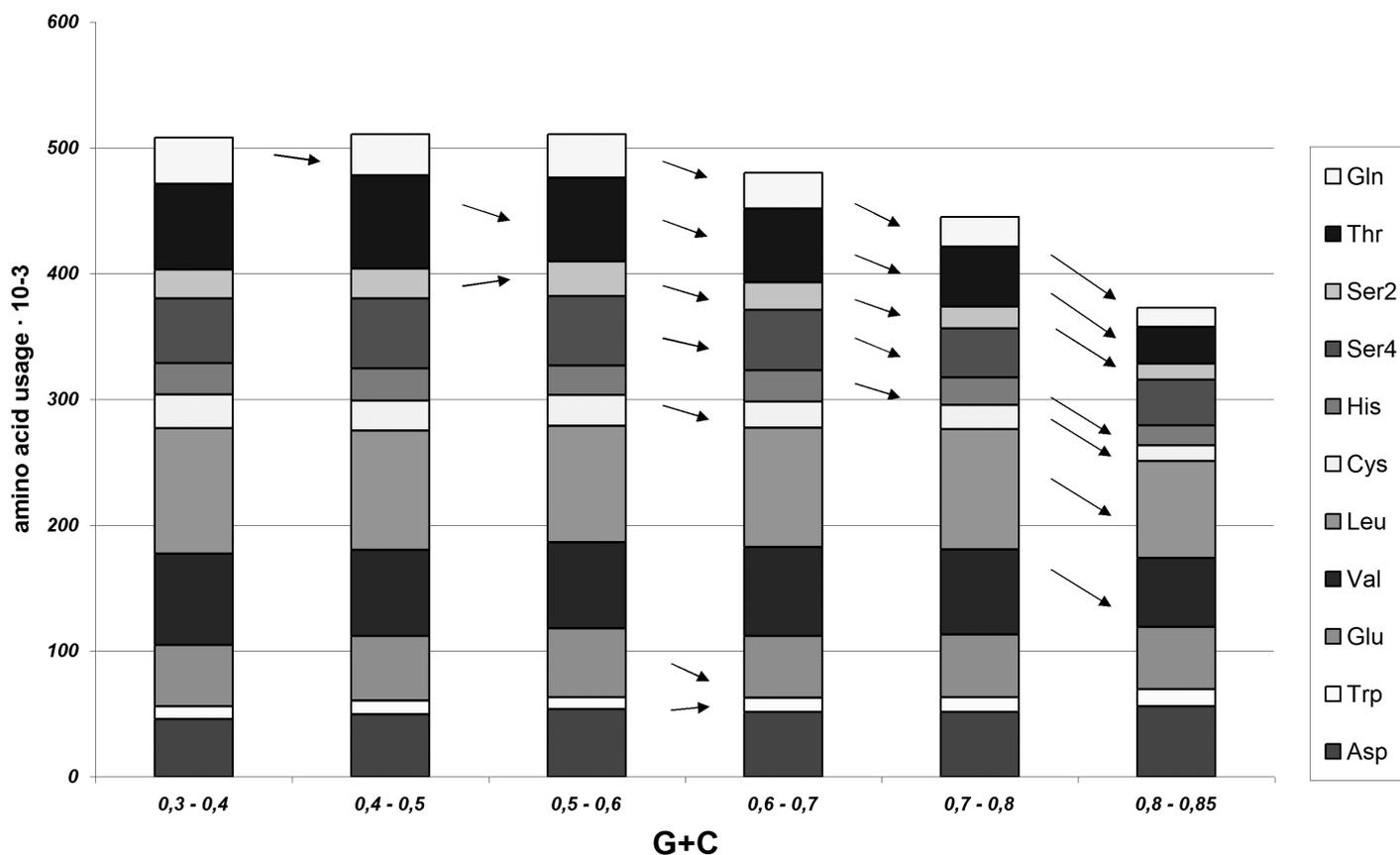


Figure 3: Levels of ten amino acid residues usage in all proteins from ten simplex and varicello viruses. Proteins are grouped according to the GC-content of genes coding for them. Significant differences in amino acid usage between two groups of proteins are marked by arrows.

With the help of Figure 3 we can estimate the relative level of neutrality/radicalism for substitutions of 10 amino acids to GARP. So, substitutions of Gln, Thr, Ser and His under the influence of GC-pressure are relatively neutral, while substitutions of Cys, Leu and Val are relatively radical. The greatest degree of radicalism has been observed for substitutions of Glu, Trp and Asp.

The final step of our in-silico experiments brought us even more concrete knowledge. We estimated the direction of substitutions in Gln, Thr, Ser and His codons under the influence of GC-pressure. To make it we aligned six capsid proteins and twelve glycoproteins of HSV1 with their homologues from PaHV2. With the help of "Codon Changes" algorithm we found out the main pathways of alphaherpesviruses proteomes simplification.

In Table 1 we placed the percentage of amino acid substitutions directions between HSV1 capsid proteins and PaHV2 capsid proteins. Codons from serine quartet (Ser4) are substituted mostly to alanine (by the way of T to G transversion in first codon positions), as well as codons

coded for threonine (by the way of A to G transitions). Codons from serine duplet are also most frequently substituted to alanine, even though this is at least two-step nucleotide substitution. These data may be the evidence of relative neutrality of Ser to Ala and Thr to Ala amino acid substitutions. Substitutions of Ser and Thr to Pro are not so neutral, as previous ones, probably, due to characteristic biochemical features of proline.

Glutamine and histidine are most frequently substituted to arginine (by the way of A to G transitions). So, we can conclude that Gln to Arg and His to Arg substitutions are more neutral than Gln to Pro and His to Pro ones.

The data in Table 2 have much in common with the data presented in Table 1. The greatest difference is in the percent of nonmutated and synonymously mutated codons. Capsid proteins are seemed to be more conserved than glycoproteins. An interesting feature of Ser2 codons is that they are rarely substituted to Arg (by the way of A to C transversions), maybe because of Ser to Arg substitution radicalism.

Amino acid, codons	Number of amino acid residues in HSV1	Nonmutated and synonymously mutated, %	Ala, %	Pro, %	Arg, %	Gly, %	other, %
Ser4 (TCX)	97	47.4	22.7 (T to G)	10.3 (T to C)	3.1	0.0	15.5 (6 AA + gaps)
Thr (ACX)	185	60.5	17.3 (A to G)	4.3 (A to C)	1.6	2.2	14.1 (10 AA + gaps)
Gln (CAA/G)	128	63.3	3.9	2.3 (A to C)	12.5 (A to G)	1.6	16.4 (8 AA + gaps)
His (CAT/C)	98	73.5	2.0	5.1 (A to C)	6.1 (A to G)	2.0	11.3 (6 AA + gaps)
Ser2 (AGT/C)	51	54.9	13.7	0.0	2.0 (A to C)	9.8 (A to G)	19.6 (4 AA + gaps)

Table 1: Amino acid substitutions in six capsid proteins, counted from HSV1 to PaHV2.

Amino acid, codons	Number of amino acid residues in HSV1	Nonmutated and synonymously mutated, %	Ala, %	Pro, %	Arg, %	Gly, %	other, %
Ser4 (TCX)	128	37.5	17.2 (T to G)	10.9 (T to C)	4.7	5.5	24.2 (10 AA + gaps)
Thr (ACX)	223	41.3	10.8 (A to G)	5.4 (A to C)	4.9	6.7	30.9 (8 AA + gaps)
Gln (CAA/G)	86	36.0	7.0	9.3 (A to C)	20.9 (A to G)	3.5	23.3 (6 AA + gaps)
His (CAT/C)	69	47.8	5.8	7.2 (A to C)	17.4 (A to G)	2.9	18.9 (6 AA)
Ser2 (AGT/C)	57	43.9	10.5	10.5	5.3 (A to C)	8.8 (A to G)	21.0 (8 AA + gaps)

Table 2: Amino acid substitutions in twelve glycoproteins, counted from HSV1 to PaHV2.

The summary of our results is the following. The entropy of amino acid content distribution decreases in alphaherpesviruses proteins coded by genes with $G+C > 0.6$ not only due to decrease in FYMINK usage, but also due to frequently fixed substitutions of Ser and Thr to Ala and Gln and His to Arg.

Discussion

In this work we showed how concretely strong GC-pressure influences amino acid content of simplex and varicello viruses' proteins. There are some amino acid substitutions that can be relatively easily fix because of their neutrality. Minimal limitations from the negative selection allow GC-content in first and second codon positions to grow mostly by the way of fixation of certain amino acid substitutions (Khrustalev and Barkovsky, 2008b). The fact of FYMINK usage decrease under the influence of GC-pressure is well known (Singer and Hickey, 2000). The fact of GARP usage increase is also described well for proteins coded by GC-rich genes (Singer and Hickey, 2000). Codons coded for FYMINK cannot mutate strictly to codons coded for GARP. At first, amino acids from FYMINK group have to be substituted to some of ten amino acids (10AA) coded by codons with "average" GC-content in their first and second codon positions. Only the second step of nucleotide substitutions can bring amino acid from GARP group in the site previously contained amino acid from FYMINK group. This model has been tested in our work. Now we can conclude that the level of GARP usage in proteins coded by genes with $G+C > 0.6$ is increased not only due to fixation of $FYMINK \rightarrow 10AA \rightarrow GARP$ substitutions but also due to fixation of $10AA \rightarrow GARP$ ones.

The method of radicalism/neutrality estimation for amino acid substitutions in proteins under the influence of GC-pressure is novel and reliable. It deals with substitutions widespread in nature. There is no doubt that Asp to Gly and Asp to Ala mutations occur frequently under the influence of mutational pressure (by the way of A to G transitions and A to C transversions, subsequently), but the level of Asp usage does not decrease significantly even in proteins coded by genes with $G+C > 0.8$. It means that Asp to Gly and Asp to Ala mutations occur but Asp to Gly and Asp to Ala substitutions are fixed rarely. We can give only one explanation of this fact: Asp to Gly and Asp to Ala mutations are eliminated by the negative selection because of their radicalism. The nature of this kind of radicalism should be biochemical. Indeed, aspartic acid has hydrophilic and negatively charged

side chain, unlike glycine that has no side chain at all and alanine that has hydrophobic nonpolar side chain (methyl group). The same kind of situation observed for the second acidic amino acid (glutamic acid) too.

Under the influence of GC-pressure the uncertainty of amino acid content distribution in subsequent proteins decreases. Now we can conclude that entropy of amino acid content distribution in proteins coded by genes with $0.6 < G+C < 0.8$ decreases because of the increase of Ala, Arg, Pro and Gly levels of usage due to decrease of Phe, Tyr, Met, Ile, Asn, Lys, Gln, Thr, Ser and His levels. In viral proteins coded by genes with extremely high GC-content ($G+C > 0.8$) level of GARP begins to grow also due to decrease of Leu, Val and Cys levels.

The simplification of amino acid content of capsid proteins and glycoproteins under the influence of GC-pressure should lead to significant changes in their physical, chemical and immunological features. Glycine and proline, according to Hopp's works (Hopp and Woods, 1983; Hopp, 1984), are the most acrophilic amino acid residues. They are situated mostly on the surface of protein globules (in water solutions) (Hopp, 1984). This is the reason why Gly and Pro are frequently included in linear and discontinuous epitopes (Hopp and Woods, 1983; Hopp, 1984). So, glycoproteins and capsid proteins enriched with Gly and Pro should contain more linear B-cells epitopes than glycoproteins and capsid proteins with average level of these two amino acid residues. We can also predict that GC-pressure should lead to formation of new and enlargement of previously existed linear epitopes.

Alanine is fairly neutral amino acid residue that can be located in both hydrophilic regions on the protein surface and in the hydrophobic areas inside. Because of the absence of side chain, glycine is the most flexible amino acid residue. Arginine has a long flexible side-chain with a positively-charged end. Proline can disrupt protein folding structures like α helix or β sheet. So, alanine is seemed to be more neutral from the biochemical point of view than any other amino acid from GARP group. That is why both threonine and serine are substituted to alanine more frequently than to proline.

In this in-silico work we proved the existence and showed the main pathways of proteomes simplification caused by mutational GC-pressure. Our findings are reliable for genomes and proteomes of simplex and varicello viruses, but

we believe that the process of proteomes simplification under the influence of mutational GC-pressure is universal.

Acknowledgment

We thank Professor Khotileva L.V., academician of National Belarussian Academy of Science, for the great support and productive conversation she always provides on our works.

References

1. Hopp TP (1984) Protein antigen conformation: folding patterns and predictive algorithms; selection of antigenic and immunogenic peptides. *Ann Sclavo collana monogr* 2: 47-60. » [PubMed](#) » [Google Scholar](#)
2. Hopp TP, Woods KR (1983) A computer program for predicting protein antigenic determinants. *Mol Immunol* 20: 483-489. » [PubMed](#) » [Google Scholar](#)
3. Khrustalev VV, Barkovsky EV (2008a) Mutational pressure in genomes of human α -herpesviruses. *Mol Gen Microb, Virol* 23: 94-100. » [CrossRef](#) » [Google Scholar](#)
4. Khrustalev VV, Barkovsky EV (2008b) An in-silico study of alphaherpesviruses ICP0 genes: Positive selection or strong mutational GC-pressure. *IUBMB Life* 60: 456-460. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
5. Kuhn JE, Dunkler G, Munk K, Braun RW (1987) Analysis of the IgM and IgG antibody response against herpes simplex virus type 1 (HSV-1) structural and nonstructural proteins. *J Med Virol* 23: 135-150. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
6. Nakamura Y, Gojobori T, Ikemura T (2000) Codon usage tabulated from the international DNA sequence databases: status for the year 2000. *Nucleic Acids Res* 28: 292. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
7. Singer GAC, Hickey DA (2000) Nucleotide bias causes a genomewide bias in the amino acid composition of proteins. *Mol Biol Evol* 17: 1581-1588. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
8. Sueoka N (1988) Directional mutation pressure and neutral molecular evolution. *Proc Natl Acad Sci USA* 85: 2653-2657. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
9. Sueoka N (2002) Wide intra-genomic G+C heterogeneity in human and chicken is mainly due to strand-symmetric directional mutation pressures: dGTP-oxidation and symmetric cytosine-deamination hypotheses. *Gene* 300: 141-154. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
10. Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol* 24: 1596-1599. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)
11. Zeeberg B (2002) Shannon information theoretic computation of synonymous codon usage biases in coding regions of human and mouse genomes. *Genome Res* 12: 944-955. » [CrossRef](#) » [PubMed](#) » [Google Scholar](#)