Essential amino acid usage and evolutionary nutrigenomics of eukaryotes - insights into the differential usage of amino acids in protein domains and extra-domains

L. Santana-Santos, F. Prosdocimi and J.M. Ortega

Laboratório de Biodados, Departamento de Bioquímica e Imunologia,
Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais,
Belo Horizonte, MG, Brasil

Corresponding author: J.M. Ortega
E-mail: miguel@icb.ufmg.br

ABSTRACT. Nutrigenomics studies the effects of nutrients on the genome, transcriptome and proteome of organisms, and here an evolutionary standpoint on this new discipline is presented. It is well known that metazoan organisms are unable to synthesize all amino acids necessary to produce their proteins and that these essential amino acids (EAA) must be acquired from the diet. Here, we tested the hypothesis that conserved regions such as protein domains (DM) have different essentiality indexes and use different sets of amino acids when compared to extra-domains (ED) and proteins without mapped domains (WD). We found that auxotrophic organisms have a tendency to use less EAAs in DM than do prototrophic ones. Looking into the amino acid usage of eukaryotic proteins downloaded from KEGG and COG, we showed that WD have a usage of amino acids closer to DM, which suggests that proteins without mapped domains behave as large domains. Using an ED index that shows the proportion of prevalent amino acids in ED, a differential usage of amino acids in domains versus extra-domains was demonstrated. Protein domains were shown to be enriched with a higher number of EAA, and it may be related to the fact that these amino acids had lost their biosynthetic pathways in metazo-

ans during a great amino acid pathway deletion, followed by a nutritional constraint that may have happened close to the conquest of the terrestrial environment. Thus, the proportion of EAA outside domains could have decreased during evolution due to nutritional constraints.

Key words: Amino acid usage; Essential amino acid; Protein domains

INTRODUCTION

Nutrigenomics studies the effects of nutrients on the genome, transcriptome and proteome of organisms (Ferguson et al., 2007; Kaput et al., 2007; Mariman, 2007). The first PubMed entry of a paper using the word nutrigenomics in its title was observed in 2001 (Peregrin, 2001), and in the last six years, 71 other papers have already been published about this topic, showing its particular relevance to the present scenario of genomic studies. Although it has always been clear that a large number of dietary chemicals alter the regulation of biological processes and, either directly or indirectly, the expression of genetic information, researchers are just now beginning to take this influence into account and trying to measure it precisely (Kaput et al., 2007; Fenech, 2008). Most of these dozens of papers in the nutrigenomics field are still theoretical, proposing possible experimental strategies and future perspectives to study interactions between genomics and nutrition (Peregrin, 2001; van Ommen and Stierum, 2002; Chavez and Munoz, 2003; Muller and Kersten, 2003; Trayhurn, 2003; Chadwick, 2004). Researchers agree that nutrigenomics shall be considered the study of both the genetic factors influencing organisms' response to diet (Low and Tai, 2007) and the effects of food on the host genome and gene expression (Ozdemir and Godard, 2007; Roy et al., 2007; Ye et al., 2007). Like pharmacogenomics when thinking about drugs, nutrigenomics shall be used in the near future to predict the very best relationship between the genome and particular nutrients, trying to measure the amounts and different kinds of nutrients that would be preferred for a given individual in order that his/her metabolism work as well as possible (Fenech, 2005; Afman and Muller, 2006; Ghosh et al., 2007). Thus, nutrigenomics raises similar ethical, legal and social issues as does pharmacogenomics, particularly in regard to how the public will access nutrigenetic tests and associated nutritional advice (Castle and Ries, 2007; Ozdemir and Godard, 2007). Current approaches in nutrigenomics have addressed a number of interesting topics in biology, such as: cancer research (Davis and Milner, 2004; Junien and Gallou, 2004; Davis, 2007; Martin, 2007), longevity (Ye et al., 2007), obesity (Palou et al., 2004), inflammation (Roy et al., 2007), cardiovascular disease (Davis, 2007; Low and Tai, 2007), eye health (Delcourt, 2007), fertility (Dawson, 2006), and the study of taste and food preferences (El-Sohemy et al., 2007). It is noteworthy that no one has yet tried to understand how diet has been influencing genome modification of organisms over time, and therefore, the present study attempts to take this evolutionary perspective into account.

In regard to amino acid utilization, organisms may be divided into prototrophic and auxotrophic. Prototrophic organisms are considered the ones able to synthesize all amino acids required to constitute their proteins, via *de novo* pathways. Besides them, auxotrophic organisms are considered the ones that have lost some of their enzymes necessary for intermediate steps during the biosynthetic pathways to produce one or more amino acids. It is well known that metazoan organisms are auxotrophic for a number of amino acids, and these so-called essential amino acids (EAA) must be obtained through the digestion of proteins from other organisms. The set of EAA might not be the same for all Metazoa, and for the purpose of this analysis the most studied

human set was adopted. When thinking about auxotrophy, a number of questions come to mind, such as: 1) Why some amino acid pathways have been lost while others have not? 2) Is there any common pattern joining together the EAA in complex metazoans? Furthermore, once these amino acids have been lost, 3) would evolutionary processes create some differentiation in the usage of amino acids in prototrophic and auxotrophic organisms? 4) Have auxotrophic metazoan organisms been modifying the amino acid content of their proteins over evolutionary time in order to use less EAA, being then more independent of diet? Initial studies conducted by our research group seem to answer "yes" to the last question. By creating an essentiality index, we have preliminarily analyzed the proteome of 9 eukaryotic organisms to evaluate differences in their EAA usage (Prosdocimi and Ortega, 2005; Prosdocimi et al., 2007). Metazoan auxotrophic organisms have been shown to use less EAA than do non-metazoan prototrophic organisms, although some metazoans - such as the worm model *Caenorhabditis elegans* - have shown a non-metazoan-like pattern, probably due to some constraints associated with its diet *C. elegans* is a saprophyte that eats bacteria, a source of all amino acids present in prototrophic organisms. Therefore, the non-metazoan pattern must be due to a lack of constraints.

Giving continuity to this research field in the study of amino acid utilization by eukaryotes, we present here an analysis of essentiality indexes of protein domains (DM). Since protein DM must show specific molecular functionalities, it would be expected that they would maintain a number of EAA necessary for their functions, while extra-domain (ED) regions would be able to change more freely their amino acid content from EAA to non-essential amino acids (NEAA). In this study, domains are operationally defined as segments aligned with RPS-BLAST to members of the Conserved Domain Database (CDD), under the alignment conditions defined above. They could coincide with functional or structural units, but this would depend on their alignment to such elements contained in the CDD. ED are defined as regions not masked by those alignments. For example, if a given polypeptide chain has two well-defined domains, the regions which participate in neither of the domains are considered an ED. After mapping domain regions in the complete proteome from a number of eukaryotic organisms downloaded from the COG (KOG division) (Tatusov et al., 2003) and KEGG (KEGG Orthology division) (Kanehisa et al., 2004) databases, we analyzed the amino acid composition of DM and ED as well as of proteins lacking any known protein domain, i.e., without mapped domains (WD). We observed that the amino acid usage in protein DM and ED is not similar for eukaryotes and an index called the ED index - demonstrating the preferential usage of amino acids in EDs - has been established to differentiate domains from extra-domains in eukaryotic genomes. It has been shown that DMs are enriched with EAAs, and we suggest that the absence of these amino acids in extra-domains could be considered an ancient character that precedes the origin of auxotrophy in eukaryotic organisms. This study also shows evidence and examples that characterize a set of EAAs as being more hydrophobic and more frequently found in the interior of globular proteins. Finally, we discuss whether these particular characteristics found in EAA may be related to the fact that their biosynthetic pathways have been lost in complex eukaryotes.

MATERIAL AND METHODS

Database download

We downloaded the complete proteome of all eukaryotic organisms from the KEGG Or-

thology (KO) (Kanehisa et al., 2004) database. The eukaryotic version of the COG database (Tatusov et al., 2003) was also downloaded.

Domain masking and filtering

Using the RPS-BLAST algorithm and the CDD (Marchler-Bauer et al., 2007), the conserved domains within these proteins were mapped. Since RPS-BLAST identifies even partial regions of protein domains, an additional filtering procedure was also performed to classify only *bona fide* domains in the proteins analyzed. Therefore, only those domains found in the CDD demonstrating coverage over 75% and identity over 50% when compared to an actual protein domain were taken into account (see Figure 1). Moreover, organisms showing less than 500 proteins with valid domains were excluded from the analysis. Thus, 17 organisms of 6 phyla (two non-metazoan phyla - Strepthophyta and Ascomycota, and four metazoan phyla - Nematoda, Arthropoda, Chordata, and Echinodermata) from the KO database were analyzed, whereas 6 organisms of 5 phyla from the KOG database were also studied (two non-metazoan phyla - Strepthophyta and Ascomycota, and three metazoan phyla - Nematoda, Arthropoda, Chordata). A total of 62,450 proteins from the KO database and 60,758 proteins from the KOG were analyzed.

PERL algorithms and formulas

The ratio of EAA usage (essentiality index) was calculated for each organism. The ED index was calculated taking into account the number of prevalent amino acids in protein extra-domains (PQS) divided by the number of amino acids found to occur more frequently in domain regions (YFIVW). The ED index was calculated for DM, ED and WD for each phylum in KO and KOG.

Stacked bars on the comparison of amino acid usage among organisms' phyla

The usage of all 20 amino acids in domain and extra-domain regions was calculated for each phylum. Then, the usage of each amino acid (by phylum) in domains was divided by their usage in extra-domains and retrieved the log2 of this ratio. Amino acids prevalent in protein domains would show a positive log2 value, whereas amino acids prevalent in extra-domains would have negative values.

Statistical analysis

The standard error of the mean was calculated and shown for data in Figures 3 and 5. An unpaired *t*-test was performed in order to test for significant differences among the data shown in Figure 3.

RESULTS

Protein domains are frequently stored in a number of distinct databases based on their shared characteristics. A combination of three of them, namely, pFAM, SMART and COG, is currently used by the program RPS-BLAST running at the NCBI website, although the RPS-BLAST

algorithm can also be downloaded and executed locally. This combination is known as the CDD (Marchler-Bauer et al., 2007), and we have used it to map domains in eukaryotic proteins using RPS-BLAST. However, before beginning the analyses of essentiality and amino acid usage in domains and extra-domains, we had searched for the best RPS-BLAST parameters to map CDD domains in actual proteins. Considering e-value maximum cut-off, identity and coverage when compared to CDD original domains, the variation of each of these parameters was tested (Figure 1 A-C), one at a time, and the number of amino acids that fell in each of the following categories was plotted: DM, ED and WD. ED regions were defined as the complete extension of the protein with exception of the regions in which domains have been mapped. The evaluation shown was conducted for proteins present in the KEGG Orthology dataset (Figure 1), and similar results were obtained for proteins from the KOG dataset (data not shown). Based on Figure 1, the RPS-BLAST parameters were defined as: i) maximum e-value equal to 1e-03; ii) minimum identity with a CDD
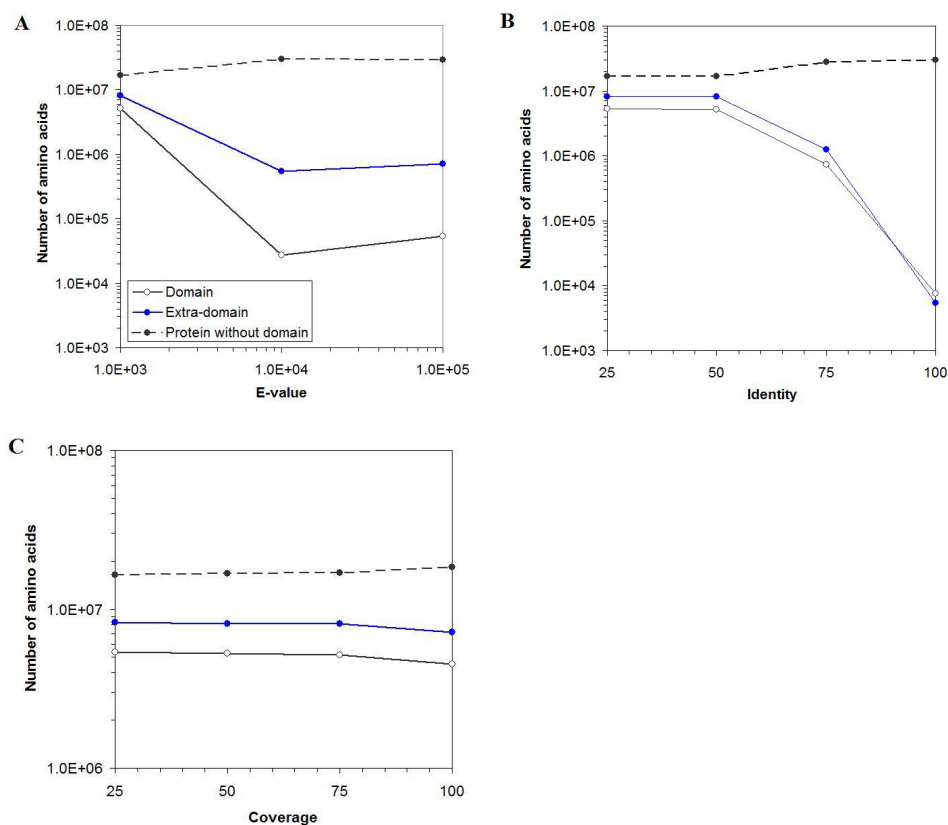

Figure 1. Study of RPS-BLAST main parameters used for domain validation. The figures show the number of amino acids present in domains, extra-domains and proteins without domains when each parameter (e-value, similarity and coverage) has been modified. Figure 1A shows the number of amino acids in each of the three categories listed when E-value cut-off is used as a variable. Figure 1B describes the variation in identity cut-off, and Figure 1C defines the variation in domain coverage. The conditions chosen were the ones that express the saturation number of amino acid residues.

consensus domain equal to 50%, and iii) minimum domain coverage equal to 75% - so that the alignment length must be at least 75% of the total domain length stored in the CDD. The saturation of the number of residues mapped at the chosen cut-offs indicates their appropriateness. Manual inspection of some alignments also supported the selection of these parameters. After domain mapping, we divided our dataset into domain versus extra-domain segments, while proteins without any mapped domain were treated separately.

Once the domains have been mapped using the validated criteria, the hypothesis was tested related to the usage of EAA in protein DM versus ED and proteins WD. For this purpose, we determined the "essentiality index" as described in a previous study (Prosdocimi and Ortega, 2005). Briefly, the index determines the ratio of EAA to all other amino acids when arginine is excluded since, besides being synthesized in higher organisms, its use during growth and proliferation is so required that it is frequently considered semi-essential (Rose et al., 1948; Nakagawa et al., 1963).

At first, the observations made here point to an unexpected result: amino acid usage in WD (essentiality index from 0.51 to 0.55; X-axis in Figure 2) is not similar to the usage
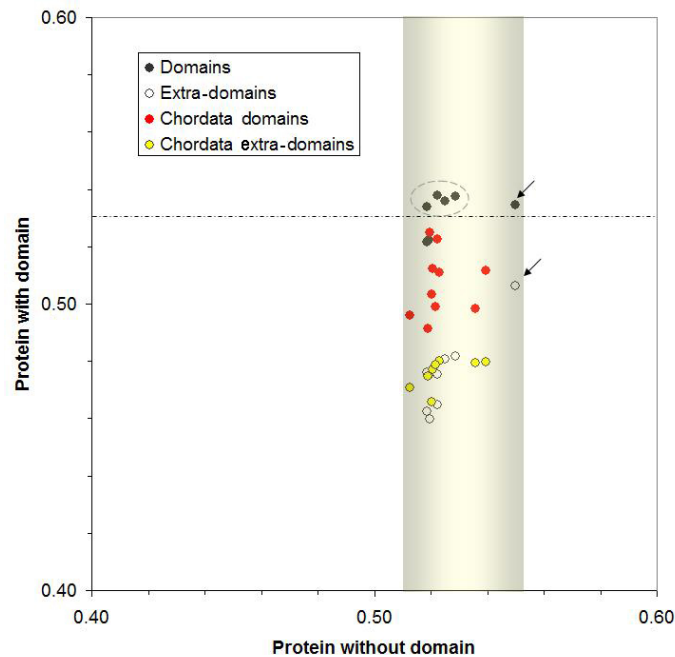
Figure 2. Essentiality index (see Methods) was determined for proteins without domains (X-axis) and for domains and extra-domains (indicated by symbols) in proteins with domains (Y-axis). Domains from CDD were mapped in all KO organisms. Each dot refers to an entire proteome. Data show a similar essentiality index value for extra-domains for all organisms (Y-axis, white and yellow dots). A smaller essentiality index was found in domains for the Chordata clade (Y-axis, red dots) than for non-Chordata (Y-axis, black dots). Higher essentiality in domains (Y-axis) was shown in the prototrophic organisms (inside the dashed circle). *Caenorhabditis elegans* (dots pointed out by arrows) showed anomalous data: high essentiality index in domains (Y-axis, black circle pointed out by arrow) and also in extra-domains (Y-axis, open circle pointed out by arrow). Dashed line separates domain essentiality index in auxotrophic organisms and *C. elegans* from the essentiality index of Coelomata (below the dashed line).

in ED (essentiality indexes below 0.49 in Y-axis; *Caenorhabditis elegans* anomalous data excluded), but it is closest to the DM one (essentiality from 0.49 to 0.55). Thus, it seems that domain mapping has been avoided in databases when the entire protein behaves as a domain. This point will be discussed below (see Figure 5). Essentiality in WD was not as broad as expected if these proteins were composed of uncovered domains (not yet described) and ED.

Conversely, they show an amino acid usage compatible with being composed of large domains. These findings made us change the focus of the investigation slightly in order to distinguish which amino acids were favored or depleted in domains, so that we might be able to further investigate the putative presence of uncovered domains in proteins without domains.

The hypothesis precluded that the usage of EAA in DM might be reflecting the conservation of critical residues, while in ED a pressure for the preferential usage of NEAA could have been applied during evolution. The data shown in Figure 2 are in agreement with this hypothesis. It is observed that ED shows the lowest essentiality index. We did not observe a bias for any high-level clade analyzed with respect to ED essentiality, and it may be seen as evidence that the amino acid usage of ED has always been around the same values, even in ancestral eukaryotic genomes, which would indicate that the differential usage in ED preceded the origin of auxotrophy. C. elegans showed the highest essentiality value of ED and also a high value of DM (see arrows in Figure 2), confirming its anomalous position in regard to EAA usage observed previously (Prosdocimi and Ortega, 2005; Prosdocimi et al., 2007). Prototrophic organisms have also been shown to have the highest levels of essentiality in DM (Figure 2, data inside the dashed circle), corroborating the theory that auxotrophic ones are changing the EAA in their proteins to NEAA, at least in protein domains. However, besides confirming the suggestion that DM would show higher essentiality compared to ED, a clear tendency was observed for a discrete clade, that is, chordates, in which the DM shows a lower essentiality. To our knowledge, this is the first attempt to characterize evolutionary usage of amino acids in protein domains, and our finding suggests that either natural selection pressure has acted on chordate domains or it only occurred due to a clade-specific radiation bias. However, it is remarkable that such bias has not affected ED usage, which tends to support (although not proving) the first possibility.

We then set up to test if the lower essentiality shown in chordate DM was biased by domains that were exclusive to the Chordata clade. For this purpose, in chordate proteomes only the chordate exclusive domains were mapped in order to calculate their essentiality. The results suggest that this bias is distributed over the entire proteome, since mapping chordate exclusive domains yield indexes sometimes higher and sometimes lower than the mapping of all domains, depending on the organism analyzed (Figure 3). However, the chordate essentiality index was shown to be always lower than 0.53 - the overall essentiality of non-Chordata domains (dashed line in Figure 2). For comparison, the inset in Figure 3 shows the proportion between exclusive and non-exclusive domains in chordate organisms. The overall essentiality measured for KO chordate domains is 0.5047, and for KOG chordate domains it is 0.5005. Thus, although it is reasonable to observe a bias in chordate proteome for domains originating in this clade, the lowest essentiality seems to be distributed over the entire set of DM regions in chordate proteins.
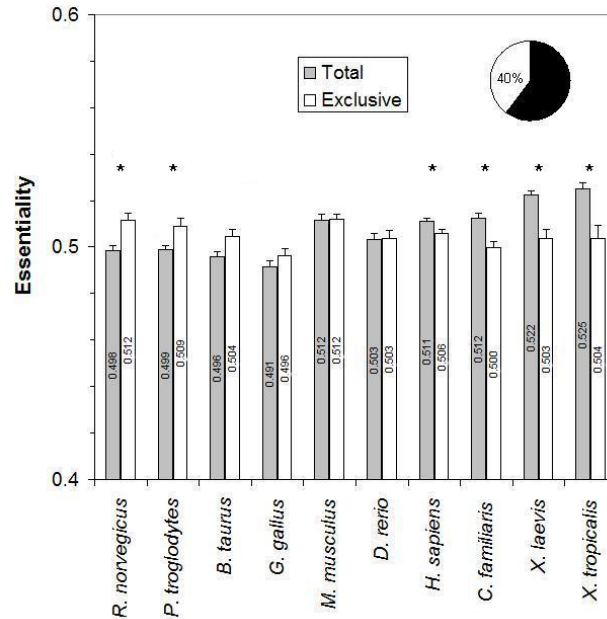
Figure 3. Essentiality index values for domains exclusive for chordates (white columns) and essentiality index values for all domains found in chordates (gray columns). Depending on the organism analyzed, the indexes yielded by domains exclusive for chordates show values either higher or lower than 0.53 (overall domain essentiality average in eukaryotes). These data indicate that the essentiality bias observed in chordate domains (Figure 2) cannot be explained by the usage of chordate-specific domains. The inset shows the overall proportion between exclusive (40%) and non-exclusive domains in Chordata organisms. *P < 0.05 for statistically significant differences.

We then set up to compare the usage of distinct amino acids in DM versus ED, and the data are presented in Figure 4. A stacked bar plot (see Methods) was used in order to better cluster the differential usage, thus the DM/ED ratio of usage is represented by the size of each colored bar. This analysis pointed out three amino acids used more in ED than in DM, yielding negative log2 ratios: P, Q, S. All them are grouped in the NEAA set enclosed in the blue box shown on the X-axis (Figure 4). Arginine (R) is not grouped since it is considered to be semi-essential, and it does not seem to show any bias. The set of amino acids most present in DM included Y, F, I, V, and W. Tyrosine (Y) is an NEAA, but since it is directly derived from F, it could behave as an EAA as well, where its synthesis could deplete the stock of F. The same fact applies to C which is directly derived from M, although M did not reach a difference that is as conspicuous as the others prevalent in DM. Thus, we decided to generate a novel index determined by the ratio of usage of PQS over YFIVW, which would reflect the prevalence of amino acid usage in ED over DM, returning positive values. This index should then reflect enrichment by residues external to the domain structure. We preliminarily named it ED index, since the biological meaning of enrichment needs further investigation (see some lines of evidence below). Figure 5 shows the ED index for DM, ED and WD in the KEGG Orthology set of proteins, and similar results were obtained with the KOG set (data not shown). For all clades, the ED index was higher for ED as compared to DM segments. Moreover, in agreement with the observations for essentiality, proteins without domains showed the lowest ED index for all clades analyzed, reinforcing the discovery that those proteins are more likely of being structured as large domains.
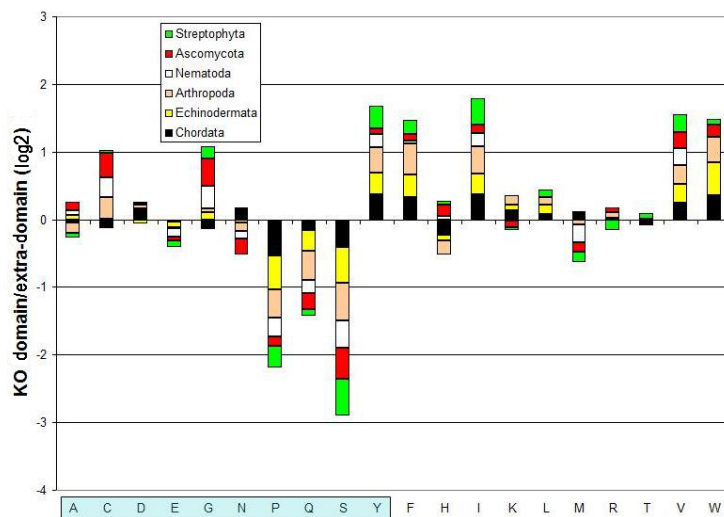
Figure 4. Comparison of the amino acid usage between domains and extra-domains. The stacked bar plot shows the following ratio log2 (DM/ED), where DM is the amino acid usage in domains and ED is the amino acid usage in extra-domains. The blue colored one-letter coded amino acids are the non-essential ones. Those bars above the X-axis indicate a higher usage in domains, whereas those below the X-axis indicate a higher usage in extra-domains. P, Q, S were found to occur more in extra-domains whereas Y, F, I, V, and W were found to occur more frequently in protein domains.
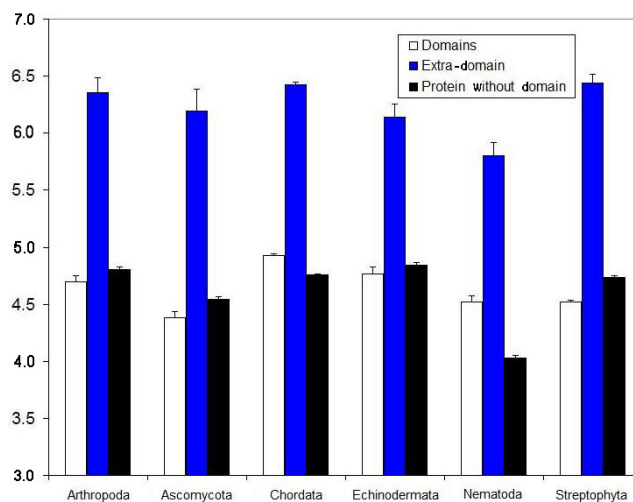
Figure 5. Extra-domains (ED index) values for domains, extra-domains and proteins without domains. The ED index expresses the ratio of the usage of those amino acids found to occur more frequently in extra-domains to the usage of those amino acids found more commonly in domains. The ED index was higher for extra-domains than for DM in all organisms, which agrees with previous observations. Interestingly, the ED index values for proteins without domains are more similar to those values found for domains, which suggests that proteins without domains look like a big single domain.

A preliminary analysis of the biological meaning of the ED index is represented in Figure 6. Two distinct views of the MAX/MAX dimmer bound to DNA are shown in Figure 6A and B. PQS were represented as space-filling blue atoms while YFIVW were represented similarly but in white. It can be noticed that blue atoms are more external to the structure. Similar representations were obtained for myoglobin (Figure 6C and D). Thus, it is plausible that a lower ED index represents the amount of residues in a core domain of a globular structure, while a higher index would represent a region in the protein with abundant contacts with the solvent. Further characterizations of these structural features are in the process of being reported elsewhere. Thus, the data seem to indicate that a set of amino acids implicated in the maintenance of domain structure such as YFIVW underwent a positive selection pressure to remain being used in DM despite not being synthesized by auxotrophic metazoan organisms. However, the origin of the NEAA character has been preferentially concentrated in a set not implicated in domain maintenance, preferentially the more externally occurring residues in a 3-D structure.
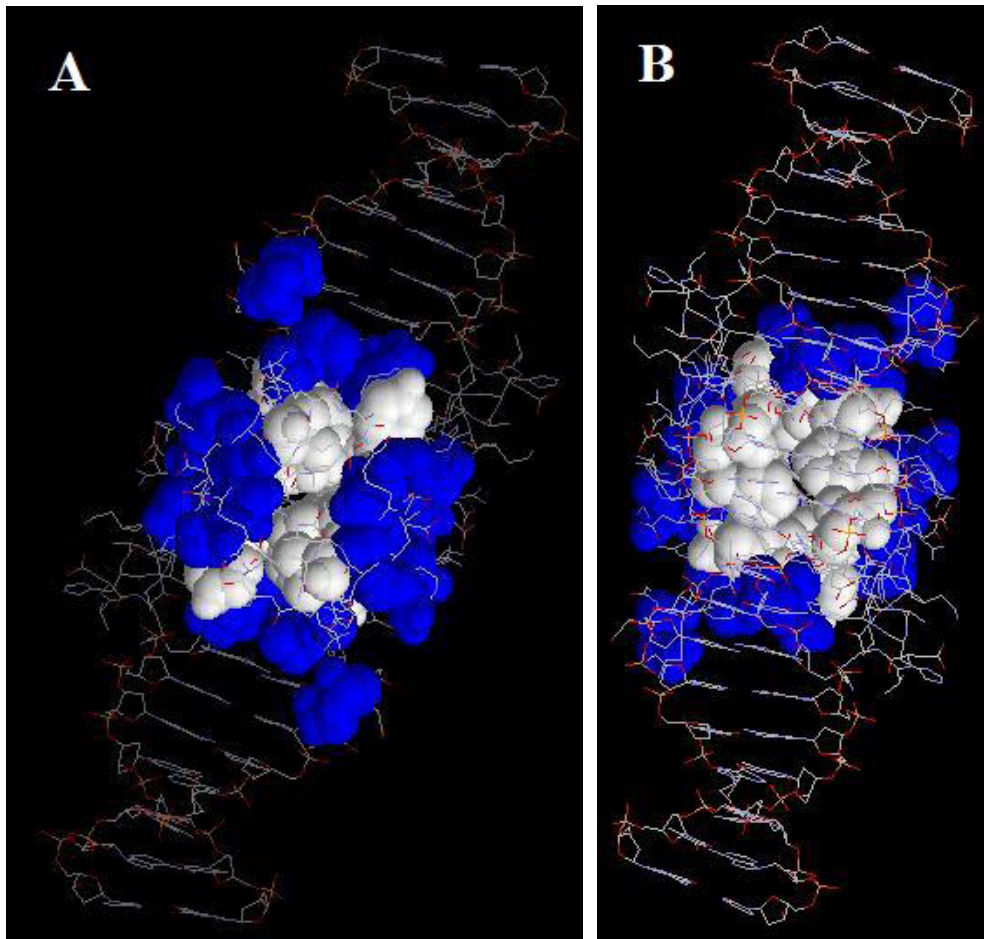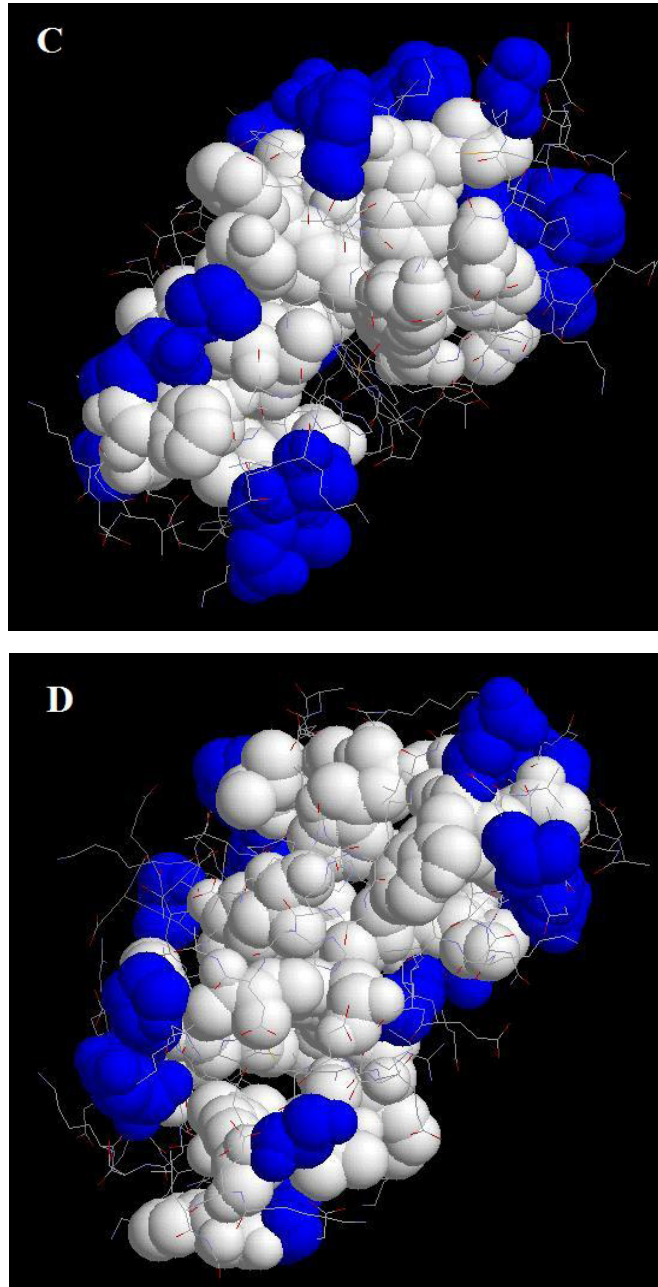

Figure 6. Continued on next page.

Figure 6. Distribution of the two distinct groups of amino acids: frequently found in domains (shown in white) and often found in extra-domains (shown in blue). **A.** and **B.** show two distinct views for MAX/MAX protein (PDB ID 1AN2A). **C.** and **D.** show two distinct views for a human myoglobin (PDBID 2MM1). In all four figures, those amino acids that are found in domains are located in the proteins' core, whereas those amino acids commonly found in extra-domains are more likely to be found in contact with the solvent.

DISCUSSION

We began our study with the hypothesis that the loss of enzymes for *de novo* synthesis of amino acids affects and modifies the usage of these pathway-depleted amino acids in auxotrophic organisms. Data provided here and elsewhere (Prosdocimi and Ortega, 2005; Prosdocimi et al., 2007) tend to support this hypothesis, although we have shown that many other interesting observations could be made from the study of amino acid usage in protein domains of eukaryotic organisms. Based on KO and KOG data about complete eukaryotic proteomes, CDD domains were mapped using *bona fide* parameters of RPS-BLAST (Figure 1) and the EAA usage in regions of DM, ED and also WD was evaluated. At first, differences between the usages of EAAs were found between prototrophic and auxotrophic taxa, mainly in regions of protein domains (Figure 2). The lower essentiality values in Chordata were shown not to be due to exclusive domains (Figure 3). Two main topics shall lead discussions about the data shown in Figure 2: i) why extra-domains show a smaller proportion of EAA when compared to domains and ii) why there is a clade-specific differentiation in EAA usage in protein domains.

Differences in amino acid usage of domains and extra domains (Figures 4, 5 and 6) may be explained by the fact that protein domains probably need to be more tri-dimensionally static, as well as protein without domains. Supporting this hypothesis, we showed evidence (Figure 6) that domain-specific amino acids are inside the globular structures of proteins while ED-specific ones are shown on the surface of proteins, allowing interaction with the solvent and other biological molecules in the medium.

Since a small difference has been shown in EAA usage among Chordata phyla, why has this difference not been demonstrated in other auxotrophic eukaryotic taxa? In fact, it has been. Only one single auxotrophic organism studied has shown an EAA usage similar to prototrophic ones, and *Drosophila melanogaster* clusters with Chordata, being the single black point below the 0.53 line in Figure 2. *C. elegans* is the single auxotrophic organism displaying an EAA usage higher than 0.53. This observation is in accordance with our previous studies (Prosdocimi and Ortega, 2005; Prosdocimi et al., 2007), and we have suggested previously that an absence of genomic modification in this worm is probably due to the fact that this organism has not been under dietary constraints during its evolutionary time. The selection pressure to change amino acids in proteins from EAA to NEAA only makes sense when some organisms fail to produce correct proteins in the absence of food. If *C. elegans*' direct ancestors have never undergone starvation along their evolutionary time, it has never been subject to a genomic pressure to change the amino acid content of their proteins. In fact, nematodes are known to eat mainly bacteria and they are a highly diverse and successful group of organisms. Moreover, *C. elegans* does not group within a broader clade, Coelomata, which includes the Chordata and other organisms where DM essentiality is below the dashed line in Figure 1. It is possible that the strongest evolutionary pressure might be related to the conquest of the terrestrial environment.

Finally, we address the question about why the amino acids shown to be concentrated in DM are the ones whose biosynthetic pathways could have been deleted in a given ancestral. Why is a group of EAA prevalent in protein domains? Would it have any relationship with joining together i) the fact that an amino acid is originally present in protein domains and ii) the fact that a highly diverse group of organisms (metazoans) has lost the biosynthetic pathways for using these amino acids? Would they be less relevant for organisms' homeostasis? Or would it be simply a random and non-adaptive pattern? We still do not know the answer

to these questions and we are currently conducting further investigations to date precisely the time at which pathway deletions occurred. Since there are no known intermediates showing only some deletions of amino acid biosynthetic pathways, we suggest a putative event called the Great Amino Acid Pathway Deletion, in which most of these pathways have been deleted to respond to some highly impacting environmental modification of an ancestral metazoan. We are currently developing phylogenies based on the clustering of amino acid usage, aiming to date precisely in which ancestral organisms this great deletion occurred. Therefore, we still need more data to understand the origin of auxotrophy in metazoans and to provide a more concise picture about this new exciting field of evolutionary nutrigenomics.

ACKNOWLEDGMENTS

REFERENCES

Afman L and Muller M (2006). Nutrigenomics: from molecular nutrition to prevention of disease. *J. Am. Diet. Assoc.* 106: 569-576.

Castle D and Ries NM (2007). Ethical, legal and social issues in nutrigenomics: the challenges of regulating service delivery and building health professional capacity. *Mutat. Res.* 622: 138-143.

Chadwick R (2004). Nutrigenomics, individualism and public health. *Proc. Nutr. Soc.* 63: 161-166.

Chavez A and Munoz de CM (2003). Nutrigenomics in public health nutrition: short-term perspectives. *Eur. J. Clin. Nutr.* 57 (Suppl 1): S97-100.

Davis CD (2007). Nutrigenomics and the prevention of colon cancer. *Pharmacogenomics* 8: 121-124.

Davis CD and Milner J (2004). Frontiers in nutrigenomics, proteomics, metabolomics and cancer prevention. *Mutat. Res.* 551: 51-64.

Dawson KA (2006). Nutrigenomics: feeding the genes for improved fertility. *Anim. Reprod. Sci.* 96: 312-322.

Delcourt C (2007). Application of nutrigenomics in eye health. *Forum Nutr.* 60: 168-175.

El-Sohemy A, Stewart L, Khataan N, Fontaine-Bisson B, et al. (2007). Nutrigenomics of taste - impact on food preferences and food production. *Forum Nutr.* 60: 176-182.

Fenech M (2005). The Genome Health Clinic and Genome Health Nutrigenomics concepts: diagnosis and nutritional treatment of genome and epigenome damage on an individual basis. *Mutagenesis* 20: 255-269.

Fenech M (2008). Genome health nutrigenomics and nutrigenetics - diagnosis and nutritional treatment of genome damage on an individual basis. *Food Chem. Toxicol.* 46: 1365-1370.

Ferguson LR, Shelling AN, Lauren D, Heyes JA, et al. (2007). Nutrigenomics and gut health. *Mutat. Res.* 622: 1-6.

Ghosh D, Skinner MA and Laing WA (2007). Pharmacogenomics and nutrigenomics: synergies and differences. *Eur. J. Clin. Nutr.* 61: 567-574.

Junien C and Gallou C (2004). Cancer nutrigenomics. *World Rev. Nutr. Diet.* 93: 210-269.

Kanehisa M, Goto S, Kawashima S, Okuno Y, et al. (2004). The KEGG resource for deciphering the genome. *Nucleic Acids Res.* 32: D277-D280.

Kaput J, Perlina A, Hatipoglu B, Bartholomew A, et al. (2007). Nutrigenomics: concepts and applications to pharmacogenomics and clinical medicine. *Pharmacogenomics* 8: 369-390.

Low YL and Tai ES (2007). Understanding diet-gene interactions: lessons from studying nutrigenomics and cardiovascular disease. *Mutat. Res.* 622: 7-13.

Marchler-Bauer A, Anderson JB, Derbyshire MK, Weese-Scott C, et al. (2007). CDD: a conserved domain database for interactive domain family analysis. *Nucleic Acids Res.* 35: D237-D240.

Mariman EC (2007). Nutrigenomics in perspective. *Pharmacogenomics* 8: 421-424.

Martin KR (2007). Using nutrigenomics to evaluate apoptosis as a preemptive target in cancer prevention. *Curr. Cancer Drug Target* 7: 438-446.

Muller M and Kersten S (2003). Nutrigenomics: goals and strategies. *Nat. Rev. Genet.* 4: 315-322.

Nakagawa I, Takahashi T, Suzuki T and Kobayashi K (1963). Amino acid requirements of children: minimal needs of tryptophan, arginine and histidine based on nitrogen balance method. *J. Nutr.* 80: 305-310.

Ozdemir V and Godard B (2007). Evidence-based management of nutrigenomics expectations and ELSIs. *Pharmacogenomics* 8: 1051-1062.

Palou A, Bonet ML, Pico C and Rodriguez AM (2004). Nutrigenomics and obesity. *Rev. Med. Univ. Navarra* 48: 36-48.

Peregrin T (2001). The new frontier of nutrition science: nutrigenomics. *J. Am. Diet. Assoc.* 101: 1306.

Prosdocimi F and Ortega JM (2005). Diet as a Pressure on the Amino Acid Content of Proteomes. In: Advances in Bioinformatics and Computational Biology. Vol. 3594. Springer, Berlin, Heidelberg, 153-159.

Prosdocimi F, Mudado MA and Ortega JM (2007). A set of amino acids found to occur more frequently in human and fly than in plant and yeast proteomes consists of non-essential amino acids. *Comput. Biol. Med.* 37: 159-165.

Rose WC, Oesterling MJ and Womack M (1948). Comparative growth on diets containing ten and nineteen amino acids, with further observations upon the role of glutamic and aspartic acid. *J. Biol. Chem.* 176: 753-762.

Roy N, Barnett M, Knoch B, Dommels Y, et al. (2007). Nutrigenomics applied to an animal model of inflammatory bowel diseases: transcriptomic analysis of the effects of eicosapentaenoic acid- and arachidonic acid-enriched diets. *Mutat. Res.* 622: 103-116.

Tatusov RL, Fedorova ND, Jackson JD, Jacobs AR, et al. (2003). The COG database: an updated version includes eukaryotes. *BMC Bioinformatics* 4: 41.

Trayhurn P (2003). Nutritional genomics - "Nutrigenomics". *Br. J. Nutr.* 89: 1-2.

van Ommen B and Stierum R (2002). Nutrigenomics: exploiting systems biology in the nutrition and health arena. *Curr. Opin. Biotechnol.* 13: 517-521.

Ye J, Cui X, Loraine A, Bynum K, et al. (2007). Methods for nutrigenomics and longevity studies in *Drosophila*: effects of diets high in sucrose, palmitic acid, soy, or beef. *Methods Mol. Biol.* 371: 111-141.