Occurrence of protein disulfide bonds in different domains of life: a comparison of proteins from the Protein Data Bank

I. Bošnjak1,4, V. Bojović2, T. Šegvić-Bubić3 and A. Bielen1,4

1Laboratory for Biology and Microbial Genetics, Department of Biochemical Engineering, Faculty of Food Technology and Biotechnology, Pierottijeva 6, 10000 Zagreb, Croatia, *NMR Centre, Ruder Bošković Institute, Bijenička cesta 54, 10000 Zagreb, Croatia and 1Laboratory for Aquaculture, Institute of Oceanography and Fisheries, Setalište Ivana Mestrovica 63, 21000 Split, Croatia

To whom correspondence should be addressed. E-mail: ibosnjak@pbf.hr (I.B.); abileen@pbf.hr (A.B.)

Received July 9, 2013; revised October 11, 2013; accepted December 10, 2013

Edited by Andrew Miranker

Disulfide bonds (SS bonds) are important post-translational modifications of proteins. They stabilize a three-dimensional (3D) structure (structural SS bonds) and also have the catalytic or regulatory functions (redox-active SS bonds). Although SS bonds are present in all groups of organisms, no comparative analyses of their frequency in proteins from different domains of life have been made to date. Using the Protein Data Bank, the number and subcellular locations of SS bonds in Archaea, Bacteria and Eukarya have been compared. Approximately three times higher frequency of proteins with SS bonds in eukaryotic secretory organelles (e.g. endoplasmic reticulum) than in bacterial periplasmic secretory pathways was calculated. Protein length also affects the SS bond frequency: the average number of SS bonds is positively correlated with the length for longer proteins (>200 amino acids), while for the shorter and less stable proteins (<200 amino acids) this correlation is negative. Medium-sized proteins (250–350 amino acids) indicated a high number of SS bonds only in Archaea which could be explained by the need for additional protein stabilization in hyperthermophiles. The results emphasize higher capacity for the SS bond formation and isomerization in Eukarya when compared with Archaea and Bacteria.

Keywords: Archaea/Bacteria/Eukarya/PDB/SS bond

Introduction

Among 20 amino acids found in living organisms, cysteine has a unique ability to form covalent disulfide bonds (SS bonds) between the thiol groups. This important post-translational modification has two main functions: structural SS bonds stabilize the mature proteins’ three-dimensional (3D) structure, and redox-sensitive SS bonds exhibit physiologically relevant redox activity (Buchanan and Balmer, 2005; Yang et al., 2007; Fan et al., 2009; Wouters et al., 2010). Disulfide bonds are strongly conserved among related proteins which emphasizes their biological importance (Thornton, 1981).

The positive effect of SS bond introduction into the protein structure is also shown by the fact that the frequency of cysteines has substantially increased during evolution of living organisms (Brooks and Fresco, 2002; Jordan et al., 2005), and it is estimated that ~50% of cysteines form disulfide bonds (Wong et al., 2011). Also, it was recently suggested that disulfide bonds are positively selected in shorter proteins (<200 residues), because they undergo stronger selection for 3D structure stabilization, due to the reduced number of contacts per residue (Bastolla and Demetrius, 2005).

In living organisms, SS bonds are formed by oxidation of thiol groups of cysteines during protein folding. After initial SS bond formation, they can be exchanged when formed between non-native cysteine pairs until the correct connectivity has been achieved. This process is called isomerization and plays a vital role in protein folding, especially for proteins with multiple cysteines (Sevier and Kaiser, 2002; Gleiter and Bardwell, 2008; Ito, 2010; Depuydt et al., 2011). Further, recent studies have shown that isomerization can also act as a regulator of protein function (Wouters et al., 2010). Therefore, intrachain disulfide bonds in mature proteins can be consecutive (between two sequential cysteines in protein primary structure) or non-consecutive (between two cysteines that are separated by sequence containing one or more disulfide-bonded cysteines) as depicted in Fig. 1.

Generally, the process of disulfide bond formation is catalyzed by thiol-disulfide oxidoreductases and is usually confined to relatively oxidizing environments. Nevertheless, there are differences in SS bond formation pathways between different types of cellular organization, with most experimental data coming from Gram-negative bacteria and eukaryotes. In Gram-negative bacteria, disulfide bonds are predominantly formed in the periplasmic space, the process being catalyzed by several thiol-disulfide oxidoreductase proteins: DsBA, DsbB, DsbC, DsbD, etc. In contrast, in eukaryotic cells SS bonds are formed mostly in the lumen of rough endoplasmic reticulum, with protein disulfide isomerase (PDI) being the key enzyme identified so far (Sevier and Kaiser, 2002; Heras et al., 2007). Much less is known about other cell types. The place and machinery for the formation of disulfide bonds in Gram-positive Actinobacteria and Firmicutes are mostly unknown, since they lack the protected compartment of periplasm, and the majority also does not have DsBA and/or DsbB homologs. Recently, eukaryotic enzyme vitamin K epoxide reductase was proposed to be a key disulfide bond formation enzyme in bacteria lacking DsbB (Dutton et al., 2008). A specific protein named protein disulfide oxidoreductase was recognized in hyperthermophilic Bacteria and Archaea as a potential catalyst of intracellular disulfide bonds (Ladenstein and Ren, 2008).
Despite different SS bond formation enzymes and compartments, disulfide bonds from different taxonomic groups were always studied separately (Dutton et al., 2008; Ladenstein and Ren, 2008; Daniels et al., 2010; Wong et al., 2011). Therefore, we aimed to compare the number and arrangement (consecutive or non-consecutive) of SS bonds in three domains of life (Archaea, Bacteria and Eukarya). We used Protein Data Bank (PDB) as the source database as it is a major protein structure repository and contains experimental data on SS bonds (Rose et al., 2011). Furthermore, we aimed to explore the effect of protein length and cellular compartment on the SS bond frequency. We have shown that eukaryotes have significantly higher SS bond formation and isomerization capacity than Bacteria and Archaea.

Materials and methods

Protein dataset

A non-redundant list of PDB protein chains with BLAST P-value cutoff of \(10 \times 10^{-7}\) as well as PDB snapshot archive were downloaded. Next, snapshot archive was filtered using the non-redundant list. Information on the disulfide bond number and type (consecutive or non-consecutive, see Fig. 1) was added to every protein chain in the dataset. First, all annotated PDB SS bonds (SSBOND tags) were taken. One outlier was excluded (SS bond distance of 61.86 Å). The mean SS bond length was 2.05 with the standard deviation of 0.11. Only SS bonds that were between 1.62 and 2.49 Å long (mean ± 4 SDs) and assigned to two different cysteine residues were taken into analysis. In the dataset, we kept only proteins that (i) were monomeric (one chain per protein); (ii) native (not synthetic, designed, artificial or hybrid); (iii) taxonomically well-defined [environmental sequences, sequences where source organism is unidentified, sequences with multiple source organisms, or unclassified sequences according to the National Center for Biotechnology Information (NCBI, taxonomy database were excluded]; and (iv) have visible side chains.

To check the identity between sequences in the final dataset, we have run CD-HIT clustering analysis (Huang et al., 2010) on sequences from our dataset, and 99% of sequences were <70% identical. By comparing our dataset with the list of redox-active proteins published previously (Fan et al., 2009), we found only 14 of redox-active proteins (1.4% of the proteins containing SS bonds in the dataset used; \(n = 984\)). Therefore, we presumed that the rest of analyzed proteins contained structural SS bonds.

Finally, to each protein in the refined dataset we have attached the data on length (total number of amino acids per chain, as present in the PDB, e.g. after processing) and taxonomy. Therefore, every protein was assigned to one of the three domains of life: Archaea, Bacteria and Eukarya.

Final dataset used for subsequent statistical analysis is shown in Supplementary File S1.

Statistical analysis

Dataset was preliminarily analyzed to demonstrate its main characteristics: total number of proteins, mean length and fraction of proteins with a minimum of one SS bond. The analysis was done for the whole dataset and for each of the three domains of life: Archaea, Bacteria and Eukarya (Table I).

Total number of SS bonds (consecutive + non-consecutive) and number of non-consecutive SS bonds were presented as mean ± SE. Two-way analysis of variance (ANOVA) was performed to test the differences in the total number of disulfide bonds and the number of non-consecutive SS bonds among protein chain lengths in investigated groups. Before using
ANOVA, normality and homogeneity of variance in the dataset were tested with Shapiro–Wilk test and Levene’s test, respectively. Proteins were grouped into five subsets based on the chain length, i.e. the number of amino acids (1: 1–100; 2: 101–200; 3: 201–300; 4: 301–400; 5: ≥401). The median length of proteins for the whole dataset and for selected groups was calculated. A model Type III sum of squares was used to cope with unbalanced design, while the Tukey–Kramer significant difference test was used to perform post-ANOVA pairwise comparisons among the groups. The probability $P < 0.01$ was considered statistically significant. The general linear model univariate procedure in SAS statistical package was used.

To test the null hypothesis of no association between the groups of organisms (Archaea, Bacteria, Eukarya) and the presence or absence of consecutive and non-consecutive disulfide bonds among all protein chain lengths, Pearson’s goodness-of-fit $\chi^2$ test was performed using Statistical Package for Social Sciences (SPSS) 16.0 (Norusis, 2008).

The two-tailed non-parametric Spearman’s correlation analysis was applied to identify the degree of association between the protein length and the frequency of disulfide bonds in general and the frequency of non-consecutive SS bonds for all groups and individually, for Eukarya, Bacteria and Archaea, respectively. In addition, datasets were divided into short (<200 amino acids) and long chains (<200 amino acids) and separately used for comparison.

**Prediction of protein localization**

To predict localization of proteins, complete protein sequences were extracted from Uniprot matching PDB entries from our database, and then analyzed with TargetP 1.1, TatP 1.0, ChloroP 1.1, SignalP 4.1, TMHMM 2.0, NetNes 1.1, SecretomeP 2.0 (Emanuelsson et al., 2007), big-PI Predictor (Eisenhaber et al., 2000; Eisenhaber et al., 2003; Eisenhaber et al., 2004), PSORTb v.3.0 (Yu et al., 2010), NucPred (Brameier et al., 2007) and LocTree2 (Goldberg et al., 2012). Results were compiled and, to avoid false predictions as much as possible, only proteins getting the same prediction results with different tools were taken for further analysis (see Supplementary File S1). Proteins from Archaea were divided into cytoplasmic, membrane and secreted, from Bacteria into cytoplasmic, inner membrane, periplasmic, outer membrane and secreted, and from Eukarya into cytoplasmic, nucleus, mitochondria, chloroplast and secretory pathway category.

**Results and discussion**

**Characterization of the dataset**

Retrieved dataset of non-redundant PDB protein chains contained 4666 proteins (Table I and Supplementary File S1). Eukarya showed the highest frequency of proteins with at least one SS bond (29.5%). This was our first experimental indication of high prevalence of this post-translational modification in eukaryotes, in comparison with other domains of life.

However, prior to detailed statistical analysis, we critically evaluated our dataset, having in mind biases present in all biological databases (Gerstein, 1998; Peng et al., 2004; Godzik, 2011). In the context of this study, we considered two important sources of bias: (i) over-representation of disulfide bonds in PDB when compared with other databases, mostly due to increased possibility of crystallization of proteins with SS bonds (Gerstein, 1998; Peng et al., 2004); and (ii) over-representation of protein structures from model organisms, such as *Escherichia coli*, *Ababidopsis thaliana* and *Mus musculus* (Godzik, 2011). Regarding (i) we thought reasonable to assume proportional over-representation of SS-bonded proteins from Archaea, Bacteria and Eukarya. Therefore, using PDB as a source database we would be able to compare SS bond frequencies between different domains of life (and not to get precise numerical data on native SS bond frequencies).

To explore further possibility (ii) we analyzed taxonomic origin of proteins in our dataset and observed that species demographics follow the distribution present in the NCBI whole-genome sequencing projects for all the three domains of life (no statistically significant difference as determined using the Mann–Whitney $U$ test, $P > 0.05$; data not shown). Archeal proteins originated from Euryarchaeota (82%) and Crenarchaeota (18%), and in the bacterial subset most common proteins were from Proteobacteria (49%), followed by Firmicutes (24%), Actinobacteria (9%) and other phyyla. Eukaryotic subset was mainly composed of proteins from Metazoa and Fungi (90%) and Viridiplantae (7%).

To check whether subgroups display similar trends in SS bond frequencies as main groups, we have subdivided bacterial and eukaryotic proteins. A previous study indicated that bacterial species exhibit diversity in their capacity for SS bond formation (e.g. at least some obligate anaerobes and obligate intracellular bacteria have reduced ability for SS bond formation, irrespective of taxonomic category). Out of 375 analyzed bacterial genomes, ~65% were predicted to be able to form SS bonds based on the significant increase of even number of cysteines in exported proteins (Dutton et al., 2008). Following this, we have divided proteins from our bacterial dataset into two groups: (i) proteins from species that can oxidatively fold cell envelope proteins according to Dutton et al. (2008) ($n = 1159$), and (ii) proteins from species that do not generally make disulfide bonds ($n = 363$). (Note that only proteins from species analyzed previously (Dutton et al., 2008) were taken into this analysis.) As expected, Group (i) exhibited greater SS bond forming capacity than Group (ii)—10.4% of proteins with a minimum of one SS bond vs. 5.0%, respectively. In comparison, when we analyzed all bacterial proteins as a single group [Group (i), Group (ii) and proteins from species that were not analyzed by Dutton et al. (2008)] fraction of proteins with a minimum of one SS bond was 11.2% (Table I), indicating that our sample was enriched in proteins from bacteria that can oxidatively fold proteins. However, this value is still much lower than for Eukarya (Table I), and taking Bacteria as a single group reflects their reduced SS bond forming capacity in comparison with Eukarya.

Eukaryotic subgroups showed high fractions of proteins with a minimum of one SS bond: 49.7% for Viridiplantae ($n = 298$), 19.7% for Fungi ($n = 458$), 33.7% for Metazoa.
and 25.0% for unicellular Alveolata, Euglenozoa and Amoebozoa (n = 100). These results prompted us to take eukaryotes as a single group that clearly shows higher incidence of SS bonds than bacteria (fraction of proteins with a minimum of one SS bond for Eukarya in total was 29.5%, see Table I). Positive selection for disulfide bonds through eukaryotic evolution and a positive correlation between the rate of half-cystine acquisition and organism complexity have been previously suggested (Wong et al., 2011).

It has to be noted that, due to small total number of archeal proteins in the dataset (n = 280), we did not subdivide this group.

**Effect of protein length and domain of life on SS bond frequencies**

Proteins were divided into length categories, and the average number of SS bonds was calculated for each category (Fig. 2). For total dataset, Bacteria or Eukarya (Fig. 2a, c and d), a similar tendency was observed: average number of SS bonds was highest for shortest (≥50 amino acids) and longest proteins (≥400 amino acids), while proteins of medium length (100–300 amino acids) had the smallest average number of SS bonds. Only proteins from archeal dataset did not follow the described trend (Fig. 2b), which could be due to the small total number of archeal proteins in the dataset. Furthermore, as in Table I, the average number of disulfide bonds is higher in eukaryotic proteins (Fig. 2d) than in other groups (Fig. 2b and c) for all length categories.

To statistically confirm observed differences in SS bond frequencies, an unbalanced two-way ANOVA was used. Based on the observed relations between the protein length and the number of SS bonds (Fig. 2), prior to ANOVA analysis, proteins were divided into five subsets based on the chain length (1: 1–100; 2: 101–200; 3: 201–300; 4: 301–400; 5: ≥401 amino acids). Results are presented in Fig. 3. The frequencies of disulfide bonds were significantly affected both by chain length and by belonging to a particular group, Bacteria, Archaea or Eukarya (two-way ANOVA on disulfide bonds: chain length: F_{4,4651} = 10.9, P < 0.001; domain: F_{2,4651} = 131.9, P < 0.001; interaction: F_{8,4651} = 11.2, P < 0.001). A Tukey–Kramer multiple comparison test for honestly significant difference (HSD) in unbalanced design confirmed that Eukarya had significantly higher number of disulfide bonds among all chain length subsets in comparison with other groups. The chain Subsets 1 and 5 accounted significantly higher number of disulfide bonds compared with the Subsets 2, 3 and 4. Among Subsets 2, 3 and 4, significant differences were not found, while Subset 5 accounted significantly higher number of disulfide bonds compared with the Subset 1. For non-consecutive bonds frequencies, a two-way ANOVA revealed a significant difference among chain length subsets (F_{4,4651} = 4.4, P < 0.001), organism group (F_{2,4651} = 85.4,
Also, Eukarya had significantly higher non-consecutive disulfide bond frequencies among all chain length subsets in comparison with the rest of groups (Tukey HSD test, \( P < 0.001 \)). Again, chain Subsets 1 and 5 accounted for significantly higher number of non-consecutive bonds compared with the Subsets 2, 3 and 4 (Tukey HSD test, \( P < 0.001 \)), while between Subsets 1 and 5 and 2, 3 and 4 significant differences were not found.

A significant association between groups of organisms (Archaea, Bacteria and Eukarya) and the presence or absence of consecutive (\( \chi^2 = 225.6, P < 0.001 \)) and non-consecutive disulfide bonds (\( \chi^2 = 310.6, P < 0.001 \)) among all protein chain lengths was found. Examination of the SS bond frequencies showed that Eukarya possess greater presence of both types of bonds than should be expected, while for Bacteria a lower presence of both types of bonds was observed in comparison with the expected count (Fig. 4).

Finally, we estimated the relation between the protein length and the frequency of disulfide bonds in different domains (Archaea, Bacteria, Eukarya) by calculating Spearman’s correlation coefficients to determine the direction and strength of relationship. Analysis was performed for all protein lengths as well as for subgroups comprising only short (<200 amino acids) or long (≥200 amino acids) chains (Table II). First, we analyzed undivided dataset (all protein lengths). For Eukarya, we observed negative correlation of a number of both disulfide bonds and non-consecutive disulfide bonds with chain length. Negative correlation was more expressed for non-consecutive bonds. For Archaea and Bacteria, the numbers of disulfide bonds were positively correlated with chain length. However, non-consecutive disulfide bonds did not show statistically significant correlations. Interesting results were obtained when short and long chains
were analyzed separately. The number of both disulfide bonds and non-consecutive disulfide bonds was positively correlated with chain length for long chains (with the non-significant exception of non-consecutive SS bonds in Archaea). Considering short chains, the correlation for the number of both disulfide bonds and non-consecutive disulfide bonds was negative and highly significant for Bacteria and Eukarya, with the exception, again, of Archaea where the correlation was non-significant, weak and positive.

Higher abundance of disulfide bonds (both consecutive and non-consecutive) was found in short and long proteins in comparison with medium-sized proteins (Figs. 2 and 3). The number of SS bonds was negatively correlated with protein size for short chains (<200 amino acids), while for longer proteins (>200 amino acids) this correlation was positive (Table II). A previous study by Bastolla and Demetrius (2005) showed positive selection for disulfide bonds in proteins shorter than 200 amino acids because they undergo stronger selection for 3D structure stabilization due to a reduced number of contacts per residue. Therefore, considering short proteins, we confirmed these results and extended it to non-consecutive disulfide bonds, where the effect was even more pronounced (e.g., more negative correlation coefficients, Table II). Contrary to our results, Bastolla and Demetrius (2005) did not find any correlation for longer proteins (>250 amino acids). However, the results cannot be completely compared because of different datasets used for the analysis. Here, we have further shown that the described phenomenon is not confined to a particular group of organisms (Figs. 2 and 3), but could be ascribed to general biophysical properties of proteins. It was recently proposed that protein stability increases linearly with length (Yen et al., 2008; Ghosh and Dill, 2009). Since SS bonds are generally regarded as stabilizing protein 3D structure, introduction of additional SS bonds in short proteins (in particular non-consecutive) could help to overcome their natural instability. Archaea is the only group where we could not confirm higher number of SS bonds in short proteins. Conversely, a relatively high SS bond frequency was observed for medium-sized proteins (250–350 amino acids) (Fig. 2b). This finding is very interesting but should be taken with caution because of the small protein sample for Archaea (n = 280). Nevertheless, many Archaea are hyperthermophilic, and it has been proposed that intracellular proteins of Pyrobaculum aerophilum and Aeropyrum pernix are stabilized by SS bonds (Mallick et al., 2002). Therefore, it could be argued that in hyperthermophiles, even medium-sized proteins could need additional stabilization by SS bonds.

We have demonstrated that eukaryotic proteins in our dataset have (i) higher fraction of proteins with a minimum of one SS bond, and (ii) elevated number of SS bonds per protein when compared with Bacteria and Archaea (Tables I and II, Figs. 2–4). We have confirmed certain differences in SS bond forming capacity in subsets of Bacteria and Eukarya (e.g., we confirmed previously reported under-representation of SS bonds in yeast (Wong et al., 2011) and some bacterial species (Dutton et al., 2008)). Interestingly enough, all bacterial subsets showed a decreased SS bond frequency in comparison with all eukaryotic subsets.

### Table II. Correlation between the protein chain length and the number of disulfide bonds for Archaea, Bacteria and Eukarya

<table>
<thead>
<tr>
<th></th>
<th>All protein lengths</th>
<th>Short chains</th>
<th>Long chains</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>DS</td>
<td>NDS</td>
</tr>
<tr>
<td>Archaea</td>
<td>280</td>
<td>0.19*</td>
<td>0.09</td>
</tr>
<tr>
<td>Bacteria</td>
<td>1916</td>
<td>0.09*</td>
<td>−0.02</td>
</tr>
<tr>
<td>Eukarya</td>
<td>2470</td>
<td>−0.07*</td>
<td>−0.16*</td>
</tr>
</tbody>
</table>

Correlation coefficients were calculated for all proteins, as well as for subsets comprising only proteins shorter or longer than 200 amino acids. n, number of protein chains; DS, disulfide bonds; NDS, non-consecutive disulfide bonds.

*Statistically significant results at \( P < 0.05 \).
Effect of protein subcellular localization on SS bond frequencies

Disulfide bonds are formed mainly in oxidizing environments—periplasm of Gram-negative bacteria and endoplasmic reticulum of eukaryotes (Sevier and Kaiser, 2002; Heras et al., 2007). Therefore, we determined the effect of protein subcellular location on the SS bond frequency. To make localization predictions as correct as possible, various available bioinformatic tools were used, and proteins for which conflicting results were obtained were excluded from the analysis. For each subcellular location, fraction of proteins containing at least one SS bond was calculated, as shown in Fig. 5. We have confirmed the expected localization of majority of disulfide-bonded proteins for Bacteria and Eukarya (Sevier and Kaiser, 2002; Heras et al., 2007) and in this way validated our experimental approach. In Bacteria, periplasmic and secreted proteins more often contained SS bonds than cytoplasmic and membrane. In Eukarya, majority of the secretory pathway proteins contained SS bonds (78%), while in other compartments proteins had less SS bonds. In other compartments, we also found SS bond containing proteins: bacterial cytoplasm (5.2%; \( n = 1083 \)), bacterial inner membrane (10.4%; \( n = 77 \)), eukaryotic cytoplasm (8.8%; \( n = 342 \)) and nucleus (16.0%; \( n = 81 \)) (Fig. 5). Some of those could be explained by erroneous SS bond and localization predictions and/or erroneous SS bond formation in proteins produced in heterologous hosts but it is known that proteins with SS bonds can exist in bacterial and eukaryotic cytoplasm (Stewart et al., 1998; Cumming et al., 2004).

Data for Archaea are not shown because there are only two recent tools that predict protein localization for this domain (Yu et al., 2010; Goldberg et al., 2012) and using this majority of archeal proteins were predicted to be cytoplasmic (237/247) with only eight membrane and two secreted proteins. Disulfide bond containing proteins were found only in the cytoplasmic fraction (14%). This is higher than the corresponding values of SS bonded protein frequency in the cytoplasm of Bacteria and Eukarya (5 and 9%, respectively). Mallick et al. (Mallick et al., 2002) also found SS bonded proteins in archeal cytoplasm using different bioinformatic and experimental methods. Cytoplasmic disulfide bonds are usually rare, and these data indicate some distinctive, but yet unexplored, mechanism for SS bond formation in (hyperthermophilic) Archaea.

Our data for the first time compare different domains in terms of SS bond formation capacity and compartment. Three times higher percentage of proteins with at least one SS bond in eukaryotic secretory pathway vs. bacterial periplasm/secretory pathway emphasizes once more the dominant capacity for SS bond formation in Eukarya. This may be caused by superior properties of eukaryotic PDI in comparison with bacterial Dsb enzymes. Other factor could be that endoplasmic reticulum presents more stable environment for SS bond formation than bacterial periplasm and/or extracellular space. However, these hypotheses need experimental confirmation.

Conclusions

By focusing on structural stabilization of proteins, in the present study we quantified and compared SS bond frequencies between Archaea, Bacteria and Eukarya, illustrating for the first time inherent differences in SS bond forming capacities in proteins from PDB. Eukaryotic proteins have much higher frequency of both consecutive and non-consecutive disulfide bonds in comparison with other groups, and these proteins are mostly confined to the secretory pathway. A comparison of catalytic properties of archeal, bacterial, and eukaryotic thiol-disulfide oxidoreductases still needs to be explored in vitro and in vivo. Generally, the average number of SS bonds is positively correlated with protein length for proteins > 200 amino acids, and negatively for shorter proteins. Only archeal proteins have elevated number of SS bonds in medium-sized proteins. That could be explained by the need for additional protein stabilization in these hyperthermophilic organisms. More experimental data on disulfide-bonded proteins from hyperthermophilic and mesophilic Archaea are needed to elucidate this further.

Supplementary data

Supplementary data are available at PEDS online.

Acknowledgements

The authors thank Tatyana Goldberg for help with LocTree2 prediction tool. Special acknowledgments go to Berislav A. Bošnjak and Christy McAffe, for helping with the linguistic issues throughout the article.

Funding

This study was supported by the Ministry of Science, Education and Sports of the Republic of Croatia (Grant numbers 058 0582261 2246, 001 0010501 0560 and 098 1770495 2919).

References

I. Bošnjak et al.