Evolutionary implication of protein secondary structure among Archaea and Bacteria

P Chellapandi, C Karthigeyen, S Sivaramakrishnan

Citation

Abstract
Molecular structures and sequences are generally more revealing of evolutionary relationships than classical phenotypes, particularly among microorganisms. Archaea are unique group of organisms among other kingdoms, which are widely diverged in metabolic pathways and have well distinguished metabolic genes, particularly involved in methanogenesis, osmoregulation, sulfur toxicity, metal detoxification and stress response. Physiochemical characteristics and secondary structure of some superprotein families from archaeal domain were compared within Archaea and with Bacteria. The results of this work revealed that many of the proteins did not show close proximity to bacterial proteins, but few of them showed evolutionary relationship to proteins with similar biochemical functions in Bacteria. Proteins involved in methanogenesis were highly unique to methanogens in archaeal domain, but proteins responsible for carbon assimilation shared their ancestral behaviors among prokaryotes with reference to similarities in secondary structural elements. CO dehydrogenase, 4-vinyl reductase, allophanate reductase, quinine oxidase, dihydrolipoamide dehydrogenase and NADH oxidase were somehow similar to prokaryotes which indicated a wide diversification and mobilization of such families between Archaea and Bacteria during evolution process. Another one noteworthy of our study is that a stable divergence and slow evolutionary process were occurred in topoisomerase family of Archaea. Perhaps, this attempt can be helpful to understand evolutionary mechanisms of some key metabolic proteins at secondary structural level in Archaea.

INTRODUCTION
The extremophilic nature of many Archaea has stimulated intense efforts to understand the physiological adaptations for living in extremes environments and to probe the potential biotechnological applications of their stable cellular components. Specific archaeal metabolites have also been purified and characterized and some of them have potential industrial uses (Alquéres et al., 2007). About 85% extremophiles are included in Archaea, still these are capable to adopt in new extreme environmental conditions, because the distinguished possessive nature of conserved amino acids and hydrophobic, and structural modulation.

Among species of Archaea, there are a variety of metabolic regimes which differ greatly from the better-known metabolic pathways of the Bacteria and eukaryotes (Brown and Doolittle, 1997; Apic et al., 2008; Ebenhöh et al., 2005). Some of these beneficial features might be either diverged within archaeal domain and or transferred to other kingdoms through horizontal gene transfer. A key to modeling and understanding the evolutionary process is the identification and characterization of the constraints that the evolution perceives as protein diverges. The limitations of widely used models of sequence evolution often prevent more refined and informative questions from being addressed (Goldman et al., 1998).

Generally, protein sequences are more conserved than nucleotide sequences in turn to rather than protein structure. The amino acid variation among closely related protein structures can reveal the presence of structural constraints or plasticity. Secondary structure in proteins consists of local inter-residue interactions mediated by hydrogen bonds. The most common secondary structures are β-helices and β-sheets. Other helices, such as the 3_{10} helix and β-helix, are calculated to have energetically favorable hydrogen-bonding patterns but are rarely if ever observed in natural proteins except at the ends of helices due to unfavorable backbone packing in the center of the helix. Other extended structures such as the polyproline helix and β-sheet are rare in native state proteins but are often hypothesized as important protein folding intermediates. Tight turns and loose, flexible loops link the more “regular” secondary structure elements. The random coil is not a true secondary structure, but is the classes of conformations that indicate an absence of regular secondary
structure. Amino acids vary in their ability to form the various secondary structure elements. Proline and glycine are sometimes known as “helix breakers” because they disrupt the regularity of a helical backbone conformation; however, both have unusual conformational abilities and are commonly found in turns. Amino acids that prefer to adopt helical conformations in proteins include methionine, alanine, leucine, glutamate and lysine; by contrast, the large aromatic residues (tryptophan, tyrosine and phenylalanine) and Cβ-branched amino acids (isoleucine, valine, and threonine) prefer to adopt β-strand conformations. However, these preferences are not strong enough to produce a reliable method of predicting secondary structure from sequence alone. An amino acid change at the protein surface usually produces only local rearrangements and can be stabilized by the reorganization of the solvent molecules. Moreover, buried residues in a protein are less affected by the environment external to that protein, which might be very different among different evolutionary lineages (Pietro et al., 1998). So that, selective constraints may act to preserve protein structure (Jeffrey et al., 1996).

The ordination of the amino acids in terms of the most frequent substitutions agrees with the conservation of the β-helix, β-sheet, and β-turn formation tendencies during evolution. The same correspondence has been demonstrated for the conservation of the physicochemical properties in the amino acid substitutions (Angélica Soto et al., 1985). Secondary structural similarities and identities may also be impacted on the conservation of amino acids in a protein. A major fluctuation in β-helix, β-turn or random coil of protein secondary structure can effect on the regeneration and degeneration of some of the metabolic genes in the divergence time (Chothia and Lesk, 1986; Chothia and Gerstein, 1997; Gerstein, 1997). Only a stretch of amino acids in proteins is conferring its conformation in secondary structure to folding process. To understand better the effects of amino acid substitution in catalytic and regulatory regions of a protein, structural biologists study and attempt to predict secondary structure (Pietro et al., 1998). Therefore, the conservation in these aspects will obviously define the evolutionary significance of this domain (Apic et al., 2008; Ebenhöh et al., 2005). In this context, we described how predicted secondary structure of selected superproteins is played a crucial role in evolutionary relationships between Archaea and Bacteria. In addition, this work was aimed to study the physicochemical properties of these proteins and to compare with homologous proteins retrieved from both Bacteria and Archaea.

MATERIALS AND METHODS

The sequences of 25 archaeal superproteins were retrieved from NCBI database, which were used to search for pairwise similarity sequences from both Archaea and Bacteria by NCBI-BLASp algorithm (Altschul et al., 1999) using default parameters. Comprehensive information of protein sequences used in this study is presented in Table 1. The physiochemical features of these proteins include molecular weight, theoretical pl, estimated half-life, instability and aliphatic index, and grand average of hydropathicity (GRAVY) were computed by using ProtParam tool (Gasteiger et al., 2005). SOPMA (Geourjon and Deléage, 1995) tool was used to predict the secondary structure of proteins. A protein sequence was uploaded in a working space of this server and run with default parameters (window width 15, similarity threshold 1 and number of states 4) for prediction of β-helix, extended strand, β-turn, and random coil in %. The result was displayed in a graphical interface. Appropriate positions of those secondary structural elements were pointed where the suitable matches were found on sequences as uploaded.

RESULTS

Physiochemical characteristics of our query proteins are listed in Table 1. It showed that Ile-tRNA synthetase was the largest molecular weight protein where as coenzyme M and F420 were smallest proteins selected in this study. The molecular masses of other proteins were ranged from 214 to 638 KD. Quinol oxidase I, CoP methyltransferase, topoisomerase VI and sarcosine oxidase were alkaline proteins because of pl of them fallen to above 8.45 and rest of them belonged to acidic proteins (ranged 4.39 to 6.22). Quinol oxidase I, 4-vinyl reductase, DB synthetase and sarcosine oxidase have high aliphatic index and GRAVY values so that more hydrophobicity would be reflected for these proteins. Apart from these other proteins had moderate hydrophobic and hydrophilic ratios.

As shown in Table 2, a homolog NADH reductase of Methanosarcina acetivorans C2A has the largest molecular weight. One of the domains of this NADH reductase/oxidase family may be similar to our query protein. Allophonate hydrolase, sarcosine oxidase and tRNA PU synthetase and quinol oxidases were alkaline proteins while other proteins were acidic proteins except G-6-P synthetase.

Hydrophobicity and aliphatic index of many proteins ranged
between 70 and 114, but maximum was 121 to DB synthetase of Methanococcus maripaludis S2. Thus, physiochemical characteristics of many superproteins were closely resembled to same protein family or domains of Archaea.

Figure 1

Table 1: Physiochemical properties of our selected archaeal superproteins (query)

As reported in Table 3, we obtained many hits which similar to archaeal superproteins at primary structure levels. A homolog tRNA PU synthetase of Rubrobacter xylanophilus DSM 9941 has the largest molecular weight and remaining were ranged from 213 to 637 KD. RUBISCO, CoP methyltransferase, Ile-tRNA transferase, topoisomerase VIB and DB synthetase belonged to alkaline proteins (range 7.77-9.24) while other proteins acted as acidic in nature. Hydrophobicity and aliphatic index of few proteins predicted above 100, maximum was 114 to NADH reductase of Solibacter utitatus Elin 6076. Accordingly, we state that physiochemical features of archaeal proteins much more similar to proteins from archaeal domain than Bacteria.

Figure 2

Table 2: Physiochemical properties of the best archaeal homologs obtained for our selected archaeal proteins (query)

As reported in Table 3, we obtained many hits which similar to archaeal superproteins at primary structure levels. A homolog tRNA PU synthetase of Rubrobacter xylanophilus DSM 9941 has the largest molecular weight and remaining were ranged from 213 to 637 KD. RUBISCO, CoP methyltransferase, Ile-tRNA transferase, topoisomerase VIB and DB synthetase belonged to alkaline proteins (range 7.77-9.24) while other proteins acted as acidic in nature. Hydrophobicity and aliphatic index of few proteins predicted above 100, maximum was 114 to NADH reductase of Solibacter utitatus Elin 6076. Accordingly, we state that physiochemical features of archaeal proteins much more similar to proteins from archaeal domain than Bacteria.

Figure 3

Table 3: Physiochemical properties of the best bacterial homologs obtained for our selected archaeal proteins (query)

As reported in Table 3, we obtained many hits which similar to archaeal superproteins at primary structure levels. A homolog tRNA PU synthetase of Rubrobacter xylanophilus DSM 9941 has the largest molecular weight and remaining were ranged from 213 to 637 KD. RUBISCO, CoP methyltransferase, Ile-tRNA transferase, topoisomerase VIB and DB synthetase belonged to alkaline proteins (range 7.77-9.24) while other proteins acted as acidic in nature. Hydrophobicity and aliphatic index of few proteins predicted above 100, maximum was 114 to NADH reductase of Solibacter utitatus Elin 6076. Accordingly, we state that physiochemical features of archaeal proteins much more similar to proteins from archaeal domain than Bacteria.

Figure 4

CO dehydrogenase
Evolutionary implication of protein secondary structure among Archaea and Bacteria

Figure 5
MH cyclohydrolase

Figure 6
Coenzyme F

Figure 7
Coenzyme M

Figure 8
RUBISCO

Figure 9

Figure 10
G-6-P synthetase

Figure 11
Quinone reductase

Figure 12
NADH oxidase
Evolutionary implication of protein secondary structure among Archaea and Bacteria

Figure 13
Quinol oxidase I

Figure 14
CoP methyltransferase

Figure 15
4 - Vinyl reductase

Figure 16
tRNA PU synthase B

Figure 17
Ile - tRNA synthetase

Figure 18
Topoisomerase VI B

Figure 19
Ribonuclease H
Evolutionary implication of protein secondary structure among Archaea and Bacteria

Figure 20
DB synthetase

Figure 21
DHL dehydrogenase

Figure 22
HN synthase

Figure 23
BA dehydrogenase

Figure 24
Choline dehydrogenase

Figure 25
Allophonate hydrolase
Evolutionary implication of protein secondary structure among Archaea and Bacteria

After predicting secondary structure of these proteins, the structural elements have been used to compare with the homologous proteins of Archaea and Bacteria (Table 4 & 5). A large portion of the secondary structure was occupied by random coil, followed by β-helix. The variations in percentile were 5-10% in β-helix and 20-25% in random coil when compared within archaeal proteins. In contrast, there was a little variation among extended coils, but many proteins showed major differences in β-turns at secondary structure level.

Table 4: Comparative analysis of protein secondary structures of our archaeal superproteins and of the best archaeal homologs

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>α-Helix</th>
<th>Extended strand</th>
<th>β-Sheets</th>
<th>Random coil</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cytosolic dihydrofolate reductase</td>
<td>0.82</td>
<td>0.77</td>
<td>0.79</td>
<td>0.80</td>
</tr>
<tr>
<td>Other dihydrofolate reductase</td>
<td>0.83</td>
<td>0.77</td>
<td>0.80</td>
<td>0.81</td>
</tr>
<tr>
<td>Other dihydrofolate reductase</td>
<td>0.84</td>
<td>0.78</td>
<td>0.81</td>
<td>0.82</td>
</tr>
<tr>
<td>Other dihydrofolate reductase</td>
<td>0.85</td>
<td>0.79</td>
<td>0.81</td>
<td>0.82</td>
</tr>
<tr>
<td>Other dihydrofolate reductase</td>
<td>0.86</td>
<td>0.80</td>
<td>0.82</td>
<td>0.83</td>
</tr>
<tr>
<td>Other dihydrofolate reductase</td>
<td>0.87</td>
<td>0.81</td>
<td>0.83</td>
<td>0.84</td>
</tr>
<tr>
<td>Other dihydrofolate reductase</td>
<td>0.88</td>
<td>0.82</td>
<td>0.84</td>
<td>0.85</td>
</tr>
<tr>
<td>Other dihydrofolate reductase</td>
<td>0.89</td>
<td>0.83</td>
<td>0.85</td>
<td>0.86</td>
</tr>
<tr>
<td>Other dihydrofolate reductase</td>
<td>0.90</td>
<td>0.84</td>
<td>0.86</td>
<td>0.87</td>
</tr>
<tr>
<td>Other dihydrofolate reductase</td>
<td>0.91</td>
<td>0.85</td>
<td>0.87</td>
<td>0.88</td>
</tr>
<tr>
<td>Other dihydrofolate reductase</td>
<td>0.92</td>
<td>0.86</td>
<td>0.88</td>
<td>0.89</td>
</tr>
<tr>
<td>Other dihydrofolate reductase</td>
<td>0.93</td>
<td>0.87</td>
<td>0.89</td>
<td>0.90</td>
</tr>
<tr>
<td>Other dihydrofolate reductase</td>
<td>0.94</td>
<td>0.88</td>
<td>0.90</td>
<td>0.91</td>
</tr>
<tr>
<td>Other dihydrofolate reductase</td>
<td>0.95</td>
<td>0.89</td>
<td>0.91</td>
<td>0.92</td>
</tr>
<tr>
<td>Other dihydrofolate reductase</td>
<td>0.96</td>
<td>0.90</td>
<td>0.92</td>
<td>0.93</td>
</tr>
<tr>
<td>Other dihydrofolate reductase</td>
<td>0.97</td>
<td>0.91</td>
<td>0.93</td>
<td>0.94</td>
</tr>
<tr>
<td>Other dihydrofolate reductase</td>
<td>0.98</td>
<td>0.92</td>
<td>0.94</td>
<td>0.95</td>
</tr>
<tr>
<td>Other dihydrofolate reductase</td>
<td>0.99</td>
<td>0.93</td>
<td>0.95</td>
<td>0.96</td>
</tr>
<tr>
<td>Other dihydrofolate reductase</td>
<td>1.00</td>
<td>0.94</td>
<td>0.96</td>
<td>0.97</td>
</tr>
</tbody>
</table>

Unlike similarities found at primary levels, secondary structure implied more conserved in nature among archaeal proteins in this study.

Similarly, secondary structure of these proteins compared with homologous proteins of Bacteria as shown in Table 5. It revealed that tRNA PU synthetase, choline dehydrogenase and allophane hydrolase of Archaea were more differed from bacterial proteins in β-turn, but other proteins closely related with Bacteria. Excluding the proteins involved in carbon assimilation other proteins such as proteins involved in sulfur metabolism, bacterial photosynthesis and halo-adaptation were quantitatively resembled with bacterial proteins. The variations in percentile were 10-15% in β-helix and 20-25% in random coil when compared within archaeal proteins. In contrast, there was a major variation among extended coils, but many proteins showed major identities in β-helices at secondary structure level.

The secondary structure of these proteins was predicted by SOPMA server, and graphical representations of the structural information including helix, sheet, coil and turn are depicted in Figure 1.

Figure 26
Sarcosine oxidase α SB

Figure 27
CoHD reductase

Figure 28
Acetyl transferase

Figure 1 showing the predicted secondary structure of superproteins used in this work (Y axis represents amino acid length).
Evolutionary implication of protein secondary structure among Archaea and Bacteria

**DISCUSSION**

The archaeal proteins with more acidic amino acids involved in methanogenesis, osmoregulation, photorespiration and urea metabolism showed close similarity to Archaea than Bacteria. Similarly, NADH reductase of Sulfolobus acidocaldarius with more alkaline amino acids was homologs to Sulfolobus solfataricus, but not to prokaryotes. The isoelectric point of CoP methyltransferase, topoisomerase VIB, allophanate hydrolase, Ile-tRNA synthetase from Archaea were not matched each other and they were independent. Slesarev et al. (2002) found that M. kandleri proteins with high content of negatively charged amino acids are determined how this methanogenic bacterium is capable of adapting to the high intracellular salinity. Thus changes of amino acid resides are to determine the physiological functions of archaeal proteins.

The instability index of CO dehydrogenase, 4-vinyl reductase, CoHD reductase, allophanate reductase, NADH reductase, quinol oxidase, DHL dehydrogenase, ribonuclease H, topoidomerase VI was above 40. Therefore, these proteins are expected to be unstable structurally on subsequent divergence revealing a chance to protein family evolution. Aliphatic index of CoHD reductase showed close proximity between Bacteria and Archaea. There are many ways to categorize amino acids by chemical properties (e.g., hydrophobicity, charge, relative size of side chain), and physicochemical distances between amino acid types have been suggested (Grantham 1974; Taylor and Jones 1993), but these categorizations or physicochemical distances may not directly reflect the differences among amino acid types that are acted upon by evolution (Jeffrey et al., 1996). An amino acid is replaced very frequently by a physiochemically similar one. In the Dayhoff model, replacement rates were derived from alignments of protein sequences that are at least 85% identical. The assignments are used to build phylogenetic trees and the internal nodes of the tree give inferred ancestral sequences (Dayhoff et al., 1972). Thus, this work states that if any amino acid change occurred protein function will be changed from which structural constraints would be reordered.

Secondary structure elements represent regularities and basic building blocks of the architecture of a protein; thus, secondary structure is much more directly related to tertiary structure than the primary structure (Pietro et al., 1998). It is also important to note that secondary structural elements are more conserved than the precise atomic structure (Mizuguchi and Go, 1995) and that protein architecture depends on constraints related to bring key residues close in space. Similarly, a computational study of the protein sequences and structures of the superfamily of archaeo-eukaryotic primases has reported by Iyer et al. (2005). Thus the results of the secondary structure of some superprotein families are clustered to address questions related the distributions of β-helix, extended strand, β-turn and random coil among Archaea and Bacteria. Although about 95% extended coil similarities of superproteins found between Archaea and Bacteria, β-helix, β-turn and random coil configuration not found to show identity to all groups; it revealed the conformational variables have to be adjusted in order to stabilize the proteins structure during evolution.

The β-Helix of ammonia monoxygenase, β-turn of coenzyme M, CoP methyltransferase, and quinol oxidase I and random coil of DB synthetase, and Ile-tRNA synthetase showed more similarity to Archaea, but secondary structural conformation of CO dehydrogenase did not show similarity as such to Archaea except β-helix. Since ribonuclease H and topoisomerase VIB possessed unique conformations and structural orientations they were diverged independently. The proteins structurally are related to Archaea and not to Bacteria as secondary structure are conserved in a specific orientation (Galagan et al., 2002). Although protein secondary structures evolve far more slowly than protein sequences, they do evolve. The major changes are at the

---

**Figure 30**

Table 5: Comparative analysis of protein secondary structures of our archaeal superproteins and of the best bacterial homologs.
Evolutionary implication of protein secondary structure among Archaea and Bacteria

boundaries of $\beta$-helixes and $\beta$-sheets. This means that the amino acids at the boundaries of $\beta$-helixes and $\beta$-sheets may experience replacement rates that are quite different from those values experienced by residues in the middle of a structure element (Pietro et al., 1998). It is interesting that the $\beta$-helix rate estimate is greater than that for loops, but the biological significance is unclear. It may be the case that helices do evolve at greater rates than loops (Jeffrey et al., 1996).

Excluding extended coil, many of the conformations of NADH reductase and Ile-tRNA synthetase have been varied consecutively like bacterial proteins. G-6-P synthetase and AMO had shown their dissimilarities in $\beta$-helix and extended coil to archaea. Topoisomerase VI and allophosphate hydrolase in $\alpha$-turn and random coil, BA dehydrogenase and ribonuclease H in $\beta$-helix and random and 4-vinyl reductase in extended and $\alpha$-turn proved maximum similarities to Bacteria. There were no considerable conformational changes in secondary structure of HN synthase ($\beta$), sarcosine oxidase ($\beta$), acetyl transferase (thiolase) ($\beta$), quinol oxidase (heme-Cu oxidase) (extended) and choline dehydrogenase (random). Thus, natural selection acts on both the secondary structure elements, because of architectural constraints, and on a few critical residues directly involved in catalysis. Probably, these key residues are also those that are responsible for compensatory variations and longer range correlations in amino acid sequences (Pietro et al., 1998).

The genomes have different frequencies of supersecondary structures, with yeast having relatively more consecutive strands, Haemophilus having more consecutive helices, and Methanococcus having more alternating helix strand structures (Chothia and M. Gerstein, 1997; Gerstein, 1997; Russell et al., 1996). Like methyl CoM reductase in methanogens, topoisomerase VIB is also having a unique feature in secondary structure. Topoisomerase VI could interact in their respective domain with protein partners that are domain specific, preventing their correct function in a foreign domain (Apic et al., 2001; Gadelle et al., 2003). This results support the existence of majestic metabolic and protein diversity in Archaea, which are not closely correspondence to bacterial metabolism. In addition, one expects analysis of structure to reveal more about distant evolutionary relationships than just sequence comparison since structure is usually more conserved than sequence. The distinction between buried and exposed residues is particularly fruitful for molecular evolutionary studies because it allows consideration of protein regions undergoing different selection pressures. One of the most important advances in the reconstruction of evolutionary trees has been the consideration of heterogeneity of evolutionary rates among sequence sites (Yang, 1996). Nick et al. (1998) reported that rate heterogeneity is strongly associated with structural environment. Exposed sites tend to experience ~2X the rate of amino acid replacements experienced by buried sites. A higher rate of replacement for exposed sites is seen for each secondary structure type. The association between accessibility status and replacement rates is a noteworthy feature of protein evolution but has received scant attention in the field of molecular evolution (Yang, 1996; Nick et al., 1998).

Reconstruction of phylogenies from sequences with known structures involves less uncertainty and is therefore expected to be more accurate than reconstruction of phylogenies from sequences with unknown structures (Jeffrey et al., 1996). Although secondary structural elements of these proteins quantitatively resembled with bacterial homologs, they are not qualitatively corresponded at specific sequence position. However, having the key residues at functional positions and structural constraints are still maintaining the functional activity and structural conservation of these proteins among Archaea and Bacteria. Thus, the results obtained in this work point to the strong necessities of better understanding of the microbial diversity, particularly archaeal domain and of its evolutionary relationships. This present attempt will obviously provide a new view to the evolutionary biologists to strengthen the conceptual idea about diversity of protein superfamilies among Archaea and Bacteria with response to quantitative secondary structure conformations during archaeal evolution.

ACKNOWLEDGEMENT

The corresponding author is grateful to the University Grants Commission, New Delhi, India for financial assistance (UGC Sanction No. 32-559/2006) to carry out the work.

References


Evolutionary implication of protein secondary structure among Archaea and Bacteria

61(4): 456–502. PMID: 9409149
PMID: 9417935
Author Information

P. Chellapandi
Department of Bioinformatics, School of Life Sciences, Bharathidasan University

C. Karthigeyen
Department of Biotechnology, School of Life Sciences, Bharathidasan University

S. Sivaramakrishnan
Department of Biotechnology, School of Life Sciences, Bharathidasan University