

# Differences in Amino Acid Composition between $\alpha$ and $\beta$ Structural Classes of Proteins

Hiroshi Nakashima\*, Yuka Saitou, Nachi Usuki

Department of Clinical Laboratory Science, Graduate Course of Medical Science and Technology,  
School of Health Sciences, Kanazawa University, Kanazawa, Japan  
Email: \*[naka@kenroku.kanazawa-u.ac.jp](mailto:naka@kenroku.kanazawa-u.ac.jp)

Received 6 July 2014; revised 21 August 2014; accepted 7 September 2014

Copyright © 2014 by authors and Scientific Research Publishing Inc.  
This work is licensed under the Creative Commons Attribution International License (CC BY).  
<http://creativecommons.org/licenses/by/4.0/>



Open Access

---

## Abstract

The amino acid composition of  $\alpha$  and  $\beta$  structural class of proteins from five species, *Escherichia coli*, *Thermotoga maritima*, *Thermus thermophilus*, yeast, and humans were investigated. Amino acid residues of proteins were classified into interior or surface residues based on the relative accessible surface area. The hydrophobic Leu, Ala, Val, and Ile residues were rich in interior residues, and hydrophilic Glu, Lys, Asp, and Arg were rich in surface residues both in  $\alpha$  and  $\beta$  proteins. The amino acid composition of  $\alpha$  proteins was different from that of  $\beta$  proteins in five species, and the difference was derived from the different contents of their interior residues between  $\alpha$  and  $\beta$  proteins.  $\alpha$ -helix content of  $\alpha$  proteins was rich in interior residues than surface ones. Similarly,  $\beta$ -sheet content of  $\beta$  proteins was rich in interior residues than surface ones. The content of Leu residues was very high, approximately 20%, in interior residues of  $\alpha$  proteins. This result suggested that the Leu residue plays an important role in the folding of  $\alpha$  proteins.

## Keywords

Amino Acid Composition,  $\alpha$  and  $\beta$  Proteins, Interior/Surface Residue, Protein Domains, Species-Specific Amino Acid Composition

---

## 1. Introduction

Nearly 30 years ago, Nakashima *et al.* [1] reported that the amino acid composition of proteins is different among the four structural classes:  $\alpha$ ,  $\beta$ ,  $\alpha/\beta$  and  $\alpha + \beta$ . Many studies [2]-[8] have confirmed that there is a corre-

---

\*Corresponding author.

lation between the amino acid composition and structural class of proteins. However, the reason behind the differences in amino acid composition among different structural classes is not clearly understood.

It is known that bacteria have species-specific nucleotide compositions in their protein-coding genes [9]–[11]. Due to the biased nucleotide composition, the amino acid composition of bacteria is also biased. The accumulation of data of the three-dimensional (3D) structure of proteins has provided the opportunity to statistically analyze the amino acid composition of proteins of different structural classes from different species. Here, we compared the amino acid composition of proteins from five species; *Escherichia coli*, *Thermotoga maritima*, *Thermus thermophilus*, yeast (*Saccharomyces cerevisiae*), and humans (*Homo sapiens*), because these species have relatively large numbers of structural proteins. The species-specific nucleotide composition is largely dependent on G + C content. The G + C content in the whole genome is 50.8% in *E. coli*, 46.2% in *T. maritima*, 69.5% in *T. thermophilus*, and 38.3% in yeast. These four species have a constant G + C content in their genomes. By contrast, the human genome has a considerably variable G + C content according to their chromosomal locations [12]. The optimal growth temperatures of *T. maritima* and *T. thermophilus* are 80°C and 85°C, respectively, as they are hyperthermophiles. It is reported that proteins of thermophiles have a different amino acid composition compared to that of mesophiles [13]–[17].

The amino acid composition of  $\alpha$  proteins is mostly distinct from that of  $\beta$  proteins among the four structural classes [1]. Therefore, comparison of the amino acid composition of  $\alpha$  and  $\beta$  proteins appears to be ideal approach to understand the differences in amino acid composition between any two structural classes of proteins. As there is a correlation between amino acid composition and structural classes, we hypothesized that there might be some basic features that facilitate the organization of  $\alpha$  or  $\beta$  proteins and that such features might be conserved. The purpose of this study was to identify such basic features from various sequences. We compared the amino acid composition of  $\alpha$  or  $\beta$  proteins from five species, and analyzed their interior or surface residues to identify the common features as well as the differences.

## 2. Materials and Methods

### 2.1. Amino Acid Sequences

The amino acid sequences of known 3D structures were obtained from the structural classification of proteins (SCOP) [18] sequence database 1.75 released on the web site <http://scop.mrc-lmb.cam.ac.uk/scop/>. Each sequence has a Protein Data Bank (PDB) [19] entry code and a SCOP structural classification code, which represents the structural class, fold, superfamily, and family. The amino acid sequences in SCOP are divided into protein domains and the structural class is depicted for individual sequences. The amino acid sequences that constitute protein domains are not always sequential, as some sequences are composed of two fragments from separate regions. Sequences longer than 100 amino acid residues were selected for this analysis. The sequences of *E. coli*, *T. maritima*, *T. thermophilus*, yeast, and humans were selected, and then the sequences of  $\alpha$  or  $\beta$  structural class proteins were further selected. The selected sequences were analyzed for their sequence similarity using the BLAST program [20]. The proteins that had more than 25% sequence identity over 100 residues were excluded to avoid any bias. Sequences with identical SCOP structural codes were included when the sequences had less than 25% sequence identity over 100 residues. The collected data included 77 sequences of *E. coli*, 22 of *T. maritima*, 23 of *T. thermophilus*, 45 of yeast, and 154 of humans for  $\alpha$  protein domains, and 90 sequences of *E. coli*, 17 of *T. maritima*, 28 of *T. thermophilus*, 35 of yeast, and 187 of humans for  $\beta$  protein domains. The number of total protein domains was 321 for  $\alpha$  and 357 for  $\beta$  proteins. The protein names, number of residues, and the SCOP and PDB codes of the sequences used in this study are listed in the supplementary data.

### 2.2. Classification of Interior/Surface Residues

The amino acid residues in the analyzed sequences were classified into two types, interior or surface residues, based on their relative solvent accessibility. Solvent accessibility and secondary structure calculated using the DSSP program [21] with the coordinate data of PDB were obtained from the European Bioinformatics Institute web site (<http://www.ebi.ac.uk/>). Amino acid residues with relative solvent accessibility greater than 25% were regarded as surface residues, and those with relative solvent accessibility less than 25% were considered as interior residues as described by Fukuchi and Nishikawa [14]. Secondary structures of a residue were classified into three states:  $\alpha$ -helix,  $\beta$ -sheet, and coil.  $\alpha$ -helices and  $\beta$ -sheets were categorized according to the definition of the

DSSP program and residues other than an  $\alpha$ -helix or  $\beta$ -sheet were considered as coils. The intrinsically disordered residues were not included in the DSSP sequences. The SCOP amino acid sequences were aligned with DSSP sequences, and the solvent accessibility and secondary structure were given for the corresponding residues between the two sequences.

### 2.3. Comparison of Amino Acid Composition

The average amino acid composition was compared between  $\alpha$  and  $\beta$  proteins, and interior and surface residues, among the five species. To analyze the amino acid composition of a protein, amino acid composition space, introduced by Nishikawa and Ooi [22]–[24], was employed. The number of each of the 20 types of amino acid residues in a protein sequence was counted and the composition was expressed in a normalized scale using the equation:

$$V_{i,k} = (C_{i,k} - AV_k) / SD_k$$

where  $V_{i,k}$  and  $C_{i,k}$  are the normalized and real composition of amino acid residues of the  $k$ -th component in a sequence  $i$ , respectively.  $AV_k$  and  $SD_k$  are the average composition and the standard deviation of the  $k$ -th component for the whole dataset, respectively.

The amino acid sequence of a protein was converted to an amino acid composition vector of 20 components and it was plotted as a point in a 20-dimensional composition space. The distribution of proteins was visualized by projecting them onto a two-dimensional (2D) plane, which was defined by two axes of the principal component analysis. A principal component analysis was conducted for the total proteins. The x-coordinate for a given sequence was calculated using the scalar product of the unit vector of the first principal component and the vector of the sequence. The y-coordinate was calculated using the scalar product of the unit vector of the second principal component and the vector of the sequence. The origin of the x-y coordinate system in the composition space was set at the average amino acid composition of all analyzed sequences.

## 3. Results

### 3.1. Amino Acid Composition

The average and standard deviation values of the amino acid composition of total proteins analyzed in this study are indicated in **Table 1**. The average amino acid compositions of interior, surface, and whole residues of  $\alpha$  and  $\beta$  structural classes of proteins are listed in **Table 1**. The number of domains used in the calculation is shown in the last row in **Table 1**. The hydrophobic Leu, Ala, Ile, and Val residues were rich in interior residues, and hydrophilic Glu, Lys, Asp, and Arg were rich in surface residues, both in  $\alpha$  and  $\beta$  proteins. This trend was observed commonly in all five species. The ratios of the interior residues against surface residues clearly indicated that Cys, Phe, Ile, Trp, and Leu were favored as interior residues, and Lys, Glu, Asp, Arg, and Gln were favored as surface residues both in  $\alpha$  and  $\beta$  proteins. This trend was also observed commonly in all five species. This result indicated that the location of a residue (interior or surface) in a protein is dependent on the character of the amino acid and independent of the structural class. The Leu residues content was very high, approximately 20%, in interior residues of  $\alpha$  proteins. To show the difference between  $\alpha$  and  $\beta$  proteins, the ratios of whole residues of  $\alpha$  proteins to those of  $\beta$  proteins were calculated. The ratios indicated that Leu, Met, Ala, and Glu residues were predominant in the  $\alpha$  proteins of the five species, while Gly, Pro, Val, and Thr residues were predominant in the  $\beta$  proteins of the five species. The favored amino acids in the two structural classes of proteins were consistent with a previous study [1].

### 3.2. Distribution of $\alpha$ and $\beta$ Proteins on a 2D Plane

The proteins of  $\alpha$  and  $\beta$  structural classes of *E. coli* were plotted on a 2D plane by using the first principal component as the x-axis and the second component as the y-axis (**Figure 1**). The distribution of  $\alpha$  proteins of *E. coli* was roughly separated from that of  $\beta$  proteins of *E. coli*. A similar plot indicated that  $\alpha$  proteins were roughly separated from that of  $\beta$  proteins in the other four species, similar to that observed in *E. coli*. However, the plot of total  $\alpha$  and total  $\beta$  proteins from the five species together indicated an overlapped distribution.

The variance of the first principal component was 12.6% and that of the second component was 12.1%. Asn, Ile, Phe, and Ser residues were largely shifted toward the positive direction, while Ala, Arg, Leu, and Glu resi-

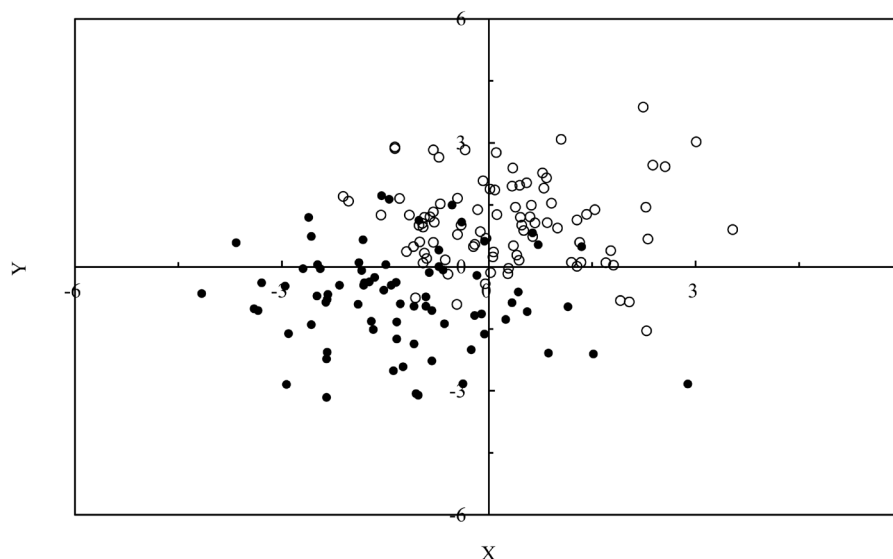
**Table 1.** Average and standard deviation (SD) of amino acid composition (%) of total proteins. Average of interior, surface, and whole residues in  $\alpha$  and  $\beta$  proteins.

Amino acid	total proteins		$\alpha$ proteins			$\beta$ proteins		
	average	SD	interior	surface	whole	interior	surface	whole
Ala	7.21	3.15	10.72	5.87	8.14	8.22	4.60	6.36
Cys	1.31	1.22	1.98	0.43	1.18	2.44	0.52	1.43
Asp	5.45	2.06	2.77	8.19	5.56	2.83	7.81	5.35
Glu	7.62	2.83	3.59	13.17	8.60	2.75	10.68	6.74
Phe	4.05	1.89	6.53	1.44	3.88	6.77	1.70	4.20
Gly	6.08	2.97	3.57	5.43	4.56	6.83	7.93	7.44
His	2.31	1.33	1.86	2.50	2.23	2.30	2.41	2.38
Ile	5.40	2.36	9.28	2.08	5.45	8.69	2.32	5.36
Lys	6.30	2.94	1.99	11.06	6.68	1.83	9.96	5.96
Leu	10.16	3.32	19.63	4.86	11.83	13.47	4.28	8.66
Met	2.17	1.40	3.44	1.62	2.49	2.24	1.44	1.88
Asn	4.25	1.99	2.48	5.75	4.20	2.54	5.95	4.29
Pro	4.73	2.28	2.57	5.08	3.87	4.08	6.80	5.50
Gln	4.45	2.24	2.59	7.15	4.98	2.22	5.62	3.98
Arg	5.54	2.55	3.08	8.47	5.92	2.79	7.42	5.20
Ser	6.11	2.70	4.18	6.87	5.67	5.18	7.50	6.50
Thr	5.28	2.28	4.16	4.90	4.59	5.26	6.43	5.91
Val	7.03	2.64	9.12	2.73	5.78	12.70	4.04	8.16
Trp	1.33	1.06	2.02	0.48	1.22	2.30	0.58	1.44
Tyr	3.22	1.60	4.44	1.92	3.17	4.56	2.01	3.26
Domains	678				321			357

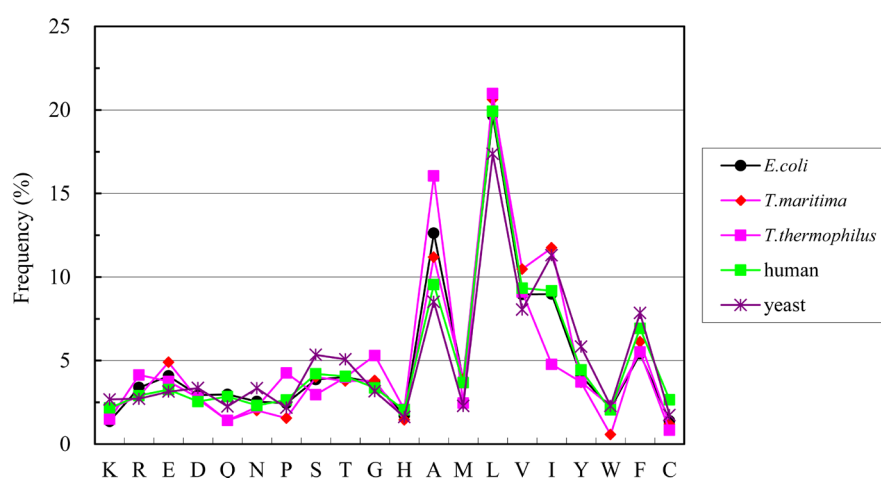
dues were largely shifted toward the negative direction along the x-axis. Asn, Ile, and Phe residues have A + T-rich codons, and Ala and Arg residues have G + C-rich codons. This result suggested that the x-axis reflects the character of the G + C content of the residues. Along the y-axis, Gly, Pro, Val, and Thr residues showed a larger positive coefficient, and Lys, Glu, Ile, and Met residues showed a larger negative coefficient. The residues with positive coefficients were the favored residues in  $\beta$  proteins, and those with negative coefficients were the residues that were mostly detected in  $\alpha$  proteins. These findings indicated that the y-axis reflects the frequency of the occurrence of amino acid residues between  $\alpha$  and  $\beta$  proteins. This is consistent with the distribution of *E. coli* proteins in **Figure 1**, where most of the  $\alpha$  proteins showed negative values of y-coordinates, and most of the  $\beta$  proteins had positive y-coordinates.

### 3.3. Differences in Interior and Surface Compositions among the Five Species

The average amino acid compositions of interior and surface residues in  $\alpha$  proteins from the five species were plotted, and the results are shown in **Figure 2** and **Figure 3**, respectively. To clearly demonstrate the differences in the amino acid content among the five species, the differences between the maximum/minimum and average composition were plotted. **Figure 4** shows the plot indicating the differences in the interior composition of  $\alpha$  proteins. Ala and Ile residues showed large differences. The Ala residue content in interior composition was highest (16.04%) in *T. thermophilus* and lowest (8.51%) in yeast in  $\alpha$  proteins (**Figure 2**). Since the average Ala



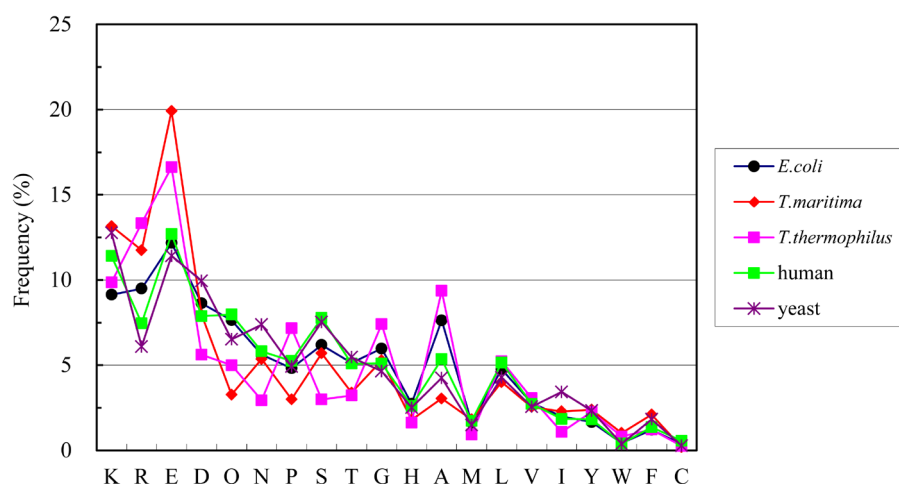
**Figure 1.** Distribution of  $\alpha$  (filled circles) and  $\beta$  proteins (open circles) of *E. coli*. The x- and y-axes represent the first and second axes determined by principal component analysis.



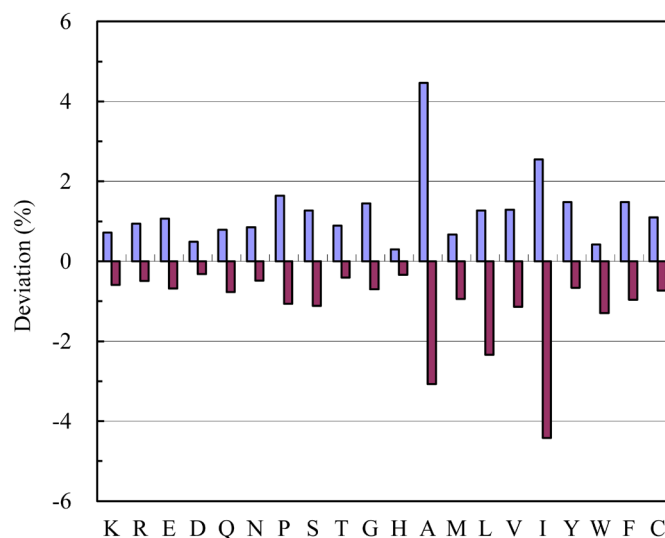
**Figure 2.** The average amino acid composition of interior residues of  $\alpha$  proteins of the five species.

composition in the five species was 11.58%, the deviation was +4.46% and -3.07% (**Figure 4**). The Ile residue content was lowest (4.77%) in *T. thermophilus* and highest (11.30%) in yeast. Since the average Ile composition in the five species was 9.19%, the deviation was +2.55% and -4.42% (**Figure 4**). Taken together, the total of Ala and Ile content was 20.81% in *T. thermophilus* and 19.81% in yeast. This result indicated that the increase or decrease of Ala content is compensated by the decrease or increase of Ile, and their contents were adjusted in the species with different G + C content. The variation in Ala and Ile contents was consistent with previous reports [25]-[29], which indicated that amino acids composed of G + C-rich or A + T-rich codons are related to the genomic G + C content of the species.

The differences in the surface composition of  $\alpha$  proteins are shown in **Figure 5**. The Glu residue showed a large difference. The average Glu residue content in the five species was 14.57%. The maximum composition was 19.92% in *T. maritima* and the minimum composition was 11.42% in yeast. Therefore, the difference in the distribution of Glu was +5.35% and -3.15% (**Figure 5**). The Glu residue has GAA and GAG codons, which are neutral with respect to G + C content. Therefore, the large differences in the Glu residue content are not explained by the G + C content. It has been reported that thermophiles have a higher content of charged residues in



**Figure 3.** The average amino acid composition of surface residues of  $\alpha$  proteins of the five species.

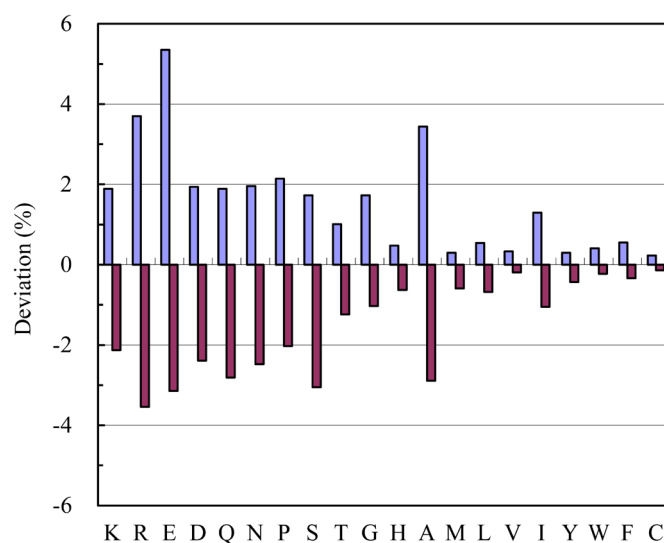


**Figure 4.** The differences in amino acid content between the maximum and average composition (in blue) and that between the minimum and average composition (in red) of interior residues of  $\alpha$  proteins among the five species.

surface composition than mesophiles [14]–[17]. The total content of charged residues of Asp, Glu, Lys, and Arg in the surface composition of  $\alpha$  proteins was 52.79% in *T. maritima*, 45.42% in *T. thermophilus*, 39.44% in *E. coli*, 39.45% in humans, and 40.21% in yeast. The total content of charged residues was higher in thermophiles than in mesophiles, and it was higher in  $\alpha$  proteins than in  $\beta$  proteins. There were more differences in the composition of surface residues in  $\alpha$  proteins than interior residues (Figure 4 and Figure 5). Similar results were observed in  $\beta$  proteins. It is empirically known that the interior residues are more conserved than the surface residues in homologous proteins. Therefore, it is reasonable that the interior composition should show less difference than the surface composition.

#### 4. Discussion

The amino acid residues were classified into interior or surface residues based on their relative solvent accessibility. The amino acid composition of interior and surface residues was dependent on the classification. In this



**Figure 5.** The differences in amino acid content between the maximum and average composition (in blue) and that between the minimum and average composition (in red) of surface residues of  $\alpha$  proteins among the five species.

study, an average of surface residues of  $\alpha$  proteins of the five species was 51%, and the average of interior residues was 49%. In  $\beta$  proteins, the average compositions of both surface and interior residues of the five species were 50%. Since we intended to classify the residues into interior and surface residues in equal proportions, the obtained result met our intended criterion.

The average content of secondary structures in  $\alpha$  and  $\beta$  proteins is shown in Table 2. The interior residues in  $\alpha$  proteins had a higher  $\alpha$ -helix content, 73%, than the surface residues. Similarly, the interior residues in  $\beta$  proteins had a higher  $\beta$ -sheet content, 54%, than the surface residues.

The hydrophobic residues, such as Leu, Ala, Ile, and Val were enriched in interior residues of both  $\alpha$  and  $\beta$  proteins. However, the percentages of  $\alpha$ -helices and  $\beta$ -sheets in interior residues were quite different. The reason behind such a big difference is not clear. Since the Ala residue has G + C-rich codons and the Ile residue has A + T-rich codons, their contents were dependent on the G + C content of the species. Therefore, the differences in Ala and Ile compositions were large compared to those of Val and Leu among the five species (Figure 4). The Val residue has GTN (N stands for all four nucleotides) codons, which are neutral with respect to G + C content. This might be the reason for the smaller differences observed in Val content. The Leu residue has neutral CTN codons and A + T-rich TTA and TTG codons. The Leu residue showed smaller deviations among species even though its content was consistently high.

In *E. coli*, the distribution of whole residues of  $\alpha$  proteins was roughly distinct from its  $\beta$  proteins. Similarly, the interior residues between  $\alpha$  and  $\beta$  proteins showed a distinct distribution. However, the plot of surface residues in *E. coli* was overlapped. Similar results were also observed in other species. This indicates that the difference in amino acid composition of whole residues between  $\alpha$  and  $\beta$  proteins was derived from the differences in the interior residues. This result was obtained by the analysis of the distribution of  $\alpha$  and  $\beta$  proteins in the amino acid composition space. However, no clear differences were noticed in the initial observations of amino acid composition of interior residues between  $\alpha$  and  $\beta$  proteins. The interior residues were rich in hydrophobic Leu, Ala, Ile, and Val residues both in  $\alpha$  and  $\beta$  proteins. To clearly demonstrate the differences in interior residues between  $\alpha$  and  $\beta$  proteins, the ratios of their amino acid composition were calculated. This indicated that Met, Leu, Glu, and Ala residues were favored in  $\alpha$  proteins and Gly, Pro, Val, and Thr were favored in  $\beta$  proteins. This trend was observed commonly in all five species. Therefore, we concluded that the existence of favorable residues is a basic common feature of  $\alpha$  and  $\beta$  proteins. The residues favored in  $\alpha$  proteins were consistent with the residues favored in  $\alpha$ -helix [30], however, Gly and Pro residues favored in  $\beta$  proteins were reported as unfavorable in  $\beta$ -sheet [30]. The Gly and Pro residues contents were not large compared to those of the hydrophobic residues, however, they were commonly favored in  $\beta$  proteins in the five species. This suggested that Gly and

**Table 2.** Average content of secondary structures (%).

	$\alpha$ proteins			$\beta$ proteins		
	whole	interior	surface	whole	interior	surface
$\alpha$ -helix	63	73	54	9	8	9
$\beta$ -sheet	2	2	2	39	54	25
coil	35	25	44	52	38	66

Pro residues are essential interior residues in the organization of  $\beta$  proteins outside of the  $\beta$ -sheet regions. The average Leu residue content was 17% in yeast, and was 20% in the other four species. This outcome suggested that the interaction between Leu residues is likely to be very large in the interior region and that it might play an important role in the folding of  $\alpha$  proteins.

## References

- [1] Nakashima, H., Nishikawa, K. and Ooi, T. (1986) Folding Type of a Protein Is Relevant to the Amino Acid Composition. *The Journal of Biochemistry*, **99**, 153-162.
- [2] Klein, P. (1986) Prediction of Protein Structural Class by Discriminant Analysis. *Biochimica et Biophysica Acta, Protein Structure and Molecular Enzymology*, **874**, 205-215. [http://dx.doi.org/10.1016/0167-4838\(86\)90119-6](http://dx.doi.org/10.1016/0167-4838(86)90119-6)
- [3] Dubchak, I., Muchnik, I., Holbrook, S.R. and Kim, S.H. (1995) Prediction of Protein Folding Class Using Global Description of Amino Acid Sequence. *Proceedings of the National Academy of Sciences of the United States of America*, **92**, 8700-8704. <http://dx.doi.org/10.1073/pnas.92.19.8700>
- [4] Bahar, I., Atilgan, A.R., Jernigan, R.L. and Erman, B. (1997) Understanding the Recognition of Protein Structural Classes by Amino Acid Composition. *Proteins*, **29**, 172-185.
- [5] Chou, K.C. (1999) A Key Driving Force in Determination of Protein Structural Classes. *Biochemical and Biophysical Research Communications*, **264**, 216-224. <http://dx.doi.org/10.1006/bbrc.1999.1325>
- [6] Ofra, Y. and Margalit, H. (2006) Proteins of the Same Fold and Unrelated Sequences Have Similar Amino Acid Composition. *Proteins: Structure, Function, and Bioinformatics*, **64**, 275-279. <http://dx.doi.org/10.1002/prot.20964>
- [7] Taguchi, Y. and Gromiha, M.M. (2007) Application of Amino Acid Occurrence for Discriminating Different Folding Types of Globular Proteins. *BMC Bioinformatics*, **8**, 404. <http://dx.doi.org/10.1186/1471-2105-8-404>
- [8] Rackovsky, S. (2009) Sequence Physical Properties Encode the Global Organization of Protein Structure Space. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 14345-14348. <http://dx.doi.org/10.1073/pnas.0903433106>
- [9] Karlin, S. and Burge, C. (1995) Dinucleotide Relative Abundance Extremes: A Genomic Signature. *Trends in Genetics*, **11**, 283-290. [http://dx.doi.org/10.1016/S0168-9525\(00\)89076-9](http://dx.doi.org/10.1016/S0168-9525(00)89076-9)
- [10] Karlin, S., Mrázek, J. and Campbell, A.M. (1997) Compositional Biases of Bacterial Genomes and Evolutionary Implications. *Journal of Bacteriology*, **179**, 3899-3913.
- [11] Nakashima, H., Ota, M., Nishikawa, K. and Ooi, T. (1998) Gene from Nine Genomes Are Separated into Their Organisms in the Dinucleotide Composition Space. *DNA Research*, **5**, 251-259. <http://dx.doi.org/10.1093/dnares/5.5.251>
- [12] Bernardi, G., Olofsson, B., Filipinski, J., Zerial, M., Salinas, J., Cuny, G., Meunier-Rotival, M. and Rodier, F. (1985) The Mosaic Genome of Warm-Blooded Vertebrates. *Science*, **228**, 953-958. <http://dx.doi.org/10.1126/science.4001930>
- [13] Kreil, D.P. and Ouzounis, C.A. (2001) Identification of Thermophilic Species by the Amino Acid Compositions Deduced from Their Genomes. *Nucleic Acids Research*, **29**, 1608-1615. <http://dx.doi.org/10.1093/nar/29.7.1608>
- [14] Fukuchi, S. and Nishikawa, K. (2001) Protein Surface Amino Acid Compositions Distinctively Differ between Thermophilic and Mesophilic Bacteria. *Journal of Molecular Biology*, **309**, 835-843. <http://dx.doi.org/10.1006/jmbi.2001.4718>
- [15] Chakravarty, S. and Varadarajan, R. (2002) Elucidation of Factors Responsible for Enhanced Thermal Stability of Proteins: A Structural Genomics Based Study. *Biochemistry*, **41**, 8152-8161. <http://dx.doi.org/10.1021/bi025523t>
- [16] Nakashima, H., Fukuchi, S. and Nishikawa, K. (2003) Compositional Changes in RNA, DNA and Proteins for Bacterial Adaptation to Higher and Lower Temperatures. *The Journal of Biochemistry*, **133**, 507-513. <http://dx.doi.org/10.1093/jb/mvg067>
- [17] Yokota, K., Satou, K. and Ohki, S. (2006) Comparative Analysis of Protein Thermostability: Differences in Amino

- Acid Content and Substitution at the Surfaces and in the Core Regions of Thermophilic and Mesophilic Proteins. *Science and Technology of Advanced Materials*, **7**, 255-262.
- [18] Murzin, A.G., Brenner, S.E., Hubbard, T. and Chothia, C. (1995) SCOP: A Structural Classification of Proteins Database for the Investigation of Sequences and Structures. *Journal of Molecular Biology*, **247**, 536-540. [http://dx.doi.org/10.1016/S0022-2836\(05\)80134-2](http://dx.doi.org/10.1016/S0022-2836(05)80134-2)
  - [19] Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N. and Bourne, P.E. (2000) The Protein Data Bank. *Nucleic Acids Research*, **28**, 235-242. <http://dx.doi.org/10.1093/nar/28.1.235>
  - [20] Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) Basic Local Alignment Search Tool. *Journal of Molecular Biology*, **215**, 403-410. [http://dx.doi.org/10.1016/S0022-2836\(05\)80360-2](http://dx.doi.org/10.1016/S0022-2836(05)80360-2)
  - [21] Kabsch, W. and Sander, C. (1983) Dictionary of Protein Secondary Structure: Pattern Recognition of Hydrogen-Bonded and Geometrical Features. *Biopolymers*, **22**, 2577-2637. <http://dx.doi.org/10.1002/bip.360221211>
  - [22] Nishikawa, K. and Ooi, T. (1982) Correlation of the Amino Acid Composition of a Protein to Its Structural and Biological Characters. *The Journal of Biochemistry*, **91**, 1821-1824.
  - [23] Nishikawa, K., Kubota, Y. and Ooi, T. (1983) Classification of Proteins into Groups Based on Amino Acid Composition and Other Characters. I. Angular Distribution. *The Journal of Biochemistry*, **94**, 981-995.
  - [24] Nishikawa, K., Kubota, Y. and Ooi, T. (1983) Classification of Proteins into Groups Based on Amino Acid Composition and Other Characters. II. Grouping into Four Types. *The Journal of Biochemistry*, **94**, 997-1007.
  - [25] Lobry, J.R. (1997) Influence of Genomic G+C Content on Average Amino-Acid Composition of Proteins from 59 Bacterial Species. *Gene*, **205**, 309-316. [http://dx.doi.org/10.1016/S0378-1119\(97\)00403-4](http://dx.doi.org/10.1016/S0378-1119(97)00403-4)
  - [26] Singer, G.A.C. and Hickey, D.A. (2000) Nucleotide Bias Causes a Genomewide Bias in the Amino Acid Composition of Proteins. *Molecular Biology and Evolution*, **17**, 1581-1588. <http://dx.doi.org/10.1093/oxfordjournals.molbev.a026257>
  - [27] Bharanidharan, D., Bhargavi, G.R., Uthamallian, K. and Gautham, N. (2004) Correlations between Nucleotide Frequencies and Amino Acid Composition in 115 Bacterial Species. *Biochemical and Biophysical Research Communications*, **315**, 1097-1103. <http://dx.doi.org/10.1016/j.bbrc.2004.01.129>
  - [28] Hu, J., Zhao, X., Zhang, Z. and Yu, J. (2007) Compositional Dynamics of Guanine and Cytosine Content in Prokaryotic Genomes. *Research in Microbiology*, **158**, 363-370. <http://dx.doi.org/10.1016/j.resmic.2007.02.007>
  - [29] Nakashima, H., Yoshihara, A. and Kitamura, K. (2013) Favorable and Unfavorable Amino Acid Residues in Water-Soluble and Transmembrane Proteins. *Journal of Biomedical Science and Engineering*, **6**, 36-44. <http://dx.doi.org/10.4236/jbise.2013.61006>
  - [30] Chou, P.Y. and Fasman, G.D. (1978) Empirical Predictions of Protein Conformation. *Annual Review of Biochemistry*, **47**, 251-276. <http://dx.doi.org/10.1146/annurev.bi.47.070178.001343>

## Supplementary Data

<i>E. coli</i>	alpha proteins			
No.	SCOP	PDB	residues	protein name
1	A.1.1.2	1GVHA1	146	Flavohemoglobin, N-terminal domain
2	A.1.2.1	1KF6B1	138	Fumarate reductase flavoprotein
3	A.2.14.1	1TJLA1	104	DnaK suppressor protein DksA, alpha-hairpin domain
4	A.4.5.8	1B9NA1	122	Transcriptional regulator ModE, N-terminal domain
5	A.4.5.10	2Z9OA1	114	(A: 21 - 143) RepE54
6	A.4.5.10	1REPC2	103	(C: 144 - 246) RepE54
7	A.4.5.10	2NRAC1	135	(C: 9 - 151) Replication initiation protein PI
8	A.4.5.10	2NRAC2	117	(C: 152 - 268) Replication initiation protein PI
9	A.4.5.27	1T0FA1	100	(A: 169 - 268) TnsA endonuclease, C-terminal domain
10	A.4.5.28	1JGSA	138	Multiple antibiotic resistance repressor
11	A.4.5.43	1OYWA1	110	(A: 407 - 516) DNA helicase RecQ DNA-binding domain
12	A.4.5.65	1T98A1	103	(A: 8 - 118) Chromosome partition protein MukF
13	A.4.6.1	1GXQA	105	Transcription activator PhoB
14	A.4.12.1	1CO0A	105	Trp repressor, TrpR
15	A.4.14.1	1R71B	114	Transcriptional repressor protein KorB
16	A.6.1.3	1Q05B	122	Transcriptional regulator CueR
17	A.7.3.1	1CHUA1	111	(A: 423 - 533) L-aspartate oxidase
18	A.7.3.1	1NEKA1	138	(A: 451 - 588) Succinate dehydrogenase
19	A.7.15.1	1XDOA1	105	Polyphosphate kinase, PPK
20	A.8.4.1	1DKXA1	101	(A: 507 - 607) Chaperone DnaK
21	A.8.4.1	1U00A1	112	(A: 504 - 615) Chaperone protein hscA (Hsc66)
22	A.8.7.1	1R8IA	187	Secretion system protein TraC
23	A.24.2.1	2ASRA	142	Aspartate receptor, ligand-binding domain
24	A.24.3.1	1APCA	106	Cytochrome b562
25	A.24.10.1	2A0BA	118	Aerobic respiration control sensor protein
26	A.24.10.4	1SR2A	116	Sensor-like histidine kinase YojN, C-terminal domain
27	A.25.1.1	1BFRA	158	Bacterioferritin (Cytochrome b1)
28	A.25.1.1	1DPSA	159	Dodecameric ferritin homolog
29	A.25.1.2	1OTKA	244	Phenylacetic acid degradation protein PaaC
30	A.25.1.2	1MXRA	339	Ribonucleotide reductase R2
31	A.24.16.4	1V4AA1	151	(A: 287 - 437) Glutamine synthase adenylyltransferase
32	A.25.1.4	2GS4A1	158	Hypothetical protein YciF
33	A.27.1.1	1U0BB1	146	(B: 316 - 461) CysteinyI-tRNA synthetase (CysRS)
34	A.29.15.1	2HI7B1	134	Disulfide bond formation protein

## Continued

35	A.29.17.1	2HGKA1	109	Hypothetical protein YqcC
36	A.43.1.5	1CMCA	104	Met repressor, MetJ
37	A.45.1.1	1N2AA1	121	(A: 81 - 201) class beta GST
38	A.45.1.1	1B8XA1	180	(A: 81 - 260) class beta GST
39	A.45.1.1	1G7OA1	140	(A: 76 - 215) Glutaredoxin 2
40	A.60.8.3	1YT3A1	101	(A: 194 - 294) Ribonuclease D
41	A.60.10.1	1EZAA1	123	(A: 22 - 144) Enzyme I of the PEP
42	A.60.16.1	3CI0K2	108	(K: 94 - 203) Pseudopilin GspK
43	A.70.1.1	1ABVA	105	ATP synthase delta subunit, N-terminal domain
44	A.78.1.1	1E2XA2	149	Fatty acid responsive transcription factor
45	A.79.1.3	1SQFA1	140	Ribosomal RNA small subunit methyltransferase
46	A.80.1.1	1JR3A1	126	(A: 243 - 368) DNA polymerase gamma subunit
47	A.80.1.1	1A5TA1	116	(A: 208 - 330) Delta prime subunit
48	A.81.1.1	1JWEA	114	DnaB helicase N-terminal domain
49	A.92.1.1	1A9XA1	153	(A: 403 - 555) Carbamoyl phosphate synthetase
50	A.96.1.2	1KG2A	224	Adenine glycosylase catalytic domain, MutY
51	A.96.1.3	1DIZA1	183	(A: 100 - 282) 3-Methyladenine DNA glycosylase II
52	A.96.1.4	1NKUA	187	3-Methyladenine DNA glycosylase I (tag)
53	A.98.1.1	5R1RA1	221	Ribonucleotide reductase R1 subunit
54	A.100.1.7	1KS9A1	124	(a: 168 - 291) Ketopantoate reductase PanE
55	A.102.1.9	2JF4A1	500	(a: 37 - 547) Periplasmic trehalase TreA
56	A.113.1.1	1WB9A1	297	(a: 270 - 566) DNA repair protein MutS, domain III
57	A.118.1.16	1OYZA	151	Hypothetical protein YibA
58	A.118.5.1	1QSAA1	450	(a: 1 - 450) 70 KDa soluble lytic transglycosylase
59	A.118.8.1	1XNFB	259	Lipoprotein NlpI
60	A.118.8.2	1HZ4A	366	Transcription factor Malt domain III
61	A.118.15.1	1L5JA1	160	Aconitase B, N-terminal domain
62	A.121.1.1	2VKVA2	138	(a: 68 - 205) Tetracyclin repressor protein
63	A.127.1.1	1TJ7A	455	Argininosuccinate lyase/delta-crystallin
64	A.128.1.1	1RQIA	300	Farnesyl diphosphate synthase
65	A.129.1.1	1AONA1	135	(a: 2 - 136, a: 410 - 525) GroEL, E domain
66	A.136.1.1	1DVOA	152	Repressor of bacterial conjugation Fino
67	A.138.1.3	1GU6A	441	Cytochrome c nitrite reductase
68	A.174.1.1	1R6OA1	142	N-terminal, clps-binding domain of ClpA
69	A.177.1.1	1OR7A2	113	RNA polymerase sigma E factor (RpoE)
70	A.177.1.1	1SIGA	205	RNA polymerase sigma factor Sigma70

## Continued

71	A.198.1.1	1QZ4A	213	Hypothetical protein YcfC
72	A.211.1.1	2PAQA1	177	5'-nucleotidase YfbR
73	A.211.1.5	1U6ZA1	197	(A: 313 - 509) Exopolyphosphatase Ppx C-terminal
74	A.223.1.1	1W26A1	185	(A: 248 - 432) Trigger factor, C-terminal domain
75	A.223.1.2	1M5YA1	140	(A: 25 - 164, a: 395 - 427) Porin chaperone SurA
76	A.233.1.1	1WPBA	168	Hypothetical protein YfbU
77	A.236.1.1	1T3WB	134	DNA primase DnaG, C-terminal domain
<i>T. maritima</i>				
1	A.4.5.28	2A61A1	139	Transcriptional regulator TM0710
2	A.4.5.61	2ESHA1	114	Hypothetical protein TM0937
3	A.7.12.1	1SUMB	224	PhoU homolog TM1734
4	A.8.3.3	2B5DX1	114	(X: 415 - 528) Alpha-amylase AmyC
5	A.24.10.3	1TQGA	105	Chemotaxis protein CheA P1 domain
6	A.24.16.3	1O3UA	120	Hypothetical protein TM0613
7	A.24.25.1	1XZPA1	172	(A: 118 - 211, A: 372 - 450) TrmE connector domain
8	A.24.29.1	2P61A1	113	Hypothetical protein TM1646
9	A.25.1.1	1VJXA	144	Hypothetical protein TM1526
10	A.25.1.1	1VLGA	164	Non-hem ferritin
11	A.29.9.1	2ETDA1	118	Hypothetical protein TM0961
12	A.60.13.1	1M6YA1	101	(A: 115 - 215) Putative methyltransferase TM0872
13	A.79.1.1	1TZVA	141	Antitermination factor NusB
14	A.100.1.10	2I76A1	103	(A: 155 - 257) Hypothetical protein TM1727
15	A.118.14.1	1LKVX	213	Flagellar motor switch protein FlgG
16	A.121.1.1	1Z77A2	122	(A: 76 - 200) Transcriptional regulator TM1030
17	A.127.1.1	1C3UA	422	Adenylosuccinate lyase
18	A.128.1.1	1V4EA	280	Octoprenyl-diphosphate synthase
19	A.149.1.1	1O0WA1	169	Rnase III endonuclease catalytic domain
20	A.152.1.2	1P8CE	111	Hypothetical protein TM1620
21	A.191.1.1	1O5HA	200	Hypothetical protein TM1560
22	A.269.1.1	2CE7A1	162	Cell division protein FtsH, C-terminal domain
<i>T. thermophilus</i>				
1	A.3.1.1	1C52A	131	Cytochrome c552
2	A.3.1.5	1NMLA1	166	(1 - 166) Di-heme cytochrome c peroxidase
3	A.3.1.5	1NMLA2	150	(167 - 326) Di-heme cytochrome c peroxidase
4	A.22.1.4	1WWIA1	145	Hypothetical protein tha1479
5	A.24.16.2	1WTYA	116	Probable nucleotidyltransferase subunit

## Continued

6	A.25.1.3	2CWLA1	299	Manganese catalase
7	A.27.1.1	1IQ0A1	126	(a: 467 - 592) Arginyl-tRNA synthetase
8	A.27.1.1	1ILEA1	180	(a: 642 - 821) Isoleucyl-tRNA synthetase
9	A.27.1.1	1A8HA1	152	(a: 349 - 500) Methionyl-tRNA synthetase
10	A.29.3.1	1WS9A1	153	(a: 235 - 387) Acyl-CoA dehydrogenase
11	A.29.12.1	2FUG11	105	(1: 334 - 438) NADH-quinone oxidoreductase chain 1
12	A.75.1.1	1FJGG	155	16S ribosomal protein S7
13	A.97.1.1	1G59A1	163	(a: 306 - 468) C-terminal Glutamyl-tRNA synthetase
14	A.99.1.1	1IQRA1	245	(a: 172 - 416) C-terminal domain of DNA photolyase
15	A.103.1.1	1IOMA	374	Citrate synthase
16	A.104.1.1	1N97A	385	Electron transport molecule cyp175a1
17	A.118.26.1	2YVXA1	125	Magnesium transporter mgtE
18	A.127.1.1	1VDKA	458	Fumarase
19	A.132.1.3	1WWMA1	180	Hypothetical protein ttha0169
20	A.144.2.1	1VSAO1	117	Ribosomal protein L20
21	A.152.1.4	2CWQA1	116	Hypothetical protein ttha0727
22	A.156.1.1	1FJGM	125	Ribosomal protein S13
23	A.174.1.1	1QVRA1	145	N-terminal domain of clpb, chaperone
Human				
1	A.1.1.2	1AJ9B	146	Hemoglobin, beta-chain
2	A.4.1.18	2DW4A1	102	(A: 172 - 273) Lysine-specific histone demethylase 1
3	A.4.3.1	1IG6A	107	MRF-2 DNA-binding domain
4	A.4.3.1	1RYUA	120	SWI-SNF complex protein p270
5	A.4.5.23	1T2KA	110	Interferon regulatory factor 3
6	A.4.5.31	2CSOA1	115	(A: 8 - 122) Pleckstrin
7	A.4.5.43	2AXLA1	144	Werner syndrome ATP-dependent helicase WRN
8	A.4.5.46	1S7AA	103	Lupus La autoantigen N-terminal domain
9	A.5.6.1	2CRUA1	105	programmed cell death protein 5
10	A.6.1.4	1L8RA	101	Retinal determination protein Dachshund
11	A.7.1.1	1OWAA	156	Spectrin alpha chain
12	A.7.1.1	1S35A2	105	(A: 1169 - 1273) Spectrin beta chain
13	A.7.1.1	1S35A1	106	(A: 1063 - 1168) Spectrin beta chain
14	A.7.1.1	1HCIA1	125	(A: 272 - 396) alpha-actinin
15	A.7.1.1	1HCIA2	115	(A: 397 - 511) alpha-actinin
16	A.7.1.1	1HCIA3	121	(A: 512 - 632) alpha-actinin
17	A.7.1.1	1HCIA4	114	(A: 633 - 746) alpha-actinin
18	A.7.4.1	1FEWA	173	Smac/diablo

## Continued

19	A.7.8.1	1NAFA	124	ADP-ribosylation factor binding protein Gga1
20	A.7.11.1	1W0BA	102	Alpha-hemoglobin stabilizing protein
21	A.7.14.1	2CPTA1	104	(A: 8 - 111) Vacuolar sorting protein 4b
22	A.11.2.1	1H4RA1	111	(A: 104 - 214) Merlin
23	A.13.1.1	2FTUA1	118	alpha-2-Macroglobulin receptor associated protein
24	A.22.1.1	2CV5C1	108	Histone H2A
25	A.22.1.3	1JFIB	135	Negative cofactor 2, beta chain
26	A.24.1.1	1EA8A	140	Apolipoprotein E
27	A.24.9.1	1RKCA1	128	Vinculin
28	A.24.9.1	1RKCA2	129	(A: 129 - 258) Vinculin
29	A.24.14.1	1K05A	135	FAT domain of focal adhesion kinase
30	A.26.1.1	1CNT1	150	Ciliary neurotrophic factor
31	A.26.1.1	1ALUA	157	Interleukin-6
32	A.26.1.1	1CD9A	171	Granulocyte-colony stimulating factor
33	A.26.1.1	1F45B	133	Heterodimeric interleukin-12 alpha chain
34	A.26.1.2	1HULA	108	Interleukin-5
35	A.26.1.2	1IJZA	113	Interleukin-13
36	A.26.1.2	1M47A	122	Interleukin-2
37	A.26.1.2	1BUYA	166	Erythropoietin
38	A.26.1.2	2OQPA1	133	Interleukin-21
39	A.26.1.3	1NIFA	153	Interleukin-19
40	A.26.1.3	1AU1A	166	Interferon-beta
41	A.26.1.3	1EKUA1	122	Interferon-gamma
42	A.29.5.1	1Y8NA1	164	(A: 13 - 176) Pyruvate dehydrogenase kinase
43	A.29.7.1	1PIIA	185	Mob1a
44	A.29.10.1	1V9VA1	101	Microtubule-associated Ser/Thr-protein kinase
45	A.35.1.7	1WH8A	111	Homeobox protein Cux-2
46	A.35.1.7	1WIZA	101	DNA-binding protein SATB2
47	A.39.1.2	1M31A	101	Calcyclin (S100)
48	A.39.1.4	1RJVA	110	Parvalbumin
49	A.39.1.5	1DGUA	183	Calcium- and integrin-binding protein
50	A.39.1.6	1IQ3A	110	Pob1
51	A.39.1.7	1EG3A1	125	(A: 85 - 209) Dystrophin
52	A.39.1.7	1TUZA	118	Diacylglycerol kinase alpha, N-terminal domain
53	A.39.1.7	2FJUB1	141	(B: 142 - 311) Phospholipase C-beta-2
54	A.39.1.8	1KFUL1	186	Calpain large subunit, C-terminal domain
55	A.40.1.1	1WJOA	124	Fimbrin (Plastin), actin-crosslinking domain

## Continued

56	A.40.1.1	1DXXA1	111	(A: 9 - 119) Dystrophin
57	A.40.1.1	1DXXB2	127	(B: 120 - 246) Dystrophin
58	A.40.1.1	1UEGA	119	Microtubule-associated protein eb1, N-domain
59	A.41.1.1	1UK0A1	137	Domain of poly(ADP-ribose) polymerase
60	A.45.1.1	1EEMA1	139	(A: 103 - 241) Class omega GST
61	A.45.1.1	1GSEA1	142	(A: 81 - 222) Class alpha GST
62	A.45.1.1	1IYHA1	124	(A: 76 - 199) Class sigma GST
63	A.45.1.1	1K0MA1	149	(A: 92 - 240) Chloride intracellular channel 1
64	A.47.1.1	1BF5A1	167	(A: 136 - 316) STAT-1, coiled coil domain
65	A.48.1.1	1YVHA2	130	(A: 48 - 177) N-terminal domain of cbl
66	A.48.2.1	1CX8A1	152	Transferrin receptor ectodomain, C-terminal domain
67	A.60.7.1	1UL1X1	140	(X: 218 - 357) Flap endonuclease-1
68	A.60.8.1	2DGZA1	100	Werner syndrome ATP-dependent helicase
69	A.60.8.4	2CPRA1	113	(A: 483 - 595) Exosome component 10
70	A.65.1.1	1W7BA	319	Annexin II
71	A.66.1.1	2EBCA1	121	Transducin (alpha subunit), insertion domain
72	A.68.1.1	1EJ5A	107	Wiscott-Aldrich syndrome protein, C-terminal domain
73	A.74.1.1	1E9HB2	123	(B: 310 - 432) Cyclin A
74	A.74.1.1	2I53A2	110	(A: 158 - 267) Cyclin K
75	A.74.1.1	2I53A1	144	(A: 14 - 157) Cyclin K
76	A.74.1.1	1JKWA1	151	(A: 11 - 161) Cyclin H
77	A.74.1.1	1JKWA2	126	(A: 162 - 287) Cyclin H
78	A.74.1.1	1W98B1	130	(B: 228 - 357) G1/S-specific cyclin-E1
79	A.74.1.1	1W98B2	140	(B: 88 - 227) G1/S-specific cyclin-E1
80	A.74.1.2	1C9BA2	109	(A: 208 - 316) Transcription factor IIB, core domain
81	A.74.1.3	1GUXA	177	Retinoblastoma tumor suppressor domains
82	A.77.1.2	1DDFA	125	Fas
83	A.77.1.2	1E3YA	104	Fadd (Mort1)
84	A.77.1.3	1CWWA	102	Apoptotic protease activating factor 1
85	A.77.1.3	3CRDA	100	Raidd CARD domain
86	A.83.1.1	1QK1A1	102	Creatine kinase, N-domain
87	A.87.1.1	1TXDA1	234	Rho guanine nucleotide exchange factor 12
88	A.87.1.1	1DBHA1	207	(A: 198 - 404) Son of sevenless-1
89	A.91.1.1	1DK8A	147	Axin RGS-homologous domain
90	A.91.1.1	1IAPA	190	p115 RhoGEF
91	A.100.1.3	1F12A1	101	Short chain L-3-hydroxyacyl CoA dehydrogenase
92	A.102.4.2	1W6JA2	279	(A: 100 - 378) Lanosterol synthase

## Continued

93	A.102.4.3	1JCQB	410	Protein farnesyltransferase, beta-subunit
94	A.102.4.4	1GHQA	307	a C3 fragment and ligand for complement receptor
95	A.114.1.1	1DG3A1	300	Interferon-induced guanylate-binding protein 1
96	A.116.1.1	1GRNB	197	Cdc42GAP
97	A.116.1.1	1PBWA	184	p85 alpha subunit RhoGAP domain
98	A.116.1.2	1WERA	324	p120GAP domain
99	A.117.1.1	1NVUS	469	Son of sevenless protein homolog 1
100	A.118.1.1	1T08A	510	beta-Catenin
101	A.118.1.2	2IE3A1	581	Phosphatase 2, Constant regulatory domain
102	A.118.1.6	1E8YA1	186	Phosphoinositide 3-kinase helical domain
103	A.118.1.7	1GW6A1	150	(A: 461 - 610) Leukotriene A4 hydrolase C-domain
104	A.118.1.8	1M8WA	340	Pumilio 1, RNA binding protein
105	A.118.1.14	1H2TC3	251	(C: 481 - 790) CBP80, 80 KDa nuclear cap-binding protein
106	A.118.1.14	1UG3A1	193	(A: 1235 - 1427) Eukaryotic initiation factor elf4G
107	A.118.1.14	1UG3A2	127	(A: 1438 - 1564) Eukaryotic initiation factor eIF4G
108	A.118.1.14	1HU3A	204	Eukaryotic initiation factor eIF4G
109	A.118.1.14	1H2TC2	190	CBP80, 80 KDa nuclear cap-binding protein
110	A.118.1.14	1H2UA1	265	CBP80, 80 KDa nuclear cap-binding protein
111	A.118.1.18	1RZ4A2	130	Eukaryotic translation initiation factor 3
112	A.118.1.19	1W9CA	321	Exportin-1 (Xpo1, Crm1)
113	A.118.1.20	2JAKA1	327	Serine/threonine-protein phosphatase 2A
114	A.118.1.21	1XQRA1	264	(A: 87 - 350) Hsp70-binding protein 1
115	A.118.6.1	1JCQA	313	Protein farnesyltransferase alpha-subunit
116	A.118.7.1	1QJAA	217	zeta isoform
117	A.118.8.1	1ELRA	128	Hop
118	A.118.8.1	1YA0A1	458	SMG-7 transcript variant 2
119	A.118.8.1	1KT0A1	155	(A: 254 - 412) FKBP51, C-terminal domain
120	A.118.9.1	1INZA	148	Epsin 1, Eps15-interacting protein
121	A.118.9.2	1ELKA	153	Tom1 protein
122	A.118.12.1	2GRNB1	157	Ran-GTPase activating protein 1, C-domain
123	A.118.16.1	1J1JA	217	Translin
124	A.118.17.1	2HYEC2	347	(C: 55 - 401) Cullin-4A
125	A.118.19.1	1RW2A	152	C-terminal domain of Ku80
126	A.118.23.1	2A9UA1	134	Ubiquitin carboxyl-terminal hydrolase 8
127	A.123.1.1	1P93A	247	Glucocorticoid receptor
128	A.123.1.1	1LV2A	225	Hepatocyte nuclear factor 4-gamma
129	A.123.1.1	1EXAA	236	Retinoic acid receptor gamma

## Continued

130	A.126.1.1	1J78A2	188	(A: 199 - 386) Vitamin D binding protein
131	A.128.1.2	1EZFA	323	Squalene synthase
132	A.132.1.1	1N3UA	214	Heme oxygenase-1 (HO-1)
133	A.143.1.2	1QKLA	127	RPB6
134	A.144.1.1	1G9LA	144	poly(A) binding protein
135	A.146.1.1	1H6OA	195	Telomeric repeat binding factor 1, TRF1
136	A.156.1.2	1TDHA1	110	(A: 132 - 246) Endonuclease VIII-like 1
137	A.158.1.1	1P22A1	109	(A: 135 - 252) F-box/WD-repeat protein 1
138	A.169.1.1	1MIIA1	305	(A: 2249 - 2553) BEACH domain of neurobeachin
139	A.173.1.1	1OU5A1	204	(A: 151 - 354) tRNA CCA-adding enzyme, c-domain
140	A.183.1.1	2OZBB1	239	U4/U6 small nuclear ribonucleoprotein Prp31
141	A.188.1.1	1MP1A	111	Ser/Arg-related nuclear matrix protein
142	A.211.1.2	1MKDA	328	Cyclic nucleotide phosphodiesterase catalytic domain
143	A.224.1.1	1SX6A	205	Glycolipid transfer protein, GLTP
144	A.235.1.1	1X9NA1	264	(A: 262 - 533) DNA ligase I
145	A.238.1.1	1X03A1	210	(A: 26 - 247) Endophilin-1
146	A.238.1.1	2ELBA1	258	(A: 6 - 273) DCC-interacting protein 13-alpha
147	A.238.1.2	1I49A	201	Arfaptin, Rac-binding fragment
148	A.238.1.4	2EFKA1	272	(A: 10 - 288) CDC42-interacting protein 4, CIP4
149	A.266.1.2	2D0TA1	373	(A: 12 - 403) Indoleamine 2,3-dioxygenase
150	A.268.1.1	2G62A1	295	Serine/threonine-protein phosphatase 2A
151	A.278.1.1	2E9XA1	144	DNA replication complex GINS protein PSF1
152	A.278.1.2	2Q9QA1	112	DNA replication complex GINS protein PSF2
153	A.283.1.1	2FMME1	116	(E: 9 - 124) Emsy
154	A.289.1.1	2Q0ZX1	176	(X: 33 - 208) Protein pro2281
Yeast				
1	A.2.17.2	2F6MB1	104	Vacuolar protein sorting-associated protein 28
2	A.3.1.1	1CHHA	108	Mitochondrial cytochrome c
3	A.3.1.3	1EZVD1	199	Cytochrome bc1 domain
4	A.4.1.6	1IGNA2	103	(A: 446 - 594) DNA-binding domain of rap1
5	A.4.3.1	1KKXA	102	Transcription regulator Adr6 (Swi1)
6	A.5.4.1	1ENWA	114	Elongation factor TFIIS domain 2
7	A.7.5.1	1QSDA	102	Tubulin chaperone cofactor A
8	A.24.10.2	1C03A	163	Phosphorelay protein ypd1
9	A.24.15.1	1JR8A	102	Thiol oxidase Erv2p
10	A.25.1.2	1JK0A	334	Ribonucleotide reductase R2
11	A.29.2.1	1E6IA	110	Transcriptional activator GCN5

## Continued

12	A.39.1.5	1M46A	148	Myosin light chain Mic1p
13	A.39.1.5	1FPWA	190	Frequenin (neuronal calcium sensor 1)
14	A.45.1.1	1G6WA1	133	Prion protein ure2p, nitrogen regulation fragment
15	A.45.1.1	1NHYA1	140	Elongation factor 1-gamma, GST-like domain
16	A.45.1.2	2HQTAl	115	GU4 nucleic-binding protein 1, Arc1p
17	A.47.2.1	1FIOA	190	Sso1
18	A.60.8.2	1Y14A	133	RNA polymerase II subunit RBP4 (RpoF)
19	A.60.9.1	1FLOA1	123	Flp recombinase
20	A.69.2.1	1FKMA1	194	(A: 249 - 442) Ypt/Rab-GAP domain of gyp1p
21	A.69.2.1	1FKMA2	128	(A: 443 - 630) Ypt/Rab-GAP domain of gyp1p
22	A.71.2.1	1M2OA1	103	(A: 524 - 626) Sec23
23	A.71.2.1	1M2VB1	107	(B: 647 - 753) Sec24
24	A.72.1.1	1DVKA	149	Functional domain of the splicing factor Prp18
25	A.80.1.1	1SXJA1	146	(A: 548 - 693) Replication factor C1
26	A.93.1.1	1CYFA	296	Cytochrome c peroxidase
27	A.102.2.1	1DL2A	510	Class I alpha-1;2-mannosidase, catalytic domain
28	A.118.1.1	1WA5C	937	Exportin Cse1p
29	A.118.1.1	1WA5B	463	Karyopherin alpha
30	A.118.1.9	1HO8A	447	Regulatory subunit H of the V-type ATPase
31	A.118.1.14	1PAQA	161	Translation initiation factor eIF-2b epsilon
32	A.118.3.1	1RE0B	195	ARF guanine-exchange factor 1
33	A.118.8.1	1Y8MA1	138	Mitochondria fission protein Fis1
34	A.118.8.1	1QQEA	281	Vesicular transport protein sec17
35	A.118.9.4	1SZAA	144	PCF11 protein
36	A.118.17.2	2D2SA1	217	(A: 525 - 753) Exocyst complex component EXO84
37	A.118.17.2	2B7MA1	513	(A: 73 - 623) Exocyst complex component EXO70
38	A.130.1.2	1CSMA	252	Allosteric chorismate mutase
39	A.158.1.1	1NEXB1	100	(B: 270 - 369) Cdc4 F-box and linker domains
40	A.160.1.1	1FA0A3	150	(A: 202 - 351) Poly(A) polymerase, middle domain
41	A.207.1.1	1UX5A	411	Bni1
42	A.210.1.1	1RF8B	100	Eukaryotic initiation factor 4f subunit eIF4g
43	A.227.1.1	1RP4A	374	Endoplasmic oxidoreductin 1, Ero1p
44	A.232.1.1	1XLYA	224	RNA-binding protein She2p
45	A.252.1.1	1YKEB1	119	RNA polymerase II holoenzyme component SRB7

<i>E. coli</i>	beta proteins			
No.	SCOP	PDB	residues	protein name
1	B.1.4.1	1BGLA2	105	(A: 626 - 730) beta-Galactosidase, domain 4
2	B.1.4.1	1BGLA1	114	(A: 220 - 333) beta-Galactosidase, domain 2
3	B.1.11.1	1N0LA1	124	Pilus chaperone PapD, N-domain
4	B.1.17.1	1VRS A1	121	Thiol:disulfide interchange protein, DsbD-alpha
5	B.1.18.2	1M7XA1	110	1,4-alpha-glucan branching enzyme N-terminal domain
6	B.1.18.2	1TXKA1	115	(A:397-511) Glucans biosynthesis protein G
7	B.1.18.17	1IX2B	102	Copper resistance protein C (CopC, PcoC)
8	B.2.3.2	2AXWA1	121	Invasin AfaD
9	B.2.3.2	2UY6B1	150	Pap fimbrial major pilin protein PapA
10	B.2.3.2	1N12A	134	PapE pilus subunit
11	B.2.3.2	1PDKB	149	PapK pilus subunit
12	B.2.3.2	1ZE3H1	118	(H: 159 - 279) Mannose-specific adhesin FimH
13	B.2.3.2	1TR7A1	158	Mannose-specific adhesin FimH
14	B.2.3.3	1J8RA	196	PapG adhesin receptor-binding domain
15	B.2.3.5	1O9VA	168	Fimbrial adhesin F17-AG lectin domain
16	B.2.3.6	1USZA	140	DraA/Afimbrial adhesin Afa-III
17	B.6.1.2	1FFT B1	166	(B: 118 - 283) Quinol oxidase (CyoA)
18	B.6.1.3	1N68A1	141	(A: 30 - 170) Multi-copper oxidase CueO
19	B.6.1.3	1N68A2	165	(A: 171 - 335) Multi-copper oxidase CueO
20	B.6.1.3	1N68A3	158	(A: 336 - 516) Multi-copper oxidase CueO
21	B.17.1.2	1FJJA	154	Hypothetical protein YbhB
22	B.18.1.5	1BGLA3	217	beta-Galactosidase
23	B.30.2.1	1D6UA1	424	(A: 301 - 724) Copper amine oxidase, domain 3
24	B.30.5.11	2F2HA2	247	Putative glucosidase YicI, N-terminal domain
25	B.33.1.3	2JO6A1	108	NADH-nitrite reductase small subunit
26	B.34.5.1	1VS6U1	102	Ribosomal proteins L24 (L24p)
27	B.34.6.2	1UB4A	103	MazF protein
28	B.35.1.2	1O89A1	143	(A: 1 - 115, A: 293 - 323) Hypothetical protein YhdH
29	B.35.1.2	1QORA1	147	(A: 2 - 112, A: 292 - 327) Quinone oxidoreductase
30	B.36.1.4	1TE0A1	100	Stress sensor protease DegS, C-terminal domain
31	B.40.2.1	1B44D	106	Heat-labile toxin type IB
32	B.40.4.1	1E1OA1	143	Lysyl-tRNA synthetase (LysRS)
33	B.40.4.3	1V1QA	110	Primosomal replication protein N, PriB
34	B.40.4.3	1QVCA	145	ssDNA-binding protein

## Continued

35	B.40.5.1	1FAJA	168	Inorganic pyrophosphatase
36	B.40.6.3	2AWNA1	135	Maltose transport protein MalK, C-terminal domain
37	B.40.9.1	1SR3A	114	Heme chaperone CcmE
38	B.43.2.1	1FUIA1	236	(A: 356 - 591) L-fucose isomerase, C-terminal domain
39	B.43.2.2	2AJTA1	170	(A: 329 - 498) L-arabinose isomerase AraA
40	B.43.4.1	1DDGA1	221	(A: 226 - 446) Sulfite reductase flavoprotein
41	B.43.4.2	1GVHA2	107	(A: 147 - 253) Flavohemoglobin, central domain
42	B.43.4.3	1I8DA2	113	(A: 94 - 206) Riboflavin synthase
43	B.45.1.3	2IDBA1	308	3-octaprenyl-4-hydroxybenzoate carboxy-lyase
44	B.46.1.1	1FMTA1	108	Methionyl-trna <sup>fmet</sup> formyltransferase
45	B.49.2.3	1KNWA1	174	(A:2-31,A:279-422) Diaminopimelate decarboxylase
46	B.52.1.4	2AE0X1	335	Membrane-bound lytic murein transglycosylase A
47	B.52.2.1	1PPYA	118	Pyruvoyl dependent aspartate decarboxylase
48	B.52.2.2	1AA6A1	132	(A: 565 - 715) Formate dehydrogenase H
49	B.52.2.2	1Q16A1	170	Respiratory nitrate reductase 1 alpha chain
50	B.53.1.2	1EUQA1	198	(A: 339 - 547) Gln-tRNA synthetase, C-terminal
51	B.58.1.2	1O65A	218	Hypothetical protein YiiM
52	B.60.1.1	1QWDA	167	Outer membrane lipoprotein Blc
53	B.60.1.4	1OEJA	187	Hypothetical protein YodA
54	B.68.4.1	1C5KA1	269	(A: 163 - 431) TolB, C-terminal domain
55	B.69.11.1	1RI6A	333	Putative isomerase YbhE
56	B.80.6.1	1VH4B	410	Stabilizer of iron transporter SufD
57	B.81.1.1	1LXAA	262	UDP N-acetylglucosamine acyltransferase
58	B.81.1.4	1HV9A1	200	N-acetylglucosamine 1-phosphate uridyltransferase
59	B.82.1.11	1RC6A	242	Hypothetical protein YlbA
60	B.82.1.12	1TQ5A1	230	Hypothetical protein YhhW
61	B.82.1.13	1XRUA1	277	5-keto-4-deoxyuronate isomerase Kdul
62	B.82.1.14	1YQCA1	159	Ureidoglycolate hydrolase AIIA
63	B.82.1.23	2D40A1	288	(A: 35 - 342) Gentisate 1,2-dioxygenase
64	B.82.2.3	1JR7A	306	Gab protein (hypothetical protein YgaT)
65	B.82.2.5	1OTJA	281	Taurine/alpha-ketoglutarate dioxygenase TauD
66	B.82.2.10	2FD8A1	199	(A: 15 - 214) Alkylated DNA repair protein AlkB
67	B.82.2.13	3BB6A1	104	Uncharacterized protein YeaR
68	B.82.3.2	1I5ZA2	132	Catabolite gene activator protein
69	B.82.4.1	1XJAC	165	Regulatory protein AraC
70	B.84.2.1	1BNCA1	116	(A: 331 - 446) Biotin carboxylase, C-domain

**Continued**

71	B.84.3.1	2F3GA	150	Glucose-specific factor III (glsIII)
72	B.85.4.1	1DUDA	136	Deoxyuridine 5'-triphosphate nucleotidohydrolase
73	B.85.4.1	1XS1A	193	Deoxycytidine triphosphate deaminase
74	B.87.1.1	1JHHB	124	LexA repressor C-terminal domain
75	B.87.1.2	1KN9A	234	Type 1 signal peptidase
76	B.92.1.2	1K6WA1	103	(A: 4 - 55, A: 376 - 426) Cytosine deaminase
77	B.92.1.7	1ONWA1	105	(A: 1 - 62, A: 347 - 389) Isoaspartyl dipeptidase
78	B.110.1.1	1RH1A1	290	(A: 10 - 312) Colicin B N-terminal domain
79	B.113.1.1	1K3WA2	124	Endonuclease VIII
80	B.122.1.7	1TE7A	103	Hypothetical protein YqfB
81	B.122.1.10	2ANEA1	110	ATP-dependent protease La, N-terminal domain
82	B.125.1.1	1IWL A	177	Outer-membrane lipoproteins carrier protein LolA
83	B.125.1.2	1IWMA	177	Outer membrane lipoprotein receptor LolB
84	B.130.1.1	1BPRA1	126	(A: 381 - 506) Chaperone DnaK
85	B.136.1.1	1OX9A	108	Stringent starvation protein B, SspB
86	B.142.1.1	1NA6A1	172	Restriction endonuclease EcoRII, N-terminal domain
87	B.150.1.1	1XSIA1	108	Putative glucosidase YicI, C-terminal domain
88	B.153.1.2	1NI5A3	118	(A: 315 - 432) tRNA-Ile-lysine synthetase, TilS
89	B.167.1.1	1ZE3D1	116	Outer membrane usher protein FimD
90	B.177.1.1	2IN5A1	195	Hypothetical lipoprotein YmcC
<i>T. maritima</i>				
1	B.1.9.2	1I82A	189	Xylanase 10A
2	B.3.1.3	2J71A1	102	Pullulanase PulA
3	B.18.1.14	1GUIA	155	Carbohydrate binding module from laminarinase 16A
4	B.23.3.1	2F4LA1	274	Putative acetamidase TM0119
5	B.29.1.19	1UYPA1	138	Beta-fructosidase (invertase), C-terminal domain
6	B.30.5.11	1ZY9A1	171	Alpha-galactosidase GalA N-terminal domain
7	B.35.1.2	1VJ0A1	154	Hypothetical protein TM0436
8	B.40.4.9	1GM5A2	175	(A: 106 - 285) RecG "wedge" domain
9	B.40.7.1	1K0SA	151	Chemotaxis protein CheW
10	B.40.11.1	2F4IA1	172	(A: 39 - 214) Hypothetical protein TM0957
11	B.47.1.1	1L1JA	226	Protease Do (DegP, HtrA), catalytic domain
12	B.62.1.3	1ZX8A1	123	Hypothetical protein TM1367
13	B.67.2.4	1VKDA	324	Hypothetical protein TM1225
14	B.80.3.1	1HF2A1	107	Cell-division inhibitor MinC, C-terminal domain
15	B.82.1.9	2F4PA1	134	Hypothetical protein TM1010

## Continued

16	B.82.1.9	1O4TA	115	Hypothetical protein TM1287
17	B.123.1.1	1NC7A	114	Hypothetical protein TM1070
<i>T. thermophilus</i>				
1	B.1.18.19	1V8HA1	106	Sulfur oxidation protein SoxZ
2	B.6.1.2	2CUAB	122	Cytochrome c oxidase
3	B.33.1.1	1NYKA	156	Soluble Rieske protein
4	B.34.5.1	1VSAS1	101	Ribosomal proteins L24
5	B.34.5.3	2HGJD1	147	Ribosomal protein L2, C-terminal domain
6	B.34.5.6	2J01T1	137	Ribosomal protein L19
7	B.39.1.1	1VSAI1	122	Ribosomal protein L14
8	B.40.4.1	1EFWA1	104	Aspartyl-tRNA synthetase
9	B.40.4.4	1GD7A	109	TRBP111 homolog CsaA
10	B.40.4.4	1B70B3	113	(B: 39 - 151) Domain B2 of PheRS-beta, PheT
11	B.40.4.5	1FJGQ	104	Ribosomal protein S17
12	B.40.4.5	2J03D2	125	Ribosomal protein L2, N-terminal domain
13	B.40.4.5	1I94L	131	Ribosomal protein S12
14	B.43.3.1	1WDTA1	103	(A: 275 - 377) Elongation factor G, domain II
15	B.43.3.1	1AIPA1	100	Elongation factor Tu, domain 2
16	B.43.3.2	2HGJE1	205	Ribosomal protein L3
17	B.45.1.2	1USCA	178	Putative styrene monooxygenase small component
18	B.51.1.1	1ILEA2	189	(A: 198 - 386) Isoleucyl-tRNA synthetase
19	B.53.1.1	1FEUD	185	Ribosomal protein TL5 (general stress protein CTC)
20	B.61.6.1	1WUBA	176	Polyisoprenoid-binding protein TTHA0802
21	B.82.1.9	1V70A	105	Hypothetical protein TTHA0104
22	B.84.1.1	1ONLA	127	Protein H of glycine cleavage system
23	B.111.1.1	1J1HA	123	Small protein B
24	B.113.1.1	1EE8A2	121	DNA repair protein MutM (Fpg)
25	B.117.1.1	1UDXA1	156	Obg GTP-binding protein N-terminal domain
26	B.122.1.3	1V47A1	132	ATP sulfurylase N-terminal domain
27	B.153.1.1	1B70B6	209	(B: 191 - 399) B3/B4 domain of PheRS, PheT
28	B.155.1.1	1VSAP1	101	Ribosomal protein L21p
human				
1	B.1.1.1	1CDBA	105	CD2, first domain
2	B.1.1.1	1AKJD	114	CD8
3	B.1.1.1	1G9MH1	129	Immunoglobulin heavy chain variable domain
4	B.1.1.1	1I85C	118	Immunoreceptor CTLA-4, N-terminal fragment

## Continued

5	B.1.1.1	1KACB	124	Coxsackie virus and adenovirus receptor
6	B.1.1.1	1O7SA	112	Sialic acid binding Ig-like lectin 7
7	B.1.1.1	1TVDA	116	T-cell antigen receptor
8	B.1.1.2	1A9BB	100	beta2-microglobulin
9	B.1.1.2	1BD2E2	129	(E: 119 - 247) T-cell antigen receptor
10	B.1.1.3	1IJ9A1	106	(A: 91 - 196) Vascular cell adhesion molecule-1
11	B.1.1.3	1ZXQA1	106	(A: 87 - 192) Intercellular cell adhesion molecule-2
12	B.1.1.4	2AVGA1	110	Cardiac myosin binding protein c, different domains
13	B.1.1.4	2CQVA1	101	Telokin
14	B.1.1.4	1BQSA2	119	Mucosal addressin cell adhesion molecule-1
15	B.1.1.4	1CVSC2	109	(C: 251 - 359) Fibroblast growth factor receptor
16	B.1.1.4	1EFXD1	100	(D: 4 - 103) Killer cell inhibitory receptor
17	B.1.1.4	1II4F1	101	(F: 150 - 250) Fibroblast growth factor receptor
18	B.1.1.4	1NBQA2	104	Junction adhesion molecule, C-terminal domain
19	B.1.1.4	1QSVA	101	Ff1t-1 receptor second domain
20	B.1.1.4	1WWAX	105	High affinity nerve growth factor receptor
21	B.1.1.4	3B5HA1	101	(A: 103 - 203) Cervical EMMPRIN
22	B.1.2.1	1WF5A1	108	(A: 8 - 115) Sidekick 2
23	B.1.2.1	1AXIB2	106	(B: 131 - 236) Growth hormone receptor
24	B.1.2.1	1WK0A	137	Fibronectin type-III domain containing protein 3a
25	B.1.2.1	1X5KA1	111	(A: 8 - 118) Neogenin
26	B.1.2.1	2CSPA1	117	Rim binding protein 2
27	B.1.2.1	2CUHA1	102	Tenascin-x
28	B.1.2.1	2HAZA1	101	(A: 489 - 589) Neural cell adhesion molecule 1
29	B.1.2.1	1EBAA2	105	(A: 117 - 224) erythropoietin receptor
30	B.1.2.1	1F42A2	119	(A: 88 - 211) The p40 domain of interleukin-12
31	B.1.2.1	1FG9C2	112	Interferon-gamma receptor alpha chain
32	B.1.2.1	1GH7A1	103	GM-CSF, IL-3 and IL-5 receptor beta-chain
33	B.1.2.1	1IIRA1	100	Cytokine receptor gp130 cytokine-binding domains
34	B.1.2.1	1J7VR2	106	(R: 101 - 206) Interleukin-10 receptor 1
35	B.1.2.1	1N26A2	102	(A: 94 - 195) Interleukin-6 receptor alpha chain
36	B.1.2.1	1N6UA2	103	Interferon-alpha/beta receptor beta chain
37	B.1.2.1	2D9QB3	107	(B: 97 - 203) Granulocyte colony-stimulating factor
38	B.1.2.1	1EBAA1	107	(A: 10 - 116) Erythropoietin receptor
39	B.1.2.1	1GH7A2	114	GM-CSF, IL-3 and IL-5 receptor beta chain
40	B.1.2.1	1N6UA1	109	Interferon-alpha/beta receptor beta chain

## Continued

41	B.1.3.1	1WGOA	123	VPS10 domain-containing receptor SorCS2
42	B.1.4.1	1BHGA1	103	(A: 226 - 328) beta-Glucuronidase
43	B.1.5.1	1EVUA2	112	Transglutaminase, two C-terminal domains
44	B.1.5.1	2Q3ZA2	114	Transglutaminase, two C-terminal domains
45	B.1.5.1	1EVUA3	100	Transglutaminase, two C-terminal domains
46	B.1.10.1	1E42A1	120	Beta2-adaptin AP2 ear domain, N-terminal subdomain
47	B.1.10.2	2DWXA1	129	ADP-ribosylation factor binding protein Gga1 domain
48	B.1.10.3	1R4XA1	163	Coatomer gamma subunit C-terminal domain
49	B.1.16.1	1IVTA	122	Lamin A/C globular tail domain
50	B.1.18.1	1OWRM1	103	(M: 576 - 678) T-cell transcription factor NFAT1
51	B.1.18.9	1L9MA1	140	Transglutaminase N-terminal domain
52	B.1.18.10	2DIBA1	115	Filamin b
53	B.1.18.10	2DMCA1	103	Filamin b
54	B.1.18.23	3ES6B1	118	Prolactin-inducible protein
55	B.2.5.2	1TSRB	194	p53 tumor suppressor, DNA-binding domain
56	B.2.5.3	1A02N2	178	T-cell transcription factor, DNA-binding domain
57	B.2.5.3	1SVCP2	208	(P: 43 - 250) p50 subunit of NF-kappa B
58	B.2.5.4	1H6FA	184	T-box protein 3
59	B.2.5.5	1BF5A2	252	(A: 317 - 568) STAT-1, DNA-binding domain
60	B.2.5.6	1E50C	125	Acute myeloid leukemia 1 protein
61	B.2.7.1	1H6EA	219	Mu2 adaptin subunit of ap2 adaptor second domain
62	B.2.9.1	1WD8A1	155	(A: 113 - 293) Ppeptidylarginine deiminase
63	B.3.3.1	1LM8V	150	VHL
64	B.3.4.1	1BM7A	114	Transthyretin (Prealbumin)
65	B.6.1.3	1KCWA1	192	Ceruloplasmin
66	B.7.1.1	1BCIA	123	Domain from cytosolic phospholipase A2
67	B.7.1.1	2EP6A1	126	Multiple C2 and transmembrane domain-containing protein
68	B.7.1.1	2FJUB2	122	(B: 678 - 799) Phospholipase C-beta-2
69	B.7.1.1	2YRBA1	142	(A: 596 - 737) Fantom
70	B.7.1.1	1D5RA1	125	Pten tumor suppressor (Phosphoinositide phosphatase)
71	B.7.1.1	1E8YA2	135	(A: 357 - 522) Phosphoinositide 3-kinase
72	B.7.1.1	2NQ3A1	131	E3 ubiquitin-protein ligase
73	B.7.1.2	2R83A1	123	(A: 271 - 393) Synaptogamin I
74	B.7.1.2	1WFMA	138	Synaptotagmin XIII
75	B.8.1.1	1FLKA1	155	(A: 350 - 504) TNF receptor associated factor 3
76	B.8.1.1	2CR2A1	146	Speckle-type poz protein

## Continued

77	B.14.1.1	1KFUL2	159	(L: 356 - 514) Calpain large subunit, middle domain
78	B.15.1.2	1EJFA	110	Co-chaperone p23
79	B.15.1.3	1WH0A	134	Ubiquitin carboxyl-terminal hydrolase 19
80	B.15.1.4	1WGVA	124	NudC domain containing protein 3
81	B.18.1.4	2BBAA1	180	(A: 17 - 196) Ephrin type-b receptor 4
82	B.18.1.8	1XNAA	151	N-terminal domain of xrcc1
83	B.18.1.9	1XPWA	143	Placental protein 25
84	B.22.1.1	1RJ7A	144	Ectodysplasin A
85	B.22.1.1	1PK6C	129	Complement c1q globular head, C chain
86	B.22.1.1	1TNRA	144	Tumor necrosis factor
87	B.22.1.1	1ALYA	146	Extracellular domain of CD40 ligand
88	B.23.1.1	1SZBA1	121	Mannose-binding protein associated serine protease 2
89	B.26.1.1	1DD1A	221	Smad4 tumor suppressor C-terminal domain
90	B.26.1.2	1R21A	100	Antigen Ki-67
91	B.26.1.2	2BRFA1	100	(A: 8 - 108) Polynucleotide kinase 3'-phosphatase
92	B.26.1.2	2G1LA1	102	(A: 498 - 599) Kinesin-like protein kif1c
93	B.26.1.2	2PIEA1	127	(A: 13 - 139) Ubiquitin ligase protein RNF8
94	B.26.1.2	1GXCA	116	Chk2 kinase
95	B.29.1.3	1A3KA	137	Galectin-3 CRD
96	B.29.1.4	1H30A2	205	(A: 461 - 678) growth-arrest-specific protein Gas6
97	B.29.1.5	1B09A	206	C-reactive protein
98	B.29.1.22	2FBEA1	188	Similar to Ret finger protein-like 1
99	B.34.2.1	1NG2A2	118	p47pox (neutrophil cytosolic factor 1)
100	B.34.9.2	1RI0A	100	Hepatoma-derived growth factor
101	B.34.9.3	1WJRA	127	Scm-like with four MBT domains protein 2
102	B.34.10.1	2CP6A1	160	(A: 8 - 167) Restin
103	B.34.10.1	1IXDA	104	Cylindromatosis tumour-suppressor
104	B.34.14.1	1SI2A	120	Eukaryotic translation initiation factor 2C 1
105	B.36.1.1	1UEQA	123	Membrane associated guanylate kinase inverted-2
106	B.36.1.1	1UF1A	128	KIAA1526 protein
107	B.36.1.2	1I16A	130	Interleukin 16
108	B.40.3.3	1UAPA	131	Procollagen c-proteinase enhancer protein
109	B.40.4.3	1EWIA	114	Replication protein A 70 KDa subunit
110	B.40.4.3	1FGUA1	118	(A: 181 - 298) Replication protein A 70 KDa subunit
111	B.40.4.3	1XJVA2	151	(A: 149 - 299) Protection of telomeres protein 1
112	B.40.4.3	2PQAB1	114	(B: 3 - 116) Rreplication protein A 14 KDa

## Continued

113	B.40.4.3	1FGUA2	128	(A: 299 - 426) Replication protein A 70 KDa subunit
114	B.40.4.3	2PI2A1	121	(A: 44 - 171) Replication protein A 32 KDa subunit
115	B.40.4.3	1XJVA1	140	(A: 6 - 145) protection of telomeres protein 1
116	B.40.4.4	1NTGA	171	Metazoan tyrosyl-tRNA synthetase, C-terminal domain
117	B.40.4.5	1WI5A	119	RRP5 protein homolog S1-domain
118	B.40.4.5	2NN6I1	125	(I: 61 - 185) Exosome component 1
119	B.40.4.5	1D7QA	143	Translation initiation factor-1a
120	B.40.4.6	1X9NA2	148	(A: 754 - 901) DNA ligase I
121	B.40.16.1	2OQ0A2	101	(A: 12 - 114) Gamma-interferon-inducible protein
122	B.42.1.1	1AXMC	128	Acidic FGF
123	B.42.1.2	1IOBA	153	Interleukin-1 beta
124	B.42.1.2	1J0SA	157	Interleukin-18
125	B.42.5.1	1DFCB3	120	(B: 2260 - 2382) Fascin
126	B.43.4.2	1UMKA1	124	(A: 30 - 153) Cytochrome b5 reductase
127	B.43.5.1	1NB0A	147	Riboflavin kinase
128	B.45.1.1	1XHNA1	170	Cellular repressor of E1A-stimulated genes
129	B.46.1.1	2BW0A1	104	10-formyltetrahydrofolate dehydrogenase domain 2
130	B.46.1.2	1F6OA	211	3-methyladenine DNA glycosylase
131	B.47.1.2	1TRNA	224	Trypsin(ogen)
132	B.54.1.1	1CL3A	138	Core binding factor beta
133	B.55.1.1	1BAKA	119	G-protein coupled receptor kinase 2
134	B.55.1.1	1DYNA	113	Dynamin
135	B.55.1.1	1V88A	130	Oxysterol binding protein-related protein 8
136	B.55.1.1	1X1FA1	136	(A: 8 - 143) Signal-transducing adaptor protein 1
137	B.55.1.1	1X1GA1	116	(A: 8 - 123) Pleckstrin-2
138	B.55.1.1	1XD4A3	131	(A: 419 - 549) Son of sevenless-1
139	B.55.1.1	2COAA1	112	Protein kinase c, d2 type
140	B.55.1.1	1B55A	163	Bruton's tyrosine kinase
141	B.55.1.1	1KI1B2	132	(B: 1439 - 1580) GEF of intersectin
142	B.55.1.1	1UNPA	119	Rac-alpha serine/threonine kinase
143	B.55.1.1	2FJUB3	131	Phospholipase c-beta-2
144	B.55.1.2	1OY2A	191	Shc adaptor protein
145	B.55.1.2	2DKQA1	147	Tensin
146	B.55.1.2	1XR0B	126	FGFR substrate 2
147	B.55.1.2	1J0WB	103	Downstream of tyrosine kinase 5
148	B.55.1.3	1K5DB	146	Ran-binding protein 1

## Continued

149	B.55.1.4	2IYBA1	111	Enabled
150	B.55.1.5	1E5WA2	148	(A: 199 - 346) Moesin
151	B.55.1.6	1MIIA2	109	(A: 2140 - 2248) PH-like domain of neurobeachin
152	B.55.1.8	1LW3A1	125	Myotubularin-related protein 2, N-terminal domain
153	B.55.1.9	1PFJA	108	TFIIH basal transcription factor complex 62 subunit
154	B.55.1.12	2HTHB1	129	Vacuolar protein sorting protein 36
155	B.59.1.1	1FU1A1	117	XRCC4, N-terminal domain
156	B.60.1.1	1XKIA	128	Von Ebner's gland protein
157	B.60.1.1	1IW2A	163	Complement protein C8 gamma
158	B.60.1.2	1B56A	133	Epidermal fatty acid binding protein
159	B.61.5.1	1K3BA	119	Dipeptidyl peptidase I (cathepsin C)
160	B.63.1.1	1QTTA	117	p13-MTGP1
161	B.65.1.1	1F3UA	118	TFIIF beta subunit, Rap30
162	B.66.1.1	1SU3A2	195	(A: 271 - 465) Collagenase, C-terminal domain
163	B.67.3.1	1S18A	317	soluble calcium-activated nucleotidase SCAN-1
164	B.68.1.1	1SNTA	352	Sialidase 2
165	B.68.5.1	1IJQA1	258	(A: 377 - 642) Low density lipoprotein receptor
166	B.68.11.1	1U6DX	288	Kelch-like ECH-associated protein 1
167	B.69.5.1	1I2MB	388	Regulator of chromosome condensation RCC1
168	B.69.8.1	1JV2A4	438	Integrin alpha N-terminal domain
169	B.69.12.1	1SHYB1	450	(B: 40 - 515) Hepatocyte growth factor receptor
170	B.69.14.1	1XKSA	374	Nuclear pore complex protein Nup133
171	B.70.3.1	1N1MA1	470	Dipeptidyl peptidase IV/CD26, N-terminal domain
172	B.74.1.1	1HUGA	256	Carbonic anhydrase
173	B.76.2.1	1H3IA1	142	Histone H3 K4-specific methyltransferase
174	B.82.1.4	1EY2A	419	Homogentisate dioxygenase
175	B.82.1.19	2IC1A1	185	Cysteine dioxygenase type I
176	B.82.2.6	1H2KA	332	Hypoxia-inducible factor inhibitor
177	B.82.2.10	2IUWA1	202	(A: 70 - 279) AlkB homolog 3
178	B.85.7.1	2F69A2	166	Histone H3 K4-specific methyltransferase
179	B.88.1.1	1FWQA	115	RabGEF Mss4
180	B.88.1.2	1YZ1A1	144	Translationally controlled tumor protein
181	B.95.1.1	2AF9A1	163	Ganglioside M2 activator
182	B.98.1.1	1GW6A2	208	Leukotriene A4 hydrolase N-terminal domain
183	B.122.1.12	3BI7A1	192	(A: 414 - 617) E3 ubiquitin-protein ligase

## Continued

184	B.131.1.1	1JEQA3	285	(A: 254 - 538) Ku70 subunit middle domain
185	B.131.1.2	1JQBI	298	(B: 242 - 542) Ku80 subunit middle domain
186	B.131.1.3	1OW1A	167	SMART/HDAC1 associated repressor protein
187	B.145.1.1	1OA8A	128	Ataxin-1
yeast				
1	B.1.8.1	1B4LA	153	Cu, Zn superoxide dismutase
2	B.1.8.1	1QUPB1	149	Copper chaperone for superoxide dismutase
3	B.1.22.1	2HUEA1	163	Anti-silencing protein 1
4	B.2.8.1	1M2OA2	158	(A: 2 - 44, A: 391 - 523) Sec23
5	B.2.8.1	1M2VB2	177	(B: 61 - 215, B: 553 - 646) Sec24
6	B.17.1.1	1WPXB1	204	(B: 501 - 719) Carboxypeptidase Y inhibitor
7	B.18.1.9	1GQPA	180	Anaphase-promoting complex APC10/DOC1 subunit
8	B.18.1.20	1R64A1	141	(A: 461 - 601) Kexin, C-terminal domain
9	B.18.1.22	1O59A1	177	(A: 0 - 187) Allantoicase
10	B.26.1.2	1G3GA	164	Phosphotyrosine binding domain of Rad53
11	B.26.1.2	1QU5A	182	Phosphotyrosine binding domain of Rad53
12	B.29.1.13	2A6VA1	218	Emp46p N-terminal domain
13	B.30.5.4	1Z45A1	342	(A: 358 - 699) Galactose mutarotase
14	B.38.4.1	2VGMA1	117	Dom34
15	B.40.4.3	1KXLA	187	CDC13 ssDNA-binding domain
16	B.40.4.5	2VNUD3	103	(D: 252 - 399) Exosome complex exonuclease RRP44
17	B.40.4.16	2VNUD4	412	(D: 495 - 910) Exosome complex exonuclease RRP44
18	B.40.5.1	1E6AA	284	Inorganic pyrophosphatase
19	B.43.3.1	1N0UA1	138	(A: 344 - 481) Elongation factor 2, domain II
20	B.44.1.1	1G7CA2	109	Elongation factor eEF-1 $\alpha$ , C-terminal domain
21	B.45.1.1	1CI0A	205	Pyridoxine 5'-phosphate oxidase
22	B.50.1.2	1DP5A	329	Acid protease
23	B.55.1.7	1Q67B	146	Dcp1
24	B.55.1.9	1Y5OA1	114	RNA polymerase II transcription factor B 73KDa
25	B.55.1.10	2GCLA1	230	FACT complex subunit POB <sub>3</sub> , middle domain
26	B.62.1.1	1ISTA	161	Cyclophilin (eukaryotic)
27	B.69.4.1	1SQ9A	378	Antiviral protein Ski8 (Ski8p)
28	B.69.4.1	1PGUA1	325	Actin interacting protein 1
29	B.69.4.1	1NEXB2	344	(B: 370 - 744) Cdc4 propeller domain
30	B.69.4.2	1YFQA	342	Cell cycle arrest protein BUB3

**Continued**

31	B.69.14.1	1XIPA	367	Nucleoporin NUP159
32	B.80.5.1	1K4ZA	157	Adenylylcyclase associated protein, C-terminal domain
33	B.82.1.16	1XE7A	186	Hypothetical protein YML079W
34	B.82.1.20	1ZVFA1	175	3-hydroxyanthranilate-3,4-dioxygenase
35	B.86.1.2	1GPPA	217	VMA1-derived endonuclease PI-Scei intein

Scientific Research Publishing (SCIRP) is one of the largest Open Access journal publishers. It is currently publishing more than 200 open access, online, peer-reviewed journals covering a wide range of academic disciplines. SCIRP serves the worldwide academic communities and contributes to the progress and application of science with its publication.

Other selected journals from SCIRP are listed as below. Submit your manuscript to us via either [submit@scirp.org](mailto:submit@scirp.org) or [Online Submission Portal](#).

