Adaptation of model proteins from cold to hot environments involves continuous and small adjustments of average parameters related to amino acid composition

Emmanuele de Vendittis, Immacolata Castellano, Roberta Cotugno, Maria Rosaria Ruocco, Gennaro Raimo, Mariorosario Masullo

www.elsevier.com/locate/yjtbi

Adaptation of model proteins from cold to hot environments involves continuous and small adjustments of average parameters related to amino acid composition

Emmanuele De Vendittis, Immacolata Castellano, Roberta Cotugno, Maria Rosaria Ruocco, Gennaro Raimo, Mariorosario Masullo

PII: S0022-5193(07)00424-9
DOI: doi:10.1016/j.jtbi.2007.09.006
Reference: YJTBI 4844

To appear in: *Journal of Theoretical Biology*

Received date: 16 May 2007
Revised date: 29 August 2007
Accepted date: 1 September 2007

Cite this article as: Emmanuele De Vendittis, Immacolata Castellano, Roberta Cotugno, Maria Rosaria Ruocco, Gennaro Raimo and Mariorosario Masullo, Adaptation of model proteins from cold to hot environments involves continuous and small adjustments of average parameters related to amino acid composition, *Journal of Theoretical Biology* (2007), doi:10.1016/j.jtbi.2007.09.006

Adaptation of model proteins from cold to hot environments involves continuous and small adjustments of average parameters related to amino acid composition

Emmanuele DE VENDITTIS[a*], Immacolata CASTELLANO[a], Roberta COTUGNO[a], Maria Rosaria RUOCCO[a], Gennaro RAIMO[b], Mariorosario MASULLO[a,c]

[a]*Dipartimento di Biochimica e Biotecnologie Mediche, Università di Napoli Federico II, Via S. Pansini, 5, I-80131 Napoli, Italy*

[b]*Dipartimento di Scienze e Tecnologie dell'Ambiente e del Territorio, Università del Molise, Contrada Fonte Lappone, I-86090, Pesche (IS), Italy*

[c]*Dipartimento di Scienze Farmacobiologiche, Università di Catanzaro "Magna Graecia" I-88021 Roccelletta di Borgia (CZ), Italy*

*Corresponding author. Tel: +39 081 7463118; Fax: +39 081 7463653. *E-mail address:* devendittis@dbbm.unina.it (E. De Vendittis)

Abstract

The growth temperature adaptation of six model proteins has been studied in forty-two microorganisms belonging to eubacterial and archaeal kingdoms, covering optimum growth temperatures from 7 to 103°C. The selected proteins include three elongation factors involved in translation, the enzymes glyceraldehyde-3-phosphate dehydrogenase and superoxide dismutase, the cell division protein FtsZ. The common strategy of protein adaptation from cold to hot environments implies the occurrence of small changes in the amino acid composition, without altering the overall structure of the macromolecule. These continuous adjustments were investigated through parameters related to the amino acid composition of each protein. The average value per residue of mass, volume and accessible surface area allowed an evaluation of the usage of bulky residues, whereas the average hydrophobicity reflected that of hydrophobic residues. The specific proportion of bulky and hydrophobic residues in each protein almost linearly increased with the temperature of the host microorganism. This finding agrees with the structural and functional properties exhibited by proteins in differently adapted sources, thus explaining the great compactness or the high flexibility exhibited by (hyper)thermophilic or psychrophilic proteins, respectively. Indeed, heat-adapted proteins incline toward the usage of heavier-size and more hydrophobic residues with respect to mesophiles, whereas the cold-adapted macromolecules show the opposite behavior with a certain preference for smaller-size and less hydrophobic residues. An investigation on the different increase of bulky residues along with the growth temperature observed in the six model proteins suggests the relevance of the possible different role and/or structure organization played by protein domains. The significance of the linear correlations between growth temperature and parameters related to the

amino acid composition improved when the analysis was collectively carried out on all model proteins.

Key words: Psychrophiles; Hyperthermophiles; Protein temperature adaptation; Average hydrophobicity; Average amino acid size

1. Introduction

Biodiversity in the Earth's biosphere includes a large proportion of organisms called extremophiles, having colonized extreme environments (Nisbet and Sleep, 2001; Cockell and Stokes, 2004). Cold and hot temperatures are hostile habitats for life, and organisms having an optimum growth under the most extreme temperature conditions are named psychrophiles and hyperthermophiles, respectively. Properties of these microorganisms belonging to eubacterial or archaeal kingdom have been extensively reviewed (Stetter, 1996; Stetter, 1998; Stetter, 1999; Hicks and Kelly, 1999; Huber et al., 2000; Deming, 2002; D'Amico et al, 2006; Cavicchioli, 2006). Indeed, it is known that some psychrophiles sustain a residual biological activity even at -20°C (Deming, 2002), whereas some hyperthermophiles are able to grow up to 113°C (Blöchl et al, 1997). The increasing discovery and characterization of this type of extremophiles, and the possibility to compare the properties of their biomolecules to that of mesophiles growing at 'usual' temperatures, offer the opportunity to study the molecular basis of life adaptation under a wide range of growth temperature.

The main targets in the environmental adaptation of extremophilic sources are proteins, the most abundant flexible macromolecules involved in the control of the whole metabolic pathways and in the structural organization of the microorganism. Several reviews summarized the properties of proteins isolated from thermophiles (Jaenicke and Zavodsky, 1990; Jaenicke, 1991; Adams, 1993; Vieille et al, 1996; Jaenicke and Böhm, 1998; Hicks et al. 1999; Niehaus et al., 1999; Vieille and Zeikus, 2001; Sterner and Liebl, 2001) and psychrophiles (Feller et al., 1997; Feller and Gerday, 1997; Gerday et al. 1997; Sanders et al., 2003; Georlette et al. 2004; Siddiqui and Cavicchioli, 2006). Comprehensive studies on crystal structures of thermophilic proteins did not reveal unusual conformations specific to the source type (Petukhov et

al., 1997; Facchiano et al., 1998; Karshikoff and Ladenstein, 1998; Szilagyi and Zavodszky, 2000; Kumar et al., 2000). A similar behavior is observed with psychrophilic proteins (Russell et al., 1998; Maes et al., 1999; Violot et al., 2005), even though the number of crystallographic structures available in this case is much lower. Therefore, it seems that in these extremophilic sources the overall structure of a protein is very similar to that possessed by the mesophilic counterpart, thus reflecting the adaptation of the specific function of the macromolecule, rather than a tolerance to the living environment in the host source. However, hyperthermophilic proteins are endowed with an extraordinary heat stability, as a consequence of a more tight compactness of the protein structure (Vieille and Zeikus, 2001). Vice versa, the psychrophilic counterparts possess an increased protein flexibility, which in most cases leads to a decreased stability compared to mesophilic proteins (D'Amico et al., 2006); in some psychrophilic enzymes the protein flexibility is enhanced in localized regions of the protein structure (Fields and Somero, 1998). Temperature adaptation of proteins is mostly relevant for the catalytic properties of the enzymes, as they must adapt the rate of the catalyzed reaction to the growth temperature of the organism. Indeed, thermophilicity studies revealed that psychrophiles synthesize cold-adapted enzymes endowed with a specific activity at low temperatures, significantly higher compared to that possessed by the mesophilic counterparts (Georlette et al., 2004; Siddiqui and Cavicchioli, 2006). On the other hand, the specific activity of hyperthermophilic enzymes reaches its optimum only at high temperatures, close to the optimum growth conditions of the source (Vieille et al., 1996; Vieille and Zeikus, 2001). Therefore, temperature adaptation of proteins reflects a multifactorial equilibrium between counteracting forces affecting flexibility, stability and activity of proteins. In particular, the similarity of the protein structure and the occurrence of a common catalytic

mechanism in proteins isolated from sources adapted from cold to hot environments indicate that the challenge to the extreme environments has been likely accomplished by a fine modulation of the amino acid composition of proteins aimed at optimizing the number of specific weak interactions inside the protein core. Indeed, the amino acid composition has been found to play an important role in determining the protein structural class (Chou and Zhang, 1994; Chou and Zhang, 1995; Chou, 1995; Chou and Maggiora, 1998), in identifying protein subcellular localization, and many other attributes (Chou and Elrod, 1999; Chou, 2002). Evidence has been presented that the amino acid composition in the proteomes of *Escherichia coli* and *Bacillus subtilis* reflects a natural selection to enhance metabolic efficiency in these microorganisms (Akashi and Gojobori, 2002). Furthermore, the analysis on several *Saccharomyces cerevisiae* genes showed a correlation between gene expression level and amino acid composition (Raghava and Han, 2005).

In order to understand the structural requirements for protein adaptation to heat or cold, the amino acid composition of proteins isolated from (hyper)thermophiles (Jaenicke, 1991; Jaenicke and Böhm, 1998; Vieille and Zeikus, 2001; Sterner and Liebl, 2001) or psychrophiles (Feller et al., 1997; Feller and Gerday, 1997; Gerday et al. 1997; Sanders et al., 2003) has long been compared with that of mesophiles. Some of these investigations have been focused on the whole genome of extremophilic microorganisms, thus considering the total protein content of the selected microbial sources. The frequencies of each amino acid residue have been derived from an average amino acid composition, in order to discover a possible bias in the amino acid usage of the considered extremophile. However, each microbial source and each protein seems to adopt only a few of possible, and even counteracting, structural trends (D'Amico et al., 2006; Sterner and Liebl, 2001). For instance, the amino acid bias discovered in some

psychrophiles (Sanders et al., 2003) is not applicable to similarly adapted, but evolutionary distant sources (Rabus et al., 2004; Medigue et al., 2005). The ambiguous results are probably related to the different content and/or representation of proteins analyzed for each microbial source. Furthermore, the genetic drift or the natural selection between different microorganisms could hide critical amino acid changes for thermal adaptation (Siddiqui and Cavicchioli, 2006). Some specific key contributions for thermal adaptation of proteins have been proposed through the comparison of the structural and functional properties of proteins in differently adapted sources or through the effect of a mutagenic analysis of a target protein on its thermal stability. These studies led to the discovery of a number of different basic mechanisms involved in thermal stability of thermophilic proteins, as previously reviewed (Vieille and Zeikus, 2001; Sterner and Liebl, 2001). For instance, surface loop depletion, an increased occurrence of hydrophobic residues with branched side chains, and an enhanced proportion of charged residues are apparently the most consistent structural factors contributing to thermostability in thermophilic proteins (Kumar and Nussinov, 2001). Furthermore, on the basis of thermodynamic differences among homologous thermophilic and mesophilic proteins, the higher stability possessed by thermophilic proteins is probably due to specific interactions, particularly electrostatic, present into the protein (Kumar et al, 2001). Finally, the unusual thermal stability of an ATP-binding cassette ATPase of mesophilic origin could be predicted on the basis of its content of polar amino acid residues (Sarin et al., 2003). However, it is not rare that the most important factor for the thermostability of a given protein is not applicable to explain the heat stability of a different one.

This article addresses the question of a possible continuum in the strategy of protein adaptation to the different growth temperatures of the host source. For this

reason the amino acid composition of model proteins has been analyzed in several microbial sources displaying an optimum growth temperature ranging from 7 to 103°C. In particular, we have analyzed the temperature dependence of average parameters related to the amino acid composition. The data obtained suggest that the average values per residue of mass, hydrophobicity, volume and accessible surface, linearly increase with the optimum growth temperature of the microbial source. This finding implies a small variation of the amino acid composition, leading to a moderate bias in the amino acid usage, depending on the growth temperature of the source. Indeed, in (hyper)thermophilic model proteins the content of heavier-size and more hydrophobic residues is increased with respect to mesophilic counterparts; vice versa, smaller-size and less hydrophobic residues are preferred in psychrophilic proteins.

2. Materials and Methods

2.1. Microbial sources

The forty-two microorganisms considered in this study have been chosen for their different adaptation to growth temperature (Table 1). They belong to the living domains of eubacteria (25 sources) and archaea (17 sources), whose complete sequenced genome is available on line, except for *Bacillus stearothermophilus* and *Pyrococcus woesei*. The selected microbial sources, whose respective optimum growth temperatures are indicated in Table 1, include psychrophiles, mesophiles, thermophiles and hyperthermophiles, thus allowing the analysis over a wide range of temperature adaptation, from 7°C of *Desulfotalea psychrophila* to 103°C of *Pyrococcus abyssi*.

2.2. Selection of model proteins

Six ubiquitous proteins have been selected as models for the analysis of the adaptation of their amino acid composition to the different growth temperatures. The list includes: three elongation factors involved in protein synthesis translation, namely the elongation factor Tu/1→(EF-Tu in eubacteria or EF-1→in archaea), the elongation factor G/2 (EF-G in eubacteria or EF-2 in archaea), and the elongation factor Ts/1→(EF-Ts in eubacteria or EF-1→in archaea); the enzymes glyceraldehyde-3-phosphate dehydrogenase (GAPDH) and superoxide dismutase (SOD); the cell division protein FtsZ (FtsZ). This choice allows covering of various aspects of the functional role played by proteins in the adaptation of metabolic pathways and structural organization of the microorganism to the growth temperature. Indeed, EF-Tu/1→and EF-G/2 are monomeric multifunctional and flexible proteins, interacting with small and large molecules in the fulfillment of their biological functions (Klink, 1985). They also display a GTPase activity, required to induce conformational changes of the protein structure, essential for their specific functions (Parmeggiani and Sander, 1981). The other translation factor EF-Ts/1→is more rigid and it is organized either as a monomer or as a homodimer (Raimo et al., 1996). It acts as an exchange factor of the guanine nucleotide bound to EF-Tu/1→, upon its binding to the specific target (Parmeggiani and Sander, 1981). GAPDH is a NAD-dependent enzyme organized as a homotetramer, whose activity is mainly involved in the glycolytic pathway (Sirover, 1999). SOD is a key enzyme in the cellular defense against oxidative stress conditions. It is organized either as a homodimer or as a homotetramer, and acts as a scavenger of the toxic superoxide anions formed during oxidative metabolism (Miller, 2004). FtsZ, a key component of the prokaryotic cytoskeleton homologous to eukaryotic tubulins, is organized as a homodimer and plays a central role in cell division. It has a GTPase

activity, required for its assemblage into the Z ring on the inner face of the cytoplasmic membrane, marking the future cell division site (Margolin, 2005).

2.3. Parameters related to the amino acid composition in proteins and their structural domains

The amino acid sequence of the selected proteins from each microbial source was downloaded on line (www.ncbi.nlm.nih.gov or www.expasy.ch). The list of 255 amino acid sequences analyzed (see Supplementary Material, Tables S1, S2, S3, S4, S5, S6) includes redundant isoforms of the model proteins in some sources. On the other hand, a few microorganisms miss one or two of the selected model proteins. The initial methionine was always included, regardless of its presence in the mature protein. The amino acid composition of the downloaded proteins has been analyzed to obtain average structural parameters. In particular, the mean values per residue of amino acid mass, hydrophobicity, volume and accessible surface have been calculated according to the following equations:

$$\text{Average mass} = \frac{\sum_{i=1}^{i=20}(\text{mass}_{aai} \bullet n_{aai}) + 18.015}{N}$$

$$\text{Average hydrophobicity} = \frac{\sum_{i=1}^{i=20}(\text{hydrophobicity}_{aai} \bullet n_{aai})}{N}$$

$$\text{Average volume} = \frac{\sum_{i=1}^{i=20}(\text{volume}_{aai} \bullet n_{aai})}{N}$$

$$\text{Average accessible surface area} = \frac{\sum_{i=1}^{i=20}(\text{surface area}_{aai} \bullet n_{aai})}{N}$$

where $mass_{aai}$, $hydrophobicity_{aai}$, $volume_{aai}$, and $surface\ area_{aai}$ are the values of these parameters referred to each amino acid residue (Table 2), n_{aai} is the content of each residue in the protein, and N the total number of residues in the polypeptide chain. Finally, in the first equation 18.015 represents the mass of a water molecule.

EF-Tu/1→and SOD are organized in three and two structural domains, respectively. A multiple alignment of the analyzed amino acid sequences of these proteins was obtained with CLUSTALW program available on-line (www.ebi.ac.uk/clustalw), and used to identify in each primary structure the respective domains. The crystal structures of EF-1→(Vitagliano et al., 2001) and SOD (Ursby et al., 1999) from the hyperthermophilic archaeon *Sulfolobus solfataricus* were used as models for the definition of domains. In particular, peptides $M_1 \rightarrow L_{224}$, $P_{225} \rightarrow G_{314}$, and $H_{315} \rightarrow K_{435}$ have been considered as constituting the domains G, M, and C of EF-1→, respectively. Concerning SOD, peptides $M_1 \rightarrow G_{98}$, and $G_{99} \rightarrow K_{211}$ represent the N- and C-domain, respectively. The identified domains were considered as small single proteins and analyzed for their mean values per residue of amino acid mass, hydrophobicity, volume and accessible surface, using the equations previously indicated.

2.4. Significance level of the temperature dependence of parameters related to the amino acid composition of proteins

For each model protein the average values per residue of mass, hydrophobicity, volume and accessible surface area were plotted versus the optimum growth temperature of the microbial source in independent analyses corresponding to each selected model protein. The dependence of the average parameter on the growth temperature was evaluated as a linear curve fit obtained with the least squares method

and the significance of the correlation was estimated from the correlation coefficient *r*. In a perfect linear correlation, r^2 approaches the value of 1.000, whereas an r^2 approaching to zero indicates the lack of any correlation in the linear regression. The significance test included the calculation of the *t*-parameter according to the following equation:

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

where *n* represents the number of pairs of scores in a two two-tailed test; therefore $n - 2$ indicates the degrees of freedom. The significance level of the linear correlation was estimated by *p*; for example, a *p* value < 0.001 indicates that a chance occurrence in the correlation is lower than 1 out of 1000.

3. Results

3.1. Correlation between growth temperature and average parameters related to the amino acid composition in six model proteins from different host microorganisms

Four average parameters related to the amino acid composition of six model proteins were considered to evaluate their dependences on the growth temperature of forty-two different micro-organisms. In particular, we have chosen the average mass, volume and accessible surface area per residue, because altogether they allow an evaluation of the usage of bulky residues in the amino acid composition. Another parameter considered was the average hydrophobicity, as it reflects the content of hydrophobic residues in the protein.

3.1.1. Average mass per residue

Independent plots related to the different model proteins show that the average mass per residue obtained for each protein slightly increases with the optimum growth temperature of the host micro-organism (Fig. 1). In all plots the scattered data apparently fit to a linear regression; equations and corresponding correlation parameters of these linear fits are reported in Table 3. The analysis of the significance test indicates that all linear regressions are significant, because *t* values range between 5.42 and 8.44, and *p* values are all lower than 0.001. In the equations reported in Table 3 the slope *b* represents how the average amino acid mass of the model protein increases per degree centigrade. These figures are different among the six model proteins; for instance, EF-Tu/1→and EF-G/2 possess a small *b* value whereas the other proteins have a significantly greater *b*. Furthermore, Table 3 reports the 'ideal' average mass per residue at 37 °C, as calculated from the corresponding linear equation of each protein. The data range from 105.6 Da of FtsZ to 112.2 Da of SOD, a finding suggesting that the preferential usage of heavier- or smaller-size residues is likely related to the specific functions played by each protein. Indeed, the interval comprising the average mass values is slightly different in the six model proteins, the highest range being related to SOD (Fig. 1D), and the lowest one to FtsZ (Fig. 1F). All these features indicate that each protein has a specific proportion of smaller and heavier residues, even though this proportion apparently changes with the temperature of the host microorganism. In particular, protein adaptation to growth temperature of the microbial source likely implies the increase of average mass of amino acid residue with temperature.

The different *b* slopes found in the model proteins probably reflect the different contribution to temperature adaptation of specific regions of the protein. For this reason, the analysis of the average mass per residue has been also performed on the structural domains of two model proteins, namely EF-Tu/EF-1→and SOD, displaying a low and

high *b* value, respectively. The plots corresponding to the EF-Tu/EF-1→domains (Fig. 2) show that the effect of temperature on the average mass per residue is different among its three domains. Indeed, the average amino acid mass consistently increases with temperature only for the catalytic G-domain (Fig. 2A), because the effect of temperature on this parameter becomes almost imperceptible for the M-domain (Fig. 2B) or the C-domain (Fig. 2C). The small decrease observed for M-domain (Fig. 2B), as well as the small increase detected for the C-domain, is very modest and could reflect randomly-scattered variations of the parameter in the overall interval of temperature. This hypothesis seems supported by the analysis of equations and corresponding correlation coefficients, which suggest an undetectable variation of the average mass along with temperature and a lower significance test for M- and C-domains (Table 3). Vice versa, the data referred to the G-domain are more significant and indicate that the temperature-dependent increase of the average amino acid mass, evaluated through the *b* slope, is almost double with respect to the entire EF-Tu/EF-1→ It is interesting that this *b* value becomes more similar to values determined for proteins with a high *b* slope (Table 3). Therefore, the low *b* slope found for the entire EF-Tu/EF-1→is explained with the fact that only the catalytic G-domain undergoes a significant increase of this parameter with temperature, whereas the M- or C-domain is apparently unaffected by temperature. The same analysis of the effect of temperature on the average amino acid mass in the two SOD domains is reported in Fig. 3. Both the N-domain (Fig. 3A) and the C-domain (Fig. 3B) undergo a consistent increase of the average mass of their residues along with temperature. The data appear significant and the calculated *b* slopes are similar to the corresponding value calculated for the entire SOD (Table 3). These findings suggest that SOD domains undergo a similar increase of the average mass of

their residues with temperature, a finding probably related to a similar involvement of both domains in the catalytic activity of the enzyme.

3.1.2. Average hydrophobicity per residue

Independent plots reporting the effect of temperature on the mean value per residue of hydrophobicity in the six model proteins are shown in Fig. 4. Also in this case the average hydrophobicity apparently increases with the optimum growth temperature of the host microbial source, and the figures related to the linear regression of the data are presented in Table 4. The significance level is consistently higher compared to the previous analysis on the average mass, because the *t* values range between 6.03 and 11.26. No great differences emerge within *b* slopes of the equations, thus indicating a similar increase of the average hydrophobicity per degree centigrade in all model proteins considered. Vice versa, the 'ideal' hydrophobicity of amino acid residue at 37 °C ranges from 4.24 kJ•mol^{-1} of FtsZ to 4.83 kJ•mol^{-1} of SOD. Indeed, the usage of hydrophobic residues slightly varies with temperature among different proteins, even though SOD (Fig. 4D) and FtsZ (Fig. 4F) contain the highest and lowest proportion of these residues, respectively. These features indicate that each model protein has its own proportion of hydrophobic residues, even though this parameter increases with the temperature to adapt the protein at the growing conditions of the host micro-organism.

The analysis of the temperature dependence of average hydrophobicity per residue has been performed also on the domains of EF-Tu/EF-1→and SOD. As shown in Table 4, the hydrophobicities of the three domains of EF-Tu/EF-1→increase with temperature, with slopes almost coincident with that of the intact protein; a similar behavior occurs also for the two SOD domains.

3.1.3. Average volume and average accessible surface area per residue

The other two parameters of the amino acid composition of model proteins, i.e. average volume and average accessible surface area per residue, are related to the size and steric occupancy of the amino acids. The independent plots showing the effect of growth temperature on the average values per residue of volume and average accessible surface area per residue are reported in Fig. 5 and Fig. 6, respectively; corresponding equations and correlation parameters of the linear fits are presented in Tables 5 and 6, respectively. The significance level of the linear regressions is even improved for all model proteins with respect to the analysis of the average mass per residue. The temperature-dependent behavior of average volume and accessible surface area is similar to that obtained from the average mass per residue. Indeed, these parameters related to the amino acid composition slightly increase with the optimum growth temperature of the host micro-organism. As in the case of the temperature-dependence of the average mass per residue, EF-Tu/1→and EF-G/EF-2 display smaller *b* values, whereas the other model proteins adopt a greater *b*. Furthermore, the highest values of average volume and accessible surface area are related to SOD (Figs. 5D and 6D), and the lowest ones to FtsZ (Figs. 5F and 6F). Therefore, the analysis on the mean values of volume and accessible surface area per residue confirm that the proportion of bulky residues, displaying higher volumes and accessible surface areas, varies among different model proteins. Nevertheless, protein adaptation to the increasing growth temperature implies the usage of bulkier amino acid residues.

3.2. Effect of the growth temperature on the parameters deduced from the average amino acid composition collectively obtained from the six model proteins

The previous results indicate that the six model proteins examined in this study have specific amino acid compositions, with proportions of bulky and hydrophobic residues probably covering the range of different usage of these residues in protein formation. For this reason, the analysis of the average amino acid composition obtained for each microbial source from the collective content of each residue in the six model proteins could be considered as representative of the amino acid usage in that source. This average amino acid composition was used to evaluate the effect of growth temperature on the parameters chosen in this study. The four plots related to the mean values of mass, hydrophobicity, volume and accessible surface area per residue are shown in Fig. 7, and the significance of the linear regressions are presented in Table 7. All the selected parameters linearly increase with the growth temperature of the microorganism. Moreover, the significance parameters consistently improve with respect to the values obtained in the analysis on single proteins. All these features confirm that the increased usage of bulky and hydrophobic residues seems a common adaptive response of protein composition in microbial sources upon the enhancement of the growth temperature.

4. Discussion

Protein adaptation to the growth temperature of host micro-organisms should imply the occurrence of small adjustments in the amino acid composition of the macromolecule, that equilibrate the required number of weak interactions, without altering the overall structure of the protein. The bias towards the usage of selected amino acid residues found in some extremophiles reflects the genetic drift of the organism rather than an adaptation to extreme environments. For this reason, the average parameters related to the amino acid composition chosen in this study appear

appropriate tools for a better understanding of protein adaptation to the growth temperature, as they mediate the differences between taxonomically distant organisms. Moreover, the analysis on six ubiquitous model proteins from several microorganisms, adapted to a wide range of growth temperature, allows the discovery of common features in the strategy of protein adaptation. The small but continuous enhancement of the average values per residue of amino acid mass, hydrophobicity, volume and accessible surface area, in a fairly good correlation with the increasing growth temperature, is in agreement with the structural and functional properties of these macromolecules. Indeed, each protein seems to adapt its amino acid composition to a determined 'container', with shape and steric hindrance aimed at a specific role played in the cell from cold– to hot–adapted sources. Among the chosen parameters, average mass, volume and accessible surface area per residue are all correlated to the proportion of bulky residues. Indeed, the results obtained from these tools are discussed altogether, as they are very similar. On the other hand, the average hydrophobicity per residue reflects the proportion of hydrophobic residues in the protein, and for this reason it is discussed separately.

The increased average mass, volume and accessible surface area per residue found in (hyper)thermophilic proteins reflects the high compactness of these macromolecules with respect to mesophilic ones, because the 'container' is filled with an increased number of bulky residues that reduce the number of cavities inside the protein core, and consequently its flexibility. These features enhance the thermostability of the macromolecule, but slow down its catalytic rate at 'normal' temperatures; indeed, the best catalytic efficiency of thermozymes is usually reached at the optimum growth temperature of the host micro-organism. Furthermore, (hyper)thermophilic proteins seem to have a greater mass in their hydrodynamic volume compared to mesophilic

ones, as demonstrated by their behavior on gel–filtration. In fact, when calibration of a gel–filtration column is made with mesophilic proteins, the hydrodynamic volume of a hyperthermophilic protein is lower than that expected on the basis of its known molecular size. Indeed, under these experimental conditions, the homotetrameric SOD from the hyperthermophile *Sulfolobus solfataricus* elutes as a protein with M_r 65,400, instead of 96,400, as expected from its amino acid sequence (Ursby et al., 1999). Vice versa, when calibration is made with hyperthermophilic protein standards, the hydrodynamic volume of this enzyme corresponds to M_r 89,000, much closer to that expected. This behaviour has been confirmed on other hyperthermophilic proteins (Raimo et al., 1996; Ruocco et al, 2004). Therefore, the greater mass of a hyperthermophilic protein is likely "contained" in a hydrodynamic volume similar to the mesophilic counterpart. For this reason we suggest that, based on gel–filtration experiments, the "density" of a hyperthermophilic protein is greater compared to its mesophilic counterpart.

Concerning psychrophilic proteins, the average values of mass, volume and accessible surface area are reduced compared to mesophiles, even though the differences are less evident because of the lower gap between cold and 'normal' temperature. Nevertheless, the reduced proportion of bulky residues in the 'container' improves the flexibility of psychrophilic proteins, because of an increased number of cavities. This leads to a decreased thermostability, but improves the catalytic rate of the cold-adapted enzymes at 'normal' temperatures, with respect to mesophiles. Usually, psychrophilic enzymes possess a sufficiently high catalysis at their optimum growth temperatures, even though their thermophilicities improve far beyond in some cases (D'Amico et al., 2006). Moreover, psychrophilic proteins have a slightly lower "density" compared to mesophilic ones, as revealed through gel-filtration experiments.

Indeed, when a gel–filtration column is calibrated with mesophilic proteins, the homodimeric SOD from the psychrophile *Pseudoalteromonas haloplanktis* elutes as a protein with M_r 46,000, instead of 42,500, as expected from its amino acid sequence (Castellano et al., 2006). The faster elution on gel-filtration was not limited to SOD, as it was described also for another psychrophilic protein (Masullo et al., 2000).

Among the six ubiquitous proteins, some proteins display a low *b* slope, whereas others possess a higher *b* slope, in the equations representing the linear increase with temperature of average mass, volume and accessible surface area. The behavior of EF-Tu/1→, possessing a low *b* slope, was explained with the fact that only domain G displays an increase of these parameters similar to those found in the other group. Domain G represents a relevant region for the catalytic properties of the molecule, as it contains the active site for GTP hydrolysis, whereas the other two domains have regulatory functions. Therefore, in this case the protein region involved in catalysis adopts the same increase of bulky residues of other proteins, to adapt its catalysis at the growth temperature of the host source. Other possible explanations could reside in the different domain organization; indeed, differently from domain G, domains M and C of EF-Tu/1→almost lack →helices. Much more information is required to better address this point; however, it is noteworthy that the two domains of SOD, both involved in catalysis and rich in →helices, display an almost similar increase in the proportion of bulky residues along with the growth temperature.

Concerning the average hydrophobicity per residue, (hyper)thermophilic and psychrophilic proteins show an opposite behavior. The first ones enhance the proportion of hydrophobic residues compared to mesophiles, and this finding agrees with the higher thermostability and thermophilicity displayed by thermozymes. In fact, more hydrophobic interactions likely take place in the protein core and this hinders the

flexibility of the macromolecule. On the other hand, in psychrophilic proteins the usage of hydrophobic residues is reduced compared to mesophiles, a finding that lowers the hydrophobic interactions and therefore improves the flexibility of the molecule. Among the six model proteins, no great differences were found in the enhancement of the average hydrophobicity along with the growth temperature, a finding that makes this parameter somehow different from those related to the usage of bulky residues. An interesting implication arising from the increased usage of hydrophobic residues with the growth temperature of the microbial source concerns the decreased strength of the hydrophobic interactions, occurring in a protein exposed at high temperatures. The higher average hydrophobicity in a (hyper)thermophilic model protein could be relevant to counteract the lower strength of hydrophobic interactions, taking place in the hot environment.

The correlation indexes obtained from the evaluation of temperature effect on the parameters related to the amino acid composition become more significant when the analysis is collectively carried out on the total amino acid composition derived from the six model proteins. The improvement of the significance of the linear correlation indicates that the average parameters appear appropriate tools to understand the effect of growth temperature on the amino acid composition in the whole interval of environmental adaptation.

The present study investigated how the classic amino acid composition of some selected proteins continuously adapted to the different growth temperatures of various host microorganisms. Therefore, the possible sequence-order adjustment of proteins during the heat/cold adapting process was lost. More insights could be derived from future studies, by incorporating the sequence-order adjustment information through the use of the so-called pseudo-amino acid composition (Chou, 2001; Chou, 2005). Indeed,

this information enhanced the success rates in predicting protein structural class (Shen el al, 2005; Shen et al, 2006), membrane type (Wang et al, 2004; Shen and Chou, 2005; Chou and Shen, 2007a), signal peptide (Chou and Shen, 2007b), and protein subcellular localization (Chou and Shen, 2007c; Chou and Shen, 2007d; Shen and Chen, 2007). Nevertheless, the results of the present investigation based on the classic amino acid composition could be helpful in protein engineering of enzymes with predefined properties of thermostability and catalytic efficiency.

Acknowledgments

This work was supported by a grants from the Ministero dell'Università, Istruzione e Ricerca (PRIN 2005, MIUR Italy).

References

Adams, M. W., 1993. Enzymes and proteins from organisms that grow near and above 100 degrees C. Annu. Rev. Microbiol. 47, 627–658.

Akashi, H., Gojobori, T., 2002. Metabolic efficiency and amino acid composition in the proteomes of *Escherichia coli* and *Bacillus subtilis*. Proc. Natl. Acad. Sci. U.S.A. 99, 3695–3700.

Auman, A. J., Breezee, J. L., Gosink, J. J., Kämpfer, P., Staley, J. T., 2006. *Psychromonas ingrahamii* sp. nov., a novel gas vacuolate, psychrophilic bacterium isolated from Arctic polar sea ice. Int. J. Syst. Evol. Microbiol. 56, 1001-1007.

Blöchl, E., Rachel, R., Burggraf, S., Hafenbradl, D., Jannasch, H. W., Stetter, K. O., 1997. *Pyrolobus fumarii*, gen. and sp. nov., represents a novel group of archaea, extending the upper temperature limit for life to 113 degrees C. Extremophiles 1, 14–21.

Castellano. I., Di Maro, A., Ruocco, M. R., Chambery, A., Parente, A., Di Martino, M. T., Parlato, G., Masullo, M., De Vendittis, E., 2006. Psychrophilic superoxide dismutase from *Pseudoalteromonas haloplanktis*: biochemical characterization and identification of a highly reactive cysteine residue. Biochimie 88, 1377–1389.

Cavicchioli, R., 2006. Cold–adapted archaea. Nat. Rev. Microbiol. 4, 331–343.

Chothia, C., 1976. The nature of the accessible and buried surfaces in proteins. J. Mol. Biol. 105, 1–12.

Chou, K. C., 1995. A novel approach to predicting protein structural classes in a (20-1)-D amino acid composition space. Proteins 21, 319–344.

Chou, K. C., 2001. Prediction of protein cellular attributes using pseudo-amino acid composition. Proteins 43, 246–255 *(Erratum: ibid., 2001, Proteins 44, 60)*.

Chou, K. C., 2002. In: Weinrer, P. W., Lu, Q. (Eds.), Gene Cloning & Expression Technologies, Chapter 4, Eaton Publishing, Westborough, MA, pp. 57–70.

Chou, K. C., 2005. Using amphiphilic pseudo amino acid composition to predict enzyme subfamily classes. Bioinformatics 21, 10–19.

Chou, K. C., Elrod, D. W., 1999. Protein subcellular location prediction. Protein Eng. 12, 107–118.

Chou, K. C., Maggiora, G. M., 1998. Domain structural class prediction. Protein Eng. 11, 523–538.

Chou, K. C., Shen, H. B., 2007a. MemType-2L: a web server for predicting membrane proteins and their types by incorporating evolution information through Pse-PSSM. Biochem. Biophys. Res. Commun. 360, 339–345.

Chou, K. C., Shen, H. B., 2007b. Signal-CF: a subsite-coupled and window-fusing approach for predicting signal peptides. Biochem. Biophys. Res. Commun. 357, 633–640.

Chou, K. C., Shen, H. B., 2007c. Large-scale plant protein subcellular location prediction. J. Cell. Biochem. 100, 665–678.

Chou, K. C., Shen, H. B., 2007d. Euk-mPLoc: a fusion classifier for large-scale eukaryotic protein subcellular location prediction by incorporating multiple sites. J. Proteome Res. 6, 1728–1734.

Chou, K. C., Zhang, C. T., 1994. Predicting protein folding types by distance functions that make allowances for amino acid interactions. J. Biol. Chem. 269, 22014–22020.

Chou, K. C., Zhang, C. T., 1995. Prediction of protein structural classes. Crit. Rev. Biochem. Mol. Biol. 30, 275–349.

Cockell, C. S., Stokes, M. D., 2004. Ecology: widespread colonization by polar hypoliths. Nature 431, 414.

Corsaro, M. M., Lanzetta, R., Parrilli, E., Parrilli, M., Tutino, M. L., Ummarino, S., 2004. Influence of growth temperature on lipid and phosphate contents of surface polysaccharides from the Antarctic bacterium *Pseudoalteromonas haloplanktis* TAC 125. J. Bacteriol. 186, 29–34.

D'Amico, S., Collins, T., Marx J. C, Feller, G., Gerday, C., 2006. Psychrophilic microorganisms: challenges for life. EMBO Rep. 7, 385–389.

Deming, J. W., 2002. Psychrophiles and polar regions. Curr. Opin. Microbiol. 5, 301–309.

Facchiano, A. M., Colonna, G., Ragone, R., 1998. Helix stabilizing factors and stabilization of thermophilic proteins: an X–ray based study. Protein Eng. 11, 753–760.

Feller, G., Gerday, C.,1997. Psychrophilic enzymes: molecular basis of cold adaptation. Cell. Mol. Life Sci. 53, 830–841.

Feller, G., Arpigny, J. L., Narinx, E., Gerday, C., 1997. Molecular adaptations of enzymes from psychrophilic organisms. Comp. Biochem. Physiol. 118A, 495–499.

Fields, P. A., Somero. G. N., 1998. Hot spots in cold adaptation: localized increases in conformational flexibility in lactate dehydrogenase A4 orthologs of Antarctic notothenioid fishes. Proc. Natl, Acad, Sci, USA 95, 11476–11481.

Georlette. D,, Blaise, V., Collins, T., D'Amico, S., Gratia, E., Hoyoux, A., Marx, J. C., Sonan, G., Feller, G., Gerday, C., 2004. Some like it cold: biocatalysis at low temperatures. FEMS Microbiol. Rev. 28, 25–42.

Gerday, C., Aittaleb, M., Arpigny, J. L., Baise, E., Chessa, J. P., Garsoux, G., Petrescu, I., Feller, G., 1997. Psychrophilic enzymes: a thermodynamic challenge. Biochim. Biophys. Acta 1342, 119–131.

Gosink, J. J., Woese, C. R., Staley, J. T., 1998. *Polaribacter* gen. nov., with three new species, *P. irgensii* sp. nov., *P. franzmannii* sp. nov. and *P. filamentus* sp. nov., gas vacuolate polar marine bacteria of the *Cytophaga-Fflavobacterium-Bacteroides* group and reclassification of *'Flectobacillus glomeratus'* as *Polaribacter glomeratus* comb. nov.. Int. J. Syst. Bateriol. 48, 223-235.

Hicks, P. M., Kelly, R. M., 1999. Thermophilic microorganisms. In: Flickinger, M. C., Drew, S. W. (Eds.), Encyclopedia of bioprocess technology: fermentation, biocatalysis, and bioseparation, John Wiley & Sons, New York, pp. 2536–2552.

Hicks, P. M., Adams, M. W., Kelly, R. M., 1999. Enzymes, extremely thermostable. In: Flickinger, M. C., Drew, S. W. (Eds.), Encyclopedia of bioprocess technology: fermentation, biocatalysis, and bioseparation, John Wiley & Sons, New York, pp. 987–1004.

Huber, R., Huber, H., Stetter, K. O., 2000. Towards the ecology of hyperthermophiles: biotopes, new isolation strategies and novel metabolic properties. FEMS Microbiol. Rev. 24, 615–623.

Jaenicke, R., 1991. Protein stability and molecular adaptation to extreme conditions. Eur. J. Biochem. 202, 715–728.

Jaenicke, R., Böhm, G., 1998. The stability of proteins in extreme environments. Curr. Opin. Struct. Biol. 8, 738–748.

Jaenicke, R., Zavodsky, P, 1990. Proteins under extreme physical conditions. FEBS Lett. 268, 344–349.

Karshikoff, A., Ladenstein, R., 1998. Proteins from thermophilic and mesophilic organisms essentially do not differ in packing. Protein Eng. 11, 867–872.

Klink, F., 1985. Elongation factors. In: Woese, C.R., Wolfe, R.S. (Eds.), The bacteria, vol VIII Archaebacteria. Academic Press, Orlando, pp. 379–407.

Kumar, S., Nussinov, R., 2001. How do thermophilic proteins deal with heat? Cell. Mol. Life Sci. 58, 1216–1233.

Kumar. S., Tsai, C. J., Nussinov, R., 2000. Factors enhancing protein thermostability. Protein Eng. 13, 179–191.

Kumar, S., Tsai, C. J., Nussinov, R., 2001. Thermodynamic differences among homologous thermophilic and mesophilic proteins. Biochemistry 40, 14152–14165.

Maes. D., Zeelen, J. P., Thanki, N., Beaucamp, N., Alvarez, M., Thi, M. H., Backmann, J., Martial, J. A., Wyns, L., Jaenicke, R., Wierenga, R. K., 1999. The crystal structure of triosephosphate isomerase (TIM) from *Thermotoga maritima*: a comparative thermostability structural analysis of ten different TIM structures. Proteins 37, 441–453.

Margolin, W., 2005. FtsZ and the division of prokaryotic cells and organelles. Nat. Rev. Mol. Cell. Biol. 6, 862–871.

Masullo, M., Arcari, P., de Paola, B., Parmeggiani, A., Bocchini, V., 2000. Psychrophilic elongation factor Tu from the antarctic Moraxella sp. Tac II 25: biochemical characterization and cloning of the encoding gene. Biochemistry 39, 15531–15539.

Medigue, C., Krin, E., Pascal, G, Barbe, V., Bernsel, A., Bertin, P. N., Cheung, F., Cruveiller, S., D'Amico, S., Duilio, A., Fang, G., Feller, G., Ho, C., Mangenot, S., Marino, G., Nilsson, J., Parrilli, E., Rocha, E. P., Rouy, Z., Sekowska, A., Tutino, M. L., Vallenet, D., von Heijne, G., Danchin, A., 2005. Coping with cold: the genome of the versatile marine Antarctica bacterium *Pseudoalteromonas haloplanktis* TAC125. Genome Res. 15, 1325–1335.

Miller, A. F., 2004. Superoxide dismutases: active sites that save, but a protein that kills. Curr. Opin. Chem. Biol. 8, 162–168.

Niehaus, F., Bertoldo, C., Kahler, M., Antranikian, G., 1999. Extremophiles as a source of novel enzymes for industrial application. Appl. Microbiol. Biotechnol. 51, 711–729.

Nisbet, E. G., Sleep, N. H., 2001. The habitat and nature of early life. Nature 409, 1083–1091.

Parmeggiani, A., Sander, G., 1981. Properties and regulation of the GTPase activities of elongation factors Tu and G, and of initiation factor 2. Mol. Cell. Biochem. 35, 129–158.

Petukhov, M., Kil, Y., Kuramitsu, S., Lanzov, V., 1997. Insights into thermal resistance of proteins from the intrinsic stability of their alpha-helices. Proteins 29, 309–320.

Rabus, R., Ruepp, A., Frickey, T., Rattei, T., Fartmann, B., Stark, M., Bauer, M., Zibat, A., Lombardot, T., Becker, I., Amann, J., Gellner, K., Teeling, H., Leuschner, W. D., Glockner, F. O., Lupas, A. N., Amann, R., Klenk, H. P., 2004. The genome of *Desulfotalea psychrophila*, a sulfate-reducing bacterium from permanently cold Arctic sediments. Environ. Microbiol. 6, 887–902.

Raghava, G. P. S., Han, J. H., 2005. Correlation and prediction of gene expression level from amino acid and dipeptide composition of its protein. BMC Bioinformatics 6, 59–72.

Raimo. G., Masullo, M., Savino, G., Scarano, G., Ianniciello, G., Parente, A., Bocchini, V., 1996. Archaeal elongation factor 1 beta is a dimer. Primary structure, molecular and biochemical properties. Biochim. Biophys. Acta 1293, 106–112.

Ruocco. M. R., Ruggiero, A., Masullo, L., Arcari, P., Masullo, M., 2004. A 35 kDa NAD(P)H oxidase previously isolated from the archaeon *Sulfolobus solfataricus* is instead a thioredoxin reductase. Biochimie 86, 883–892.

Russell, R. J., Gerike, U., Danson, M. J., Hough, D. W., Taylor. G. L., 1998. Structural adaptations of the cold–active citrate synthase from an Antarctic bacterium. Structure 6, 351–361.

Sarin, J., Raghava, G. P. S., Chakraborti, P. K., 2003. Intrinsic contributions of polar amino acid residues towards thermal stability of an ABC-ATPase of mesophilic origin. Protein Sci. 12, 2118–2120.

Saunders, N. F. W., Thomas, T., Curmi, P. M. G., Mattick, J. S., Kuczek, E., Slade, R., Davis, J., Franzmann, P. D., Boone, D., Rusterholtz, K., Feldman, R., Gates, C., Bench, S., Sowers, K., Kadner, K., Aerts, A., Dehal, P., Detter, C., Glavina, T., Lucas, S., Richardson, P., Larimer, F., Hauser, L., Land, M., Cavicchioli, R., 2003. Mechanisms of thermal adaptation revealed from the genomes of the Antarctic

archaea *Methanogenium frigidum* and *Methanococcoides burtonii*. Genome Res. 13, 1580–1588.

Shen, H. B. Chou, K. C., 2005. Using optimized evidence-theoretic K-nearest neighbor classifier and pseudo-amino acid composition to predict membrane protein types. Biochem. Biophys. Res. Commun. 334, 288–292.

Shen, H. B., Chou, K. C., 2006. Ensemble classifier for protein fold pattern recognition, Bioinformatics 22, 1717–1722.

Shen, H. B., Chou, K. C., 2007. Hum-mPLoc: an ensemble classifier for large-scale human protein subcellular location prediction by incorporating samples with multiple sites. Biochem. Biophys. Res. Commun. 355, 1006–1011.

Shen, H. B., Yang, J., Liu, X. J., Chou, K. C., 2005. Using supervised fuzzy clustering to predict protein structural classes. Biochem. Biophys. Res. Commun. 334 577–581.

Siddiqui, K. S., Cavicchioli, R., 2006. Cold–adapted enzymes. Annu. Rev. Biochem. 75, 403–433.

Sirover, M. A., 1999. New insights into an old protein: the functional diversity of mammalian glyceraldehyde–3–phosphate dehydrogenase. Biochim. Biophys. Acta 1432, 159–184.

Sterner, R., Liebl, W., 2001. Thermophilic adaptation of proteins. Crit. Rev. Biochem. Mol. Biol. 36, 39–106.

Stetter, K. O., 1996. Hyperthermophilic procaryotes. FEMS Microbiol. Rev. 18, 149–158.

Stetter, K. O., 1998. Hyperthermophiles: isolation, classification, and properties. In: Horikoshi, K., Grant, E. D. (Eds.), Extremophiles: microbial life in extreme environments, Wiley–Liss, New York, pp. 1–24.

Stetter, K. O., 1999. Extremophiles and their adaptation to hot environments. FEBS Lett. 452, 22–25.

Szilagyi, A., Zavodszky. P., 2000. Structural differences between mesophilic, moderately thermophilic and extremely thermophilic protein subunits: results of a comprehensive survey. Structure 8, 493–504.

Tanford, C., 1962. Contribution of hydrophobic interactions to the stability of the globular conformation of proteins. J. Am. Chem. Soc. 84, 4240–4247.

Tomita, K., Kuroki, Y., Hayashi, N., Komukai, Y., 2000. Isolation of a thermophile degrading poly(butylene succinate–*co*–butylene adipate). J. Biosci. Bioeng. 90, 350–352.

Ursby, T., Adinolfi, B. S., Al-Karadaghi, S., De Vendittis, E., Bocchini, V., 1999. Iron superoxide dismutase from the archaeon *Sulfolobus solfataricus*: analysis of structure and thermostability. J. Mol. Biol. 286, 189–205.

Vieille, C., Zeikus, G. J., 2001. Hyperthermopilic enzymes: sources, uses, and molecular mechanisms for thermostability. Microbiol. Mol. Biol. Rev. 65, 1–43.

Vieille, C., Burdette, D. S., Zeikus, J. G., 1996. Thermozymes. Biotechnol. Annu. Rev. 2, 1–83.

Violot. S,, Aghajari, N., Czjzek, M., Feller, G., Sonan, G. K., Gouet, P., Gerday, C., Haser, R., Receveur-Brechot, V., 2005. Structure of a full length psychrophilic cellulase from *Pseudoalteromonas haloplanktis* revealed by X–ray diffraction and small angle X–ray scattering. J. Mol. Biol. 348, 1211–1224.

Vitagliano, L., Masullo, M., Sica, F., Zagari, A., Bocchini, V., 2001. The crystal structure of *Sulfolobus solfataricus* elongation factor 1alpha in complex with GDP reveals novel features in nucleotide binding and exchange. EMBO J. 20, 5305–5311.

Wang, M., Yang, J., Liu, G. P., Xu, Z. J., Chou, K.C., 2004. Weighted-support vector machines for predicting membrane protein types based on pseudo-amino acid composition. Protein Eng. Des. Sel. 17, 509–516.

Zamyatnin, A. A., 1972. Protein volume in solution. Prog. Biophys. Mol. Biol. 24, 107–123.

Zillig, W., Holz, I., Klenk, H. P., Trent, J., Wunderl, S., Janekovic, D., Imsel, E., Haas, B., 1987. *Pyrococcus woesei*, sp. nov., an ultra–thermophilic marine archaebacterium, representing a novel order, *Thermococcales*. Syst. Appl. Microbiol. 9, 62–70.

Figure legends

Fig. 1. Effect of growth temperature on the average mass per residue in different model proteins. The amino acid composition of each protein was used to calculate the average mass per residue as described in the Experimental Procedure. (A) EF-Tu/1→; (B) EF-G/2; (C) EF-Ts/1→; (D) SOD; (E) GAPDH; (F) FtsZ.

Fig. 2. Effect of growth temperature on the average mass per residue in the three domains of EF-Tu/1→ The amino acid composition of each domain was used to calculate the average mass per residue as described in the Experimental Procedure. (A) Domain G; (B) Domain M; (C), Domain C.

Fig. 3. Effect of growth temperature on the average mass per residue in the two domains of SOD. The amino acid composition of the two domains was used to calculate the average mass per residue as described in the Experimental Procedure. (A) Domain N; (B) Domain C.

Fig. 4. Effect of growth temperature on the average hydrophobicity per residue in different model proteins. The amino acid composition of each protein was used to calculate the average hydrophobicity per residue as described in the Experimental Procedure. (A) EF-Tu/1→; (B) EF-G/2; (C) EF-Ts/1→; (D) SOD; (E) GAPDH; (F) FtsZ.

Fig. 5. Effect of growth temperature on the average volume per residue in different model proteins. The amino acid composition of each protein was used to calculate the

average volume per residue as described in the Experimental Procedure. (A) EF-Tu/1→; (B) EF-G/2; (C) EF-Ts/1→; (D) SOD; (E) GAPDH; (F) FtsZ.

Fig. 6. Effect of growth temperature on the average accessible surface area per residue in different model proteins. The amino acid composition of each protein was used to calculate the average accessible surface area per residue as described in the Experimental Procedure. (A) EF-Tu/1→; (B) EF-G/2; (C) EF-Ts/1→; (D) SOD; (E) GAPDH; (F) FtsZ.

Fig. 7. Effect of growth temperature on parameters deduced from the average amino acid composition collectively obtained from six model proteins. The average amino acid composition of a group of six ubiquitous proteins including EF-Tu/1→; EF-G/2, EF-Ts/1→; SOD, GAPDH and FtsZ was used to calculate the following average parameters per residue as described in the Experimental Procedure. (A) Average mass; (B) average hydrophobicity; (C) average volume; (D) average accessible surface area.

Table 1

Microbial sources and their respective optimum growth temperatures

Microorganism	Domain	Optimum growth temperature [a] (°C)
Desulfotalea psychrophila strain LSv54	bacteria	7
Colwellia psychrerythraea strain 34H	bacteria	8
Psychrobacter arcticus strain 273-4	bacteria	8.75 [b]
Polaribacter irgensii strain 23-P	bacteria	10 [c]
Psychromonas ingrahamii 37	bacteria	10 [d]
Pseudoalteromonas haloplanktis TAC125	bacteria	15 [e]
Photobacterium profundum strain SS9	bacteria	15
Leifsonia xyli strain CTCB07	bacteria	22.5 [b]
Methanococcoides burtonii strain DSM 6242	archaea	23.4
Azoarcus sp. EbN1	bacteria	26
Xylella fastidiosa strain Temecula1	bacteria	27 [b]
Pseudomonas aeruginosa 2192	bacteria	27.5 [b]
Oceanobacillus iheyensis strain HTE831	bacteria	30
Lactobacillus acidophilus strain NCFM	bacteria	30 [b]
Escherichia coli strain K12	bacteria	37
Salmonella typhimurium strain LT2	bacteria	37
Streptococcus thermophilus strain LMG 18311	bacteria	45
Methylococcus capsulatus strain Bath	bacteria	45
Chlorobium tepidum strain TLS	bacteria	48
Moorella thermoacetica strain ATCC 39073	bacteria	58
Thermoplasma acidophilum strain DSM 1728	archaea	59
Symbiobacterium thermophilum strain IAM 14863	bacteria	60
Picrophilus torridus strain DSM 9790	archaea	60
Bacillus stearothermophilus	bacteria	63 [f]
Methanothermobacter thermautotrophicus strain Delta H	archaea	67.5 [b]
Thermus thermophilus strain HB27	bacteria	68
Sulfolobus acidocaldarius strain DSM 639	archaea	72.5 [b]

Thermoanaerobacter tengcongensis strain MB4	bacteria	75
Sulfolobus tokodaii strain 7	archaea	80
Thermotoga maritima strain MSB8	bacteria	80
Archaeoglobus fulgidus strain DSM 4304	archaea	83
Sulfolobus solfataricus strain P2	archaea	85
Thermofilum pendens strain Hrk 5	archaea	88
Staphylothermus marinus strain F1	archaea	92
Aeropyrum pernix strain K1	archaea	92.5 [b]
Aquifex aeolicus strain VF5	bacteria	96
Pyrococcus horikoshii strain OT3	archaea	98
Pyrococcus furiosus strain DSM 3638	archaea	100
Pyrobaculum aerophilum strain IM2	archaea	100
Hyperthermus butylicus strain DSM 5456	archaea	100.5 [b]
Pyrococcus woesei	archaea	101.5 [g]
Pyrococcus abyssi strain GE5	archaea	103

[a] Unless otherwise indicated, optimum growth temperatures were derived from the National Center for Biotechnology Information database (www.ncbi.nlm.nih.gov)

[b] Mean calculated from the interval of optimum growth temperature indicated in www.ncbi.nlm.nih.gov

[c] Value according to Gosink et al., 1998.

[d] Value according to Auman et al., 2006.

[e] Value according to Corsaro et al., 2004.

[f] Value according to Tomita et al., 2000.

[g] Mean value calculated from an interval of optimum growth temperature Zillig et al., 1987.

Table 2.

Values of mass, hydrophobicity, volume, and accessible surface area of amino acid residues

Amino acid residue	Mass (Da)	Hydrophobicity [a] (kJ•mol^{-1})	Volume [b] (Å^3)	Accessible surface area [c] (Å^2)
Alanine	71.08	3.14	88.6	115
Arginine	156.20	3.14	173.4	225
Aspartic acid	115.09	0	111.1	150
Asparagine	114.11	0	117.7	160
Cysteine	103.14	4.19	108.5	135
Glutamic acid	129.12	0	138.4	190
Glutamine	128.14	0	143.9	180
Glycine	57.06	0	60.1	75
Histidine	137.15	0	153.2	195
Isoleucine	113.17	12.35	166.7	175
Leucine	113.17	10.05	166.7	170
Lysine	128.18	6.28	168.6	200
Methionine	131.21	5.44	162.9	185
Phenylalanine	147.18	11.10	189.9	210
Proline	97.12	10.89	122.7	145
Serine	87.08	0	89.0	115
Threonine	101.11	1.88	116.1	140
Triptophan	186.21	12.56	227.8	255
Tyrosine	163.18	11.93	193.6	230
Valine	99.14	7.12	140.0	155

[a] Values according to Tanford, 1962.

[b] Values according to Zamyatnin, 1972.

[c] Values according to Chothia, 1976.

Table 3

Significance level of the linear dependence on growth temperature of the average mass per residue

Protein, domain name	*n*	Linear curve fit ($y = a + b \bullet x$)					Ideal average mass at 37°C (Da)
		Intercept at 0°C (Da)	Slope β 10^2 (Da•°C^{-1})	Correlation index			
				r	*t*	*p*	
EF-Tu/1β	44	108.84	2.328	0.629	8.44	< 0.001	109.70
EF-G/2	51	109.71	2.342	0.572	8.09	< 0.001	110.58
EF-Ts/1β	40	105.77	7.309	0.624	7.94	< 0.001	108.47
SOD	37	110.15	5.608	0.456	5.42	< 0.001	112.22
GAPDH	42	106.79	3.923	0.516	6.53	< 0.001	108.24
FtsZ	41	103.99	4.426	0.467	5.85	< 0.001	105.63
EF-Tu/1β, domain G	44	108.66	4.452	0.568	7.43	< 0.001	110.31
EF-Tu/1β, domain M	44	109.41	–1.351	0.107	2.24	< 0.05	108.91
EF-Tu/1β, domain C	44	109.08	1.216	0.173	2.96	< 0.01	109.53
SOD, domain N	37	111.03	4.859	0.265	3.55	< 0.001	112.83
SOD, domain C	37	109.62	6.244	0.400	4.83	< 0.001	111.93

n is the number of amino acid sequences.

In the equation $y = a + b \bullet x$, *y* is the average mass per amino acid residue, *a* the intercept at 0°C, *b* the slope of the equation, *x* the optimum growth temperature.

Table 4

Significance level of the linear dependence on growth temperature of the average hydrophobicity per residue

Protein, domain name	*n*	Linear curve fit ($y = a + b \bullet x$)					Ideal average hydrophobicity at 37°C (kJ•mol^{-1})
		Intercept at 0°C (kJ•mol^{-1})	Slope β 10^3 (kJ•mol^{-1}•°C^{-1})	Correlation value			
				r	*t*	*p*	
EF-Tu/1β	44	4.40	6.063	0.610	8.11	< 0.001	4.62
EF-G/2	51	4.37	6.356	0.724	11.34	< 0.001	4.61
EF-Ts/1β	40	4.14	7.668	0.543	6.72	< 0.001	4.42
SOD	37	4.60	6.087	0.510	6.04	< 0.001	4.83
GAPDH	42	4.34	6.118	0.699	9.64	< 0.001	4.57
FtsZ	41	3.94	8.224	0.721	10.04	< 0.001	4.24
EF-Tu/1β, domain G	44	4.38	5.755	0.626	8.38	< 0.001	4.59
EF-Tu/1β, domain M	44	4.20	6.049	0.411	5.41	< 0.001	4.42
EF-Tu/1β, domain C	44	4.63	6.263	0.457	5.95	< 0.001	4.86
SOD, domain N	37	4.54	5.451	0.240	3.32	< 0.01	4.74
SOD, domain C	37	4.65	6.605	0.454	5.39	< 0.001	4.89

n is the number of amino acid sequences.

In the equation $y = a + b \bullet x$, *y* is the average hydrophobicity per amino acid residue, *a* the intercept at 0°C, *b* the slope of the equation, *x* the optimum growth temperature.

Table 5

Significance level of the linear dependence on growth temperature of the average volume per residue

Protein name	*n*	Linear curve fit ($y = a + b \bullet x$)					Ideal average volume at 37°C (Å^3)
		Intercept at 0°C (Å^3)	Slope β 10^2 (Å^3•°C^{-1})	Correlation value			
				r	*t*	*p*	
EF-Tu/1β	44	131.3	5.010	0.746	11.11	< 0.001	133.2
EF-G/2	51	132.1	5.099	0.757	12.35	< 0.001	134.0
EF-Ts/1β	40	129.0	9.291	0.678	8.94	< 0.001	132.4
SOD	37	132.3	8.818	0.549	6.53	< 0.001	135.6
GAPDH	42	129.1	6.491	0.710	9.90	< 0.001	131.5
FtsZ	41	125.3	7.699	0.652	8.55	< 0.001	128.1

n is the number of amino acid sequences.

In the equation $y = a + b \bullet x$, *y* is the average volume per amino acid residue, *a* the intercept at 0°C, *b* the slope of the equation, *x* the optimum growth temperature.

Table 6

Significance level of the linear dependence on growth temperature of the average accessible surface area per residue

Protein name	*n*	Linear curve fit ($y = a + b \bullet x$)					Ideal average accessible surface area at 37°C ($Å^2$)
		Intercept at 0°C ($Å^2$)	slope β 10^2 ($Å^2 \bullet °C^{-1}$)	Correlation value			
				r	*t*	*P*	
EF-Tu/1β	44	158.0	4.737	0.746	11.11	< 0.001	159.8
EF-G/2	51	159.2	4.719	0.714	11.06	< 0.001	160.9
EF-Ts/1β	40	155.8	9.873	0.589	7.38	< 0.001	159.5
SOD	37	159.1	9.383	0.543	6.45	< 0.001	162.6
GAPDH	42	154.8	7.327	0.667	8.95	< 0.001	157.5
FtsZ	41	151.2	7.582	0.575	7.26	< 0.001	154.0

n is the number of amino acid sequences.

In the equation $y = a + b \bullet x$, *y* is the average accessible surface area per amino acid residue, *a* the intercept at 0°C, *b* the slope of the equation, *x* the optimum growth temperature.

Table 7

Significance level of the linear dependence on growth temperature of average parameters related to the collective amino acid composition of six model proteins in forty-two microorganisms

Average parameter	Linear curve fit ($y = a + b \bullet x$)					Ideal average parameter at 37°C
	Intercept at 0°C	Slope β 10^2	Correlation value			
			R	*t*	*p*	
Mass	108.09 Da	3.331 Da•°C^{-1}	0.742	10.73	< 0.001	109.32 Da
Hydrophobicity	4.28 kJ•mol^{-1}	0.7076 kJ•mol^{-1}•°C^{-1}	0.770	11.57	< 0.001	4.54 kJ•mol^{-1}
Volume	130.3 $Å^3$	6.209 $Å^3$•°C^{-1}	0.815	13.27	< 0.001	132.6 $Å^3$
Accessible surface area	157.1 $Å^2$	6.006 $Å^2$•°C^{-1}	0.816	13.32	< 0.001	159.3 $Å^2$

In the equation $y = a + b \bullet x$, y is the average parameter per residue, a the intercept at 0°C, b the slope of the equation, x the optimum growth temperature.

Fig. 1

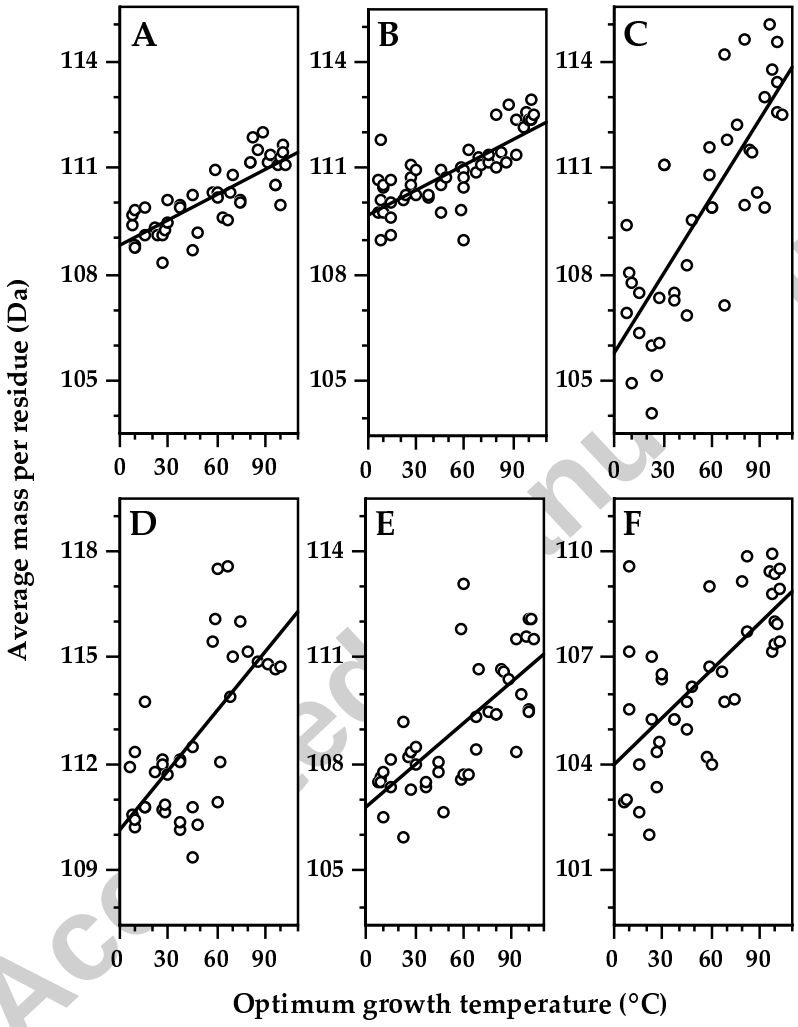

Fig. 2

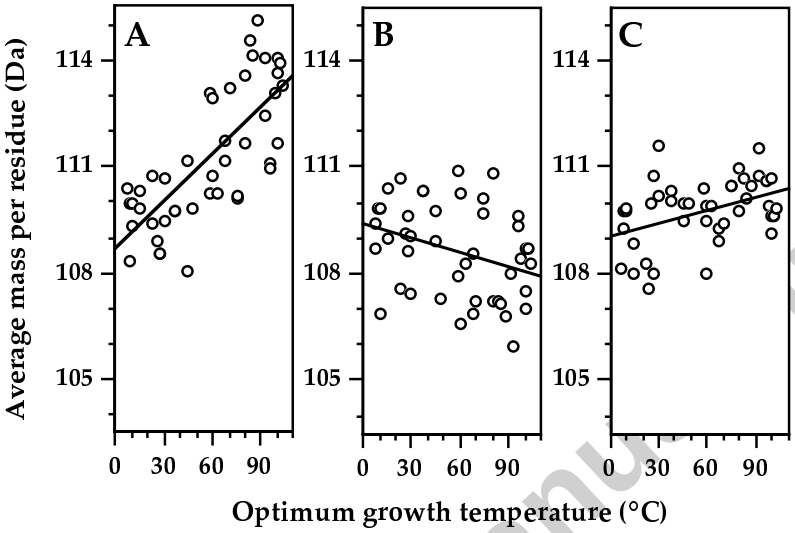

Fig. 3

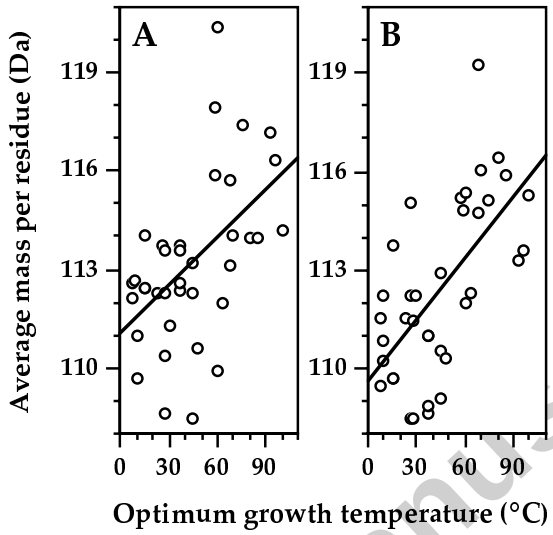

Fig. 4

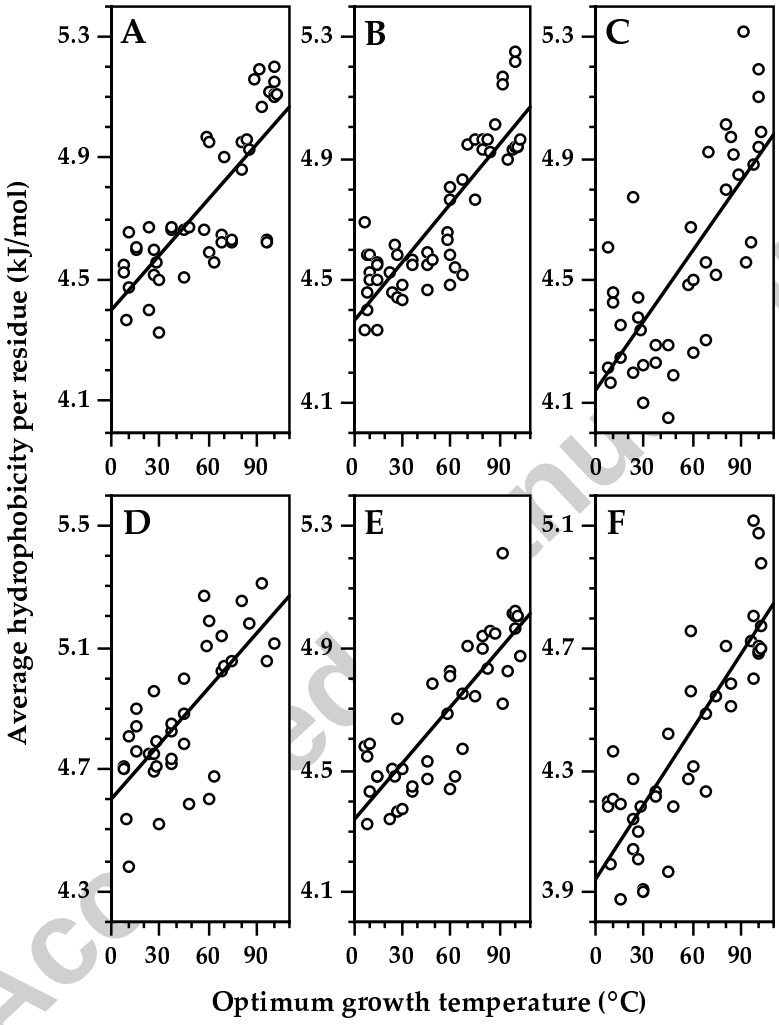

Fig. 5

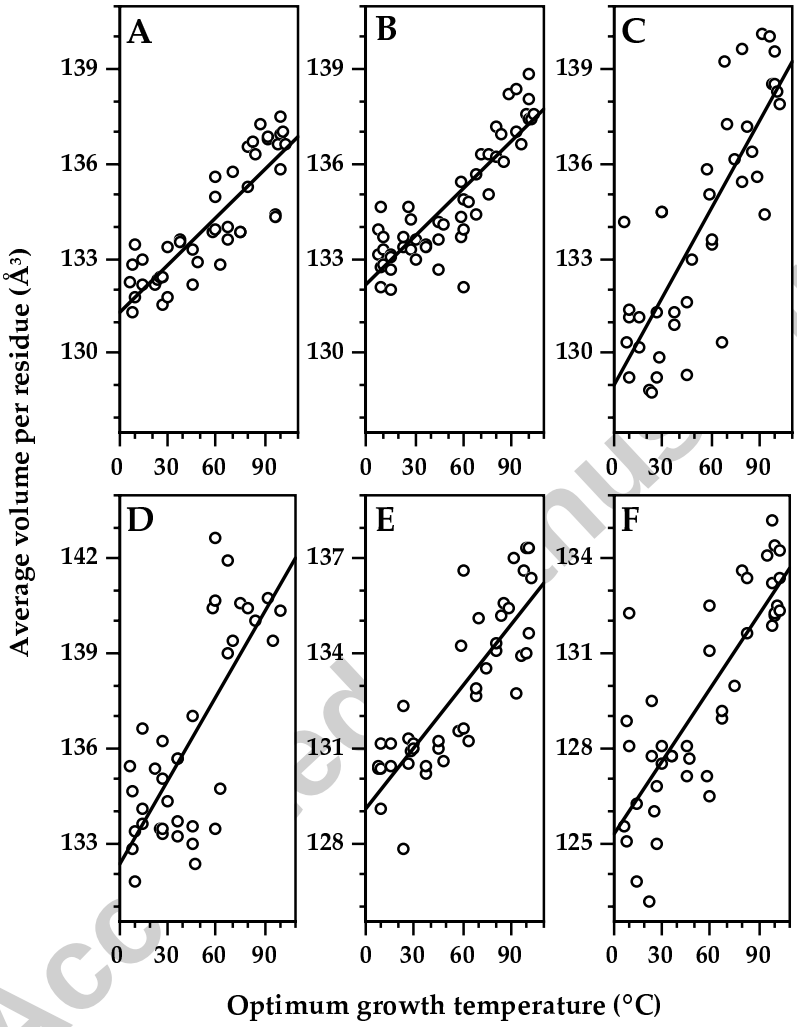

Fig. 6

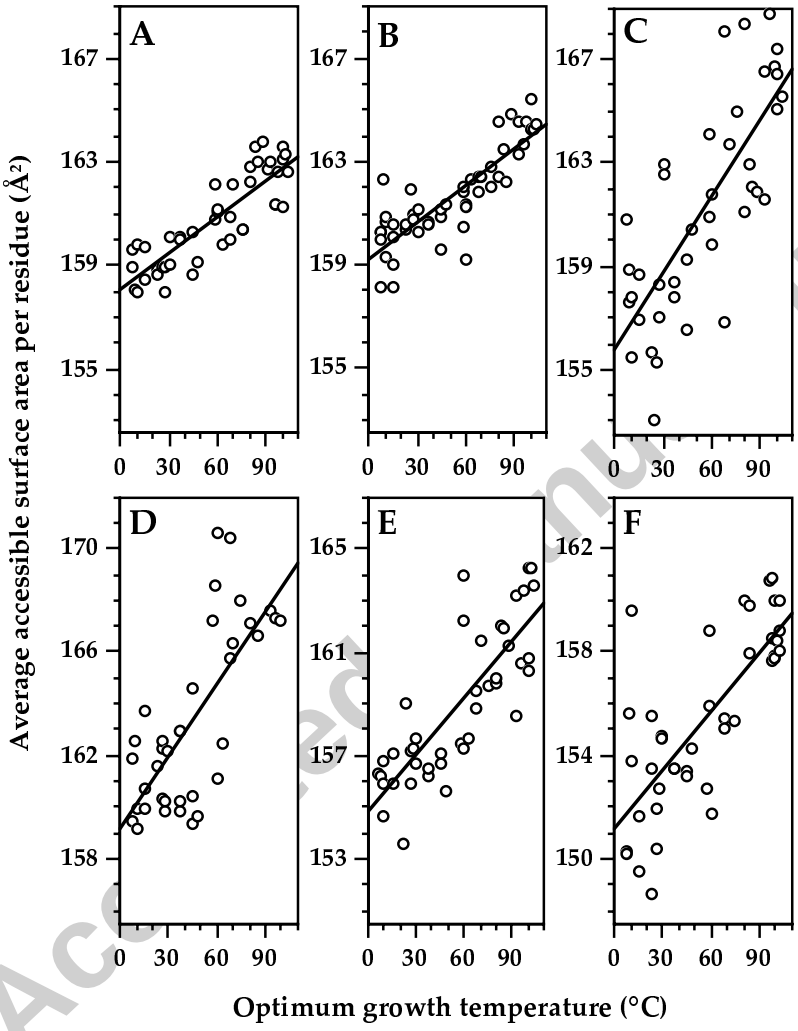

Fig. 7

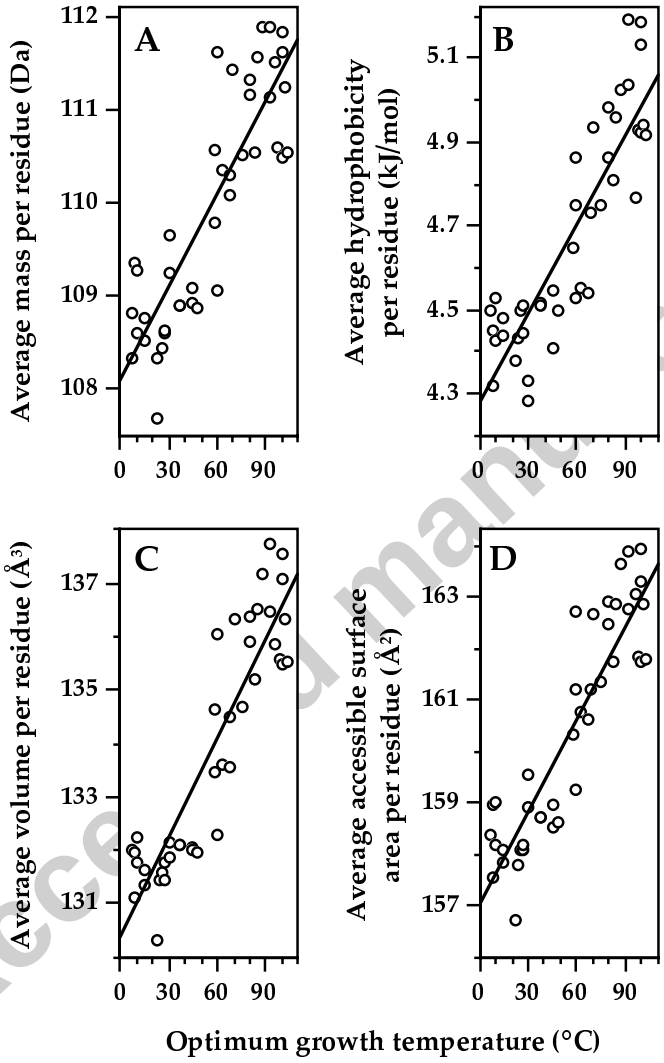

Table S1: Analyzed sequences of elongation factor Tu/1β

Microbial source	No residues	Accession number
Desulfotalea psychrophila strain LSv54	396	CAG35851
Colwellia psychrerythraea strain 34H	394	YP_271423
Psychrobacter arcticus strain 273-4	396	YP_265176
Polaribacter irgensii strain 23-P	395	ZP_01117987
Psychromonas ingrahamii 37	394	YP_944720
Pseudoalteromonas haloplanktis TAC125	393	CAB65285
Photobacterium profundum strain SS9	394	CAG18755
Leifsonia xyli strain CTCB07	397	Q6ACZ0
Methanococcoides burtonii strain DSM 6242	422	YP_565843
Azoarcus sp. EbN1	396	CAI08280
Xylella fastidiosa strain Temecula 1	396	NP_780188
Pseudomonas aeruginosa 2192	397	P09591
Oceanobacillus iheyensis strain HTE831	395	NP_691038
Lactobacillus acidophilus strain NCFM	396	YP_193737
Escherichia coli strain K12	394	P0A6N1
Salmonella typhimurium strain LT2	394	NP_463015
Streptococcus thermophilus strain LMG 18311	398	AAV60197
Methylococcus capsulatus strain Bath	396	AAU92683
Chlorobium tepidum strain TLS	393	NP_663065
Moorella thermoacetica strain ATCC 39073	400	YP_431287
Thermoplasma acidophilum strain DSM 1728	424	NP_393922
Symbiobacterium thermophilum strain IAM 14863	395	YP_076903
Picrophilus torridus strain DSM 9790	424	YP_023193
Bacillus stearothermophilus	395	CAA03976
Methanothermobacter thermautotrophicus strain Delta H	413	NP_276188
Thermus thermophilus strain HB27	406	YP_005703
Sulfolobus acidocaldarius strain DSM 639	435	YP_255358
Thermoanaerobacter tengcongensis strain MB4	400	NP_623833
" "	400	NP_623847
Sulfolobus tokodaii strain 7	435	NP_376127
Thermotoga maritima strain MSB8	400	NP_229302
Archaeoglobus fulgidus strain DSM 4304	423	NP_069770
Sulfolobus solfataricus strain P2	435	P35021
Thermophilum pendens strain Hrk 5	433	YP_920052
Staphylothermus marinus strain F1	438	YP_001040819
Aeropyrum pernix strain K1	437	NP_148207
Aquifex aeolicus strain VF5	405	NP_212987
" "	405	NP_214323
Pyrococcus horikoshii strain OT3	428	NP_143347
Pyrococcus furiosus strain DSM 3638	428	NP_579104
Pyrobaculum aerophilum strain IM2	444	NP_560418
Hyperthermus butylicus strain DSM 5456	440	YP_001013747
Pyrococcus woesei	430	CAA42517
Pyrococcus abyssi strain GE5	428	CAB49596

Table S2: Analyzed sequences of elongation factor G/2

Microbial source	No residues	Accession number
Desulfotalea psychrophila strain LSv54	692	CAG35850
" "	695	CAG34736
Colwellia psychrerythraea strain 34H	699	YP_271409
" "	701	YP_267624
Psychrobacter arcticus strain 273-4	708	YP_265177
Polaribacter irgensii strain 23-P	705	ZP_01119277
Psychromonas ingrahamii 37	697	YP_944721
" "	698	YP_941819
Pseudoalteromonas haloplanktis TAC125	704	CAE00448
" "	694	CAI87975
Photobacterium profundum strain SS9	695	CAG19645
" "	698	CAG18754
Leifsonia xyli strain CTCB07	700	Q6ACY9
Methanococcoides burtonii strain DSM 6242	730	YP_565844
Azoarcus sp. EbN1	683	CAI06555
Xylella fastidiosa strain Temecula 1	705	AAO29826
Pseudomonas aeruginosa 2192	702	EAZ57945
Oceanobacillus iheyensis strain HTE831	692	Q8ETY5
Lactobacillus acidophilus strain NCFM	697	AAV42182
Escherichia coli strain K12	704	P0A6M8
Salmonella typhimurium strain LT2	704	AAL22309
Streptococcus thermophilus strain LMG 18311	693	AAV61388
Methylococcus capsulatus strain Bath	694	AAU93266
" "	698	YP_114791
Chlorobium tepidum strain TLS	704	NP_663066
Moorella thermoacetica strain ATCC 39073	680	YP_429999
" "	692	YP_431288
Thermoplasma acidophilum strain DSM 1728	732	P26752
Symbiobacterium thermophilum strain IAM 14863	694	YP_076904
" "	695	YP_076966
Picrophilus torridus strain DSM 9790	732	Q6L200
Bacillus stearothermophilus	692	CAC09927
Methanothermobacter thermautotrophicus strain Delta H	730	O27131
Thermus thermophilus strain HB27	691	Q72I01
Sulfolobus acidocaldarius strain DSM 639	724	AAY79995
Thermoanaerobacter tengcongensis strain MB4	690	NP_623834
" "	700	AAM25474
Sulfolobus tokodaii strain 7	724	BAB65426
Thermotoga maritima strain MSB8	692	P38525
Archaeoglobus fulgidus strain DSM 4304	728	NP_070719
Sulfolobus solfataricus strain P2	736	AAK41025
Thermophilum pendens strain Hrk 5	734	YP_920020
Staphylothermus marinus strain F1	736	YP_001040874
Aeropyrum pernix strain K1	736	NP_147939
Aquifex aeolicus strain VF5	699	NP_212986

Pyrococcus horikoshii strain OT3	732	O59521
Pyrococcus furiosus strain DSM 3638	732	NP_579741
Pyrobaculum aerophilum strain IM2	740	NP_558538
Hyperthermus butylicus strain DSM 5456	738	YP_001013036
Pyrococcus woesei	732	P61878
Pyrococcus abyssi strain GE5	732	CAB49200

Table S3: Analyzed sequences of elongation factor Ts/1β

Microbial source	No residues	Accession number
Desulfotalea psychrophila strain LSv54	196	YP_064891
Colwellia psychrerythraea strain 34H	282	YP_268296
Psychrobacter arcticus strain 273-4	294	YP_263654
Polaribacter irgensii strain 23-P	321	ZP_01117587
Psychromonas ingrahamii 37	292	YP_944276
Pseudoalteromonas haloplanktis TAC125	283	YP_340534
Photobacterium profundum strain SS9	284	Q6LN25
Leifsonia xyli strain CTCB07	276	YP_062194
Methanococcoides burtonii strain DSM 6242	89	YP_565987
Azoarcus sp. EbN1	296	CAI09540
Xylella fastidiosa strain Temecula 1	292	NP_780140
Pseudomonas aeruginosa 2192	289	EAZ59793
Oceanobacillus iheyensis strain HTE831	294	NP_692508
Lactobacillus acidophilus strain NCFM	341	YP_194131
Escherichia coli strain K12	283	NP_414712
Salmonella typhimurium strain LT2	283	NP_459222
Streptococcus thermophilus strain LMG 18311	351	YP_138618
Methylococcus capsulatus strain Bath	293	YP_113087
Chlorobium tepidum strain TLS	288	NP_662659
Moorella thermoacetica strain ATCC 39073	203	YP_429890
Thermoplasma acidophilum strain DSM 1728	88	Q9HKN1
Symbiobacterium thermophilum strain IAM 14863	304	YP_075321
Picrophilus torridus strain DSM 9790	101	AAT43749
Methanothermobacter thermautotrophicus strain Delta H	89	O27734
Thermus thermophilus strain HB27	196	YP_004483
Sulfolobus acidocaldarius strain DSM 639	90	Q4JAN4
Thermoanaerobacter tengcongensis strain MB4	200	NP_623026
Sulfolobus tokodaii strain 7	91	NP_376060
Thermotoga maritima strain MSB8	199	NP_229405
Archaeoglobus fulgidus strain DSM 4304	88	NP_069408
Sulfolobus solfataricus strain P2	91	NP_341733
Thermophilum pendens strain Hrk 5	91	YP_920066
Staphylothermus marinus strain F1	90	YP_001040727
Aeropyrum pernix strain K1	90	NP_148638
Aquifex aeolicus strain VF5	290	NP_213490
Pyrococcus horikoshii strain OT3	91	NP_142049
Pyrococcus furiosus strain DSM 3638	91	NP_579694
Pyrobaculum aerophilum strain IM2	92	NP_558776
Hyperthermus butylicus strain DSM 5456	92	YP_001013529
Pyrococcus abyssi strain GE5	95	CAB48952

Table S4: Analyzed sequences of superoxide dismutase

Microbial source	No residues	Accession number
Desulfotalea psychrophila strain LSv54	197	YP_064052
Colwellia psychrerythraea strain 34H	194	YP_270150
Psychrobacter arcticus strain 273-4	209	YP_265220
Polaribacter irgensii strain 23-P	202	ZP_01117062
Psychromonas ingrahamii 37	193	YP_943778
Pseudoalteromonas haloplanktis TAC125	193	CAI86290
Photobacterium profundum strain SS9	194	CAG20957
" "	203	CAG20948
Leifsonia xyli strain CTCB07	208	YP_062104
Azoarcus sp. EbN1	195	CAI09001
Xylella fastidiosa strain Temecula 1	203	NP_780168
" "	230	NP_779088
Pseudomonas aeruginosa 2192	193	EAZ62265
" "	203	EAZ62159
Oceanobacillus iheyensis strain HTE831	203	BAC13888
Escherichia coli strain K12	193	NP_416173
" "	206	NP_418344
Salmonella typhimurium strain LT2	193	NP_460394
" "	206	NP_462936
Streptococcus thermophilus strain LMG 18311	220	YP_139228
Methylococcus capsulatus strain Bath	194	AAU91441
" "	210	AAU91964
Streptococcus thermophilus strain LMG 18311	200	NP_662101
Methylococcus capsulatus strain Bath	226	YP_430759
Thermoplasma acidophilum strain DSM 1728	205	NP_393491
Symbiobacterium thermophilum strain IAM 14863	204	YP_074876
Picrophilus torridus strain DSM 9790	205	YP_023258
Bacillus stearothermophilus	204	P00449
Methanothermobacter thermautotrophicus strain Delta H	205	AAB84666
Thermus thermophilus strain HB27	204	YP_004164
Sulfolobus acidocaldarius strain DSM 639	211	YP_254907
Thermoanaerobacter tengcongensis strain MB4	188	NP_622509
Sulfolobus tokodaii strain 7	211	NP_378284
Sulfolobus solfataricus strain P2	211	NP_341862
Aeropyrum pernix strain K1	214	NP_147461
Aquifex aeolicus strain VF5	213	NP_214035
Pyrobaculum aerophilum strain IM2	211	NP_558493

Table S5: Analyzed sequences of glyceraldehyde-3-phosphate dehydrogenase

Microbial source	No residues	Accession number
Desulfotalea psychrophila strain LSv54	334	YP_064558
Colwellia psychrerythraea strain 34H	334	YP_269060
Psychrobacter arcticus strain 273-4	481	YP_264787
Polaribacter irgensii strain 23-P	333	ZP_01119221
Psychromonas ingrahamii 37	330	YP_943703
Pseudoalteromonas haloplanktis TAC125	334	CAI86443
Photobacterium profundum strain SS9	339	YP_130406
Leifsonia xyli strain CTCB07	336	YP_062105
Methanococcoides burtonii strain DSM 6242	335	YP_565552
Azoarcus sp. EbN1	338	CAI06702
Xylella fastidiosa strain Temecula 1	336	NP_779817
Pseudomonas aeruginosa 2192	334	EAZ59323
Oceanobacillus iheyensis strain HTE831	335	BAC14394
Lactobacillus acidophilus strain NCFM	338	YP_193604
Escherichia coli strain K12	331	NP_416293
Salmonella typhimurium strain LT2	331	NP_460256
Streptococcus thermophilus strain LMG 18311	345	YP_140202
Methylococcus capsulatus strain Bath	336	YP_115003
Chlorobium tepidum strain TLS	334	NP_662365
Moorella thermoacetica strain ATCC 39073	335	YP_429140
Thermoplasma acidophilum strain DSM 1728	338	NP_394562
Symbiobacterium thermophilum strain IAM 14863	336	YP_075474
Picrophilus torridus strain DSM 9790	341	Q6L125
Bacillus stearothermophilus	335	P00362
Methanothermobacter thermautotrophicus strain Delta H	337	NP_276144
Thermus thermophilus strain HB27	331	YP_004524
Sulfolobus acidocaldarius strain DSM 639	343	YP_255984
Thermoanaerobacter tengcongensis strain MB4	335	NP_623352
Sulfolobus tokodaii strain 7	343	NP_377309
Thermotoga maritima strain MSB8	333	NP_228497
Archaeoglobus fulgidus strain DSM 4304	340	NP_070560
Sulfolobus solfataricus strain P2	340	NP_342058
Thermophilum pendens strain Hrk 5	342	YP_920162
Staphylothermus marinus strain F1	348	YP_001040262
Aeropyrum pernix strain K1	343	NP_147019
Aquifex aeolicus strain VF5	342	NP_213724
Pyrococcus horikoshii strain OT3	334	NP_143662
Pyrococcus furiosus strain DSM 3638	334	NP_579603
Pyrobaculum aerophilum strain IM2	344	NP_559513
Hyperthermus butylicus strain DSM 5456	343	YP_001013132
Pyrococcus woesei	334	P61880
Pyrococcus abyssi strain GE5	334	NP_126077

Table S6: Analyzed sequences of cell division protein FtsZ

Microbial source	No residues	Accession number
Desulfotalea psychrophila strain LSv54	420	YP_066629
Colwellia psychrerythraea strain 34H	386	YP_271107
Psychrobacter arcticus strain 273-4	398	YP_265027
Polaribacter irgensii strain 23-P	639	ZP_01116810
Psychromonas ingrahamii 37	388	YP_942588
Pseudoalteromonas haloplanktis TAC125	416	CAI87547
Photobacterium profundum strain SS9	394	CAG21517
Leifsonia xyli strain CTCB07	382	YP_062434
Methanococcoides burtonii strain DSM 6242	368	YP_566573
" "	394	YP_565059
Azoarcus sp. EbN1	379	YP_157811
Xylella fastidiosa strain Temecula 1	411	NP_780044
Pseudomonas aeruginosa 2192	394	EAZ62223
Oceanobacillus iheyensis strain HTE831	391	NP_692394
Lactobacillus acidophilus strain NCFM	452	YP_193706
Escherichia coli strain K12	383	NP_414637
Salmonella typhimurium strain LT2	383	NP_459138
Streptococcus thermophilus strain LMG 18311	440	YP_139243
Methylococcus capsulatus strain Bath	382	YP_114837
Chlorobium tepidum strain TLS	434	NP_660936
Moorella thermoacetica strain ATCC 39073	355	YP_429710
Thermoplasma acidophilum strain DSM 1728	345	NP_393984
" "	395	O59635
Symbiobacterium thermophilum strain IAM 14863	354	YP_075048
Methanothermobacter thermautotrophicus strain Delta H	381	NP_276787
Thermus thermophilus strain HB27	352	YP_004699
Thermoanaerobacter tengcongensis strain MB4	357	NP_623237
Thermotoga maritima strain MSB8	351	NP_228645
Archaeoglobus fulgidus strain DSM 4304	368	NP_069371
" "	392	NP_069404
Aquifex aeolicus strain VF5	367	NP_213369
Pyrococcus horikoshii strain OT3	365	NP_143219
" "	372	NP_142027
" "	414	NP_142705
Pyrococcus furiosus strain DSM 3638	366	NP_579236
" "	372	NP_579717
" "	408	NP_578254
Pyrococcus woesei	366	Q52630
Pyrococcus abyssi strain GE5	365	CAB49728
" "	372	NP_125696
" "	413	NP_126968