Amino Acids (2008) 34: 25–33
DOI 10.1007/s00726-007-0589-x
Printed in The Netherlands

Amino Acids

Differences in amino acids composition and coupling patterns between mesophilic and thermophilic proteins

Review Article

X.-X. Zhou[1], **Y.-B. Wang**[2], **Y.-J. Pan**[1], and **W.-F. Li**[3]

[1] Institute of Chemical Biology and Pharmaceutical Chemistry, Zhejiang University, Hangzhou, China
[2] Food Quality and Safety Department, Zhejiang Gongshang University, Hangzhou, China
[3] Microbiology Division, College of Animal Science, Zhejiang University, Hangzhou, China

Received May 31, 2007
Accepted July 12, 2007
Published online August 21, 2007;

Summary. Thermophilic proteins show substantially higher intrinsic thermal stability than their mesophilic counterparts. Amino acid composition is believed to alter the intrinsic stability of proteins. Several investigations and mutagenesis experiment have been carried out to understand the amino acid composition for the thermostability of proteins. This review presents some generalized features of amino acid composition found in thermophilic proteins, including an increase in residue hydrophobicity, a decrease in uncharged polar residues, an increase in charged residues, an increase in aromatic residues, certain amino acid coupling patterns and amino acid preferences for thermophilic proteins. The differences of amino acids composition between thermophilic and mesophilic proteins are related to some properties of amino acids. These features provide guidelines for engineering mesophilic protein to thermophilic protein.

Keywords: Thermophilic protein – Thermostability – Amino acid composition – Amino acid coupling pattern

Introduction

According to optimum growth temperature (OGT), organisms are classified as four groups: psychrophiles with OGT in the range of −5 to 20 °C, mesophiles with OGT in the range of 15 to 45 °C, thermophiles with OGT in the range of 45 to 80 °C, and hyperthermophiles (or extreme thermophile) with OGT above 80 °C (Vieille and Zeikus, 2001; Li et al., 2005). Hyperthermophiles were first isolated more than 20 years ago. Proteins originated from mesophiles are mesophilic proteins while that from thermophiles and hyperthermophile are thermophilic proteins (Pantazaki et al., 2007). Most hyperthermophiles belong to the phylogenetical domain of Archaeas, and they are members of the genera *Pyrobaculum, Pyrodictium*, *Pyrococcus* and *Methanopyrus*, whereas only two families, *Thermotoga maritime* and *Aquifex pyrophilus*, belong to the kingdom of Bacteria (Stetter, 1996; Trivedi et al., 2006).

Thermophilic proteins show substantially higher intrinsic thermal stability than their mesophilic counterparts with retaining the basic fold characteristic of the particular protein family (Jaenicke, 1996; Pack and Yoo, 2004). The term protein thermostability refers to the preservation of the unique chemical and spatial structure of a polypeptide chain under extremes of temperature conditions. Thermophilic proteins can exhibit higher core hydrophobicity (Schumann et al., 1993), greater numbers of ionic interactions (Vetriani et al., 1998), increased packing density (Russell et al., 1997), additional networks of hydrogen bonds (Jaenicke and Bohm, 1998), decreased lengths of surface loops (Thompson and Eisenberg, 1999), stabilization by heatstable chaperones (Haslbeck et al., 2005), an increase in disulfide bond formation (Beeby et al., 2005) and a general shortening of length (Das and Gerstein, 2000; Tekaia et al., 2002). Protein characteristic is believed to be related with amino acids composition and some of these structural factors seem to be obtained with the exchange of some amino acids (Diao et al., 2007; Scandurra et al., 1998). And in studying proteins from extremophiles, researchers have found no new amino acids, covalent modifications or structural motifs that ex-

plain the ability of these molecules to function in such harsh environments. Rather, subtle redistributions of the same intramolecular interactions required for protein stabilization at moderate temperatures are sufficient to maintain structural integrity at hot or cold extremes (Fields, 2001). Thus, the modifications in primary structure, that is, amino acid composition, are believed to alter the intrinsic stability of proteins, and comparison of sequence and structure of thermophilic and mesophilic proteins has formed the basis of theoretical efforts in elucidating the thermostability mechanisms.

Kumar et al. (2000) found that, despite high sequence homology, the differences in amino acid distributions in the thermophilic and mesophilic proteins are highly significant. While some of the differences in the amino acid distributions are likely to be the outcome of phylogenetic differences between thermophiles and mesophiles, others correlate with protein thermostability. It was observed that Gly, Ser, Lys, and Asp in mesophiles are generally substituted by Ala, Thr, Arg and Glu, respectively, in thermophiles to enhance the stability (Argos et al., 1979). Ponnuswamy et al. (1982) found a correlation between melting temperature and amino acid composition of a stabilizing and destabilizing group of amino acids and reported an empirical relation between thermostability and amino acid content of proteins. Much more investigations have been carried out recently to understand the amino acid composition for the thermostability of proteins (Singer and Hickey, 2003; Pack and Yoo, 2004, 2005; Liang et al., 2005; Sadeghi et al., 2006; Trivedi et al., 2006; Yokota et al., 2006; Lin et al., 2007). This review will present some generalized features of amino acid composition and amino acid coupling pattern found in thermophilic proteins.

Difference of amino acid composition between mesophilic and thermophilic protein

The first statistical analyses comparing amino acid compositions in mesophilic and thermophilic proteins indicated that the properties most correlated with the proteins of the thermophile include higher residue volume, higher residue hydrophobicity, more charged amino acids (especially Glu, Arg, and Lys), and fewer uncharged polar residues (Ser, Thr, Asn, and Gln) (Haney et al., 1999). As more experimental data accumulate (in particular, complete genome sequences), more results of the correlation of amino acid and thermostability are gained. The results demonstrated the properties of side chain of amino acids were determinants for thermostability of protein though thermophilic and mesophilic proteins have both similar polar and nonpolar contribution to the surface area and compactness (Sadeghi et al., 2006).

According to the properties of side chain of amino acids, amino acids are classified as four groups: (1) nonpolar (2) polar, uncharged (3) polar charged or (4) aromatic. The difference of these four groups of amino acids composition between mesophilic and thermophilic protein is highlighted below.

Non-polar amino acids

The nonpolar amino acids are characterized by having no polar atoms (only carbon and hydrogen) in their side chains. They include Glycine (Gly, G), Ala (Alanine, A), Val (Valine, V), Leu (Leucine, L), Ile (Isoleucine, I), Pro (Proline, P), and Met (Methionine, M). A general feature of globular proteins is that such hydrophobic residues are found in the protein interior, while polar residues occur on the surface.

The hydrophobic amino acids content is marginally higher in thermophilies than that in mesophilies and these residues can increase rigidity and hydrophobicity of proteins (Chakravarty and Varadarajan, 2000). Pack and Yoo (2004) found that there were the lower frequency in exposed state and the higher frequency in well-buried state of Ala (with one methyl group) in thermophilic proteins than those in mesophilic ones. It indicated that, in thermophilic proteins, the amino acids with the short alkyl group would tend to interact more closely with neighboring residues and have better packed form in protein structure. A higher Ala content in thermophilic proteins reflects the fact that Ala is the best helix-forming residue (Argos et al., 1979). Although the helices from thermophilic proteins contain a smaller fraction of beta-branched residues (Val, Ile, and Thr) than helices in mesophilic proteins, and beta-branched residues were found to destabilize α-helix, most systematical analyses showed that thermophilic proteins had higher frequency of Ile and Val compared with mesophilic ones (Chakravarty and Varadarajan, 2000; Kumar et al., 2000).

Among hydrophobic residues, Ala, Val, Leu, and Ile belong to the aliphatic amino acids. It has been widely accepted that the aliphatic amino acids would contribute to the hydrophobic interaction, which is main force for maintaining conformational stability in inner part of protein (Creighton, 1997). The aliphatic index is defined as the relative volume of a protein occupied by aliphatic side chains. The aliphatic index of a protein is calculated

according to the following formula: Aliphatic index = X(Ala) + a × X(Val) + b × (X(Ile) + X(Leu)), where X(Ala), X(Val), X(Ile), and X(Leu) are mole percent (100 × mole fraction) of alanine, valine, isoleucine, and leucine. The coefficients a and b are the relative volume of valine side chain (a = 2.9) and of Leu/Ile side chains (b = 3.9) to the side chain of alanine. The aliphatic index can be regarded as a positive factor for the increase of thermostability of globular proteins (Ikai, 1980). Lu et al. (1998) compared the difference of amino acid composition between 110 pairs of homologous thermophilic and mesophilic proteins and found that thermophilic proteins have higher average hydropathy and aliphatic index due to higher Leu composition.

Gly is known as the residue for making void volume or cavity in inner part of protein structure (Creighton, 1997). Thermophilic proteins have fewer Gly in a particular region of the structure (Panasik et al., 2000). However, another study showed that there is no typical pattern of Gly occurrence in thermophilic proteins (Pack and Yoo, 2004).

Pro residue, with their pyrolidine ring, can only adopt a few configurations and has the lowest conformational entropy, and thus restrict the configurations allowed for the preceding residue. It is known as the residue for making rigid conformation or turn conformation in protein structure (Watanabe et al., 1997). Some analysis showed that thermophilic proteins have higher frequency of Pro (Xu et al., 2003; Pack and Yoo, 2004; Sadeghi et al., 2006). Pro has been used to increase the protein stability in the several mutational studies (Veltman et al., 1996; Van den Burg et al., 1998) and hence an increase of Pro content may be due to the increase of the thermophilic protein rigidity. This indicated that the rigid conformation or turn conformation of thermophilic proteins might have better packed forms than those of mesophilic ones.

The side chain of Met includes a sulfur atom but remains hydrophobic in nature. Met is known as thermolabile amino acid due to its tendency to undergo oxidation at high temperature. Some systematical analyses reported that thermophilic proteins have lower frequency of Met compared with mesophilic proteins (Kumar et al., 2000; Xu et al., 2003). Beta-glucosidase A (BglA) from thermostable bacteria shows lower content of Met than do BglA and other enzymes of the family from mesophilic organisms (Lopez-Camacho et al., 1996).

Increase of overall hydrophobic group in thermophilic proteins could be due to the role of the hydrophobic effect that destabilizes the unfolded forms and it will increase with the temperature (Ikai, 1980; Britton et al., 1995).

Polar, uncharged amino acids

These are amino acids that possess oxygen, sulfur and/or nitrogen in the side chain and are therefore polar, but cannot have their side chain ionized and thus do not carry an overall charge. The polar nature of the side chain means that these amino acids are ready to interact with water (hydrophilic). This group includes Asparagine (Asn, N), Glutamate (Gln, Q), Cysteine (Cys, C), Serine (Ser, S) and Threonine (Thr, T).

Asn and Gln are known as thermolabile amino acids due to their tendency to undergo deamination at high temperature (Catanzano et al., 1997). Ser and Thr are known as the best residues for interacting with the waters surrounding protein (Mattos, 2002). Since the water that is interacted with Ser and Thr would be released at higher temperature, the local protein structure around water-binding site could be changed to be unstable enough to evoke protein instability (Denisov et al., 1997; Nagendra et al., 1998). There have been several reports that the uncharged polar residues contents, especially Ser, Thr, Gln, were much lower in thermophiles than that in mesophiles (Chakravarty and Varadarajan, 2000; Kumar et al., 2000; Pack and Yoo, 2004). These replacements may also minimize problems related to deamidation of Asn and Gln by Ser and Thr at high temperatures (Tomazic and Klibanov, 1988).

The side group of Cys also contains a sulfur atom. Unlike Met, the sulfur group in Cys comes at the end of the hydrocarbon chain and it therefore has the potential to be more reactive. Like Met, Cys is also known as thermolabile amino acids due to undergo oxidation at high temperature (Russell et al., 1994). So it has lower frequency in thermophilic proteins (Kumar et al., 2000; Xu et al., 2003). Gromiha et al. (1999) found that strong preference of replacements is observed for the mutations from meso to thermophile, Met → Ala, Cys → Ala, Trp → Tyr, Met → Leu, Cys → Val and Cys → Ile. Interestingly, five of the mutations are strongly favored from Cys and Met residues to other hydrophobic residues. This may be due to the fact that the size of the sulfur atoms is so different from other atoms, and they might not pack easily among the other more similarly sized atoms or due to the higher reactivity of sulfur atoms than that of carbon atoms. The substitution of Met and Cys by more stable amino acids has proved to be effective in the stabilization of different proteins, as it is the case of subtilisin (Estell et al., 1985), lysozymes (Perry and Wetzel, 1986), and BglA (Lopez-Camacho et al., 1996).

The mesophile–mesophile homologue comparison showed that the reduction of the uncharged polar residues

has specific role in the thermostability of thermophilic proteins (Sadeghi et al., 2006). The decrease in the content of uncharged polar residues in thermophilic proteins is likely to minimize deamination, oxidation and backbone cleavage involving Asn, Gln and Cys that are induced by temperature (Tomazic and Klibanov, 1988). It is also assumed that a reduction in uncharged polar residues will diminish the hydrogen bonding capability, while some could be retained by switching to charged amino acids.

Polar charged amino acids

Five amino acids, Arginine (Arg, R), Lysine (Lys, K), Histidine (His, H), Aspartate (Aspartic acid, Asp, D) and Glutamate (Glutamic acid, Glu, E) belong to this group. The side chains of them are not only polar but can also carry a positive charge or negative charge and are therefore highly hydrophilic. Lys and Arg both have pKs around 10.0 and are therefore always positively charged at neutral pH. His, with a pK of 6.5, can be uncharged or positively charged depending upon its local environment. Asp and Glu are only two amino acids with negatively charged side chains. In each case the pK of the side chain carboxyl group is about 4.4, so confer a negative charge on the proteins. The charged amino acids would contribute to the electrostatic interaction, which is an important force for maintaining conformational stability in the outer part of protein (Dill, 1990; Ladbury et al., 1995; Creighton, 1997; Vogt et al., 1997; Kumar et al., 2000). More charged residues are found in hyperthermophilic proteins than in mesophilic proteins, mostly at the expense of uncharged polar residues (in particular Gln), except His having lower frequency in thermophilies (Chakravarty and Varadarajan, 2000; Pack and Yoo, 2004).

An increase in charged residues would provide less labile residues while retains the hydrogen bonding capacity (Sadeghi et al., 2006). Charged residues may be involved in larger numbers of stabilizing ion pairs and networks, although location of ion pairs within molecular structures also appears to be important in determining stability (Xiao and Honig, 1999).

The most striking feature of thermostable proteins, when compared to mesophilic proteins, is the decrease in the number of lysine residues, residues that are mainly replaced with arginine and glutamic acid residues. Higher frequency of Arg was known as the characteristic feature of thermophilic proteins, especially in exposed state, which would stabilize the exposed part of protein structure (Chakravarty and Varadarajan, 2000; Kumar et al., 2000; Das et al., 2006). The reason for this is that Arg δ-guanido moiety has a reduced chemical reactivity due to its high pKa and its resonance stabilization. The δ-guanido moiety provides more surface area for charged interactions than Lys amino group does. Arg has higher tendency to participate in salt-bridge interaction (ion pair) and the salt-bridge involved in Arg shows more stabilizing effect on protein structure. So Arg residues would be better adapted to high temperatures than Lys residues (Mrabet et al., 1992). A comparison between thermophiles and hyperthermophiles shows that the number of charged amino acids (which may include Lys) is very high in hyperthermophiles, whereas thermophiles have a preference of Arg over Lys (Trivedi et al., 2006).

Asp is unstable at high temperatures and therefore its percentage decreases in hyperthermophiles (Szilágyi and Závodszky, 2000). This finding contradicts Tanaka et al. (2004) who report that Asp interactions are critical for thermal stability particularly in hyperthermophiles. Although Asp residues particularly in Asp-Pro combination may be susceptible to hydrolysis of peptide bonds, they get protection by either substitution or by higher conformational rigidity (Vieille and Zeikus, 2001).

But thermophilic proteins showed higher frequency of Glu both in buried and exposed state. Related with another higher frequency of Arg, the higher frequency of Glu could be explained as counter trend for making salt-bridge in thermophilic proteins (Gromiha et al., 1999). Farias and Bonato (2003) found that the $E+K/Q+H$ ratio was >4.5 in hyperthermophiles, <2.5 in the mesophiless, and 3.2–4.6 in thermophiles. The index is used to identify thermophilic proteins (Farias et al., 2004).

Thermal stability of proteins in thermophiles through comprehensive genome comparison, focusing on the occurrence of salt bridges, also showed that thermophiles have a greater content of charged residues than mesophiles, both at the overall genomic level and in alpha helices, and the charged residues in helices tend to be preferentially arranged with a 1–4 helical spacing and oriented so that intra-helical charge pairs agree with the helix dipole (Das and Gerstein, 2000). These results imply that intra-helical salt bridges are more prevalent in thermophiles than mesophiles and thus suggest that they are an important factor stabilizing thermophilic proteins.

Hence, higher ratios for charged amino acids, especially at the protein surface, increase ion interactions and enhanced occurrence of salt bridges and ion pairs in thermophilic proteins which provide thermal stability to proteins (Xiao and Honig, 1999; Szilágyi and Závodszky, 2000; Fukuchi et al., 2003; Nakashima et al., 2003; Saunders

et al., 2003; Suhre and Claverie, 2003; Tanaka et al., 2004). The ratios of these amino acids are important for flexibility or lack of it in proteins (Parthasarathy and Murthy, 2000); they help in tetramerization (L-isoaspartyl-O-methyltransferase from Sulfolobus tokodaii) and are critical for thermal stability (Tanaka et al., 2004). However, a recent study by Sadeghi et al. (2006) showed that there's also an increase in charged amino acid content in some mesophilic proteins. This will eliminate their role as a possible mean to enhance the protein thermostability, and an increase in charged residue content could be an indirect consequence of a decrease in the uncharged polar residues.

Aromatic amino acids

Phenylalanine (Phe, F), Tryptophan (Trp, W), and Tyrosine (Tyr, Y) are aromatic amino acids. They have a benzene-like ring structure in their side chain. Phe is highly hydrophobic and is found buried within globular proteins. Tyr is Phe with an extra hydroxyl (−OH) group that can act as a donor or an acceptor in hydrogen bond, and in consequence has significantly different properties. It is polar and very weakly acidic. Trp is far more like Phe than Tyr. It is a hydrophobic residue with a bulky double rings side chain and tends to be found buried inside globular proteins. The presence of the nitrogen group makes Trp a little less hydrophobic than Phe. They are known as the participants for cation-pi interaction, which is another interaction for maintaining conformational stability in protein structure (Ma and Dougherty, 1997).

The study on 24 pairs of structurally similar thermophilic and mesophilic proteins has shown that the thermophilic proteins have a large number of pair-wise aromatic interactions compared with the mesophilic homologue. The additional aromatic clusters identified in the thermophiles are smaller in size and are largely found on the protein surface. The aromatic clusters are found to be relatively rigid regions of the surface and often the additional aromatic cluster is located close to the active site of the thermophilic enzyme (Kannan and Vishveshwara, 2000). Thermophilic proteins showed higher frequency of Trp in well-buried state (Pack and Yoo, 2004). Hyperthermophilic proteins also contain slightly more aromatic residues than mesophilic proteins do, especially Tyr, while Trp occurs with a similar proportion in both thermophilic and mesophilic chains (Kumar et al., 2000). Due to large side chain, Tyr may be useful both in short range local interactions and in long range interactions.

The differences of amino acids composition between thermophilic and mesophilic proteins mentioned above are related to some properties of amino acids. Gromiha et al. (1999) found that the increase in shape(s) and the decrease in G_{hn} of amino acid increase the thermostability of proteins in 14 families. Shape(s) is location of branch point in side chain of amino acids of β-branched amino acid is 1, γ-branched amino acid is 2; (G_{hn}) is Gibbs free energy change of hydration of amino acid, and the decrease in G_{hn} theoretically increases the stability of protein by increasing the exposure of polar atoms. A good correlation is observed between these two properties. The increase in shape, which tends to increase with increasing number of carbon both for polar and non-polar residues, may generate more packing and compactness. On the other hand, the increase in G_{hn} in thermophilic proteins increases the solubility of the proteins. This can explain that Gly, Ser, Lys, Asp, and Trp in mesophiles are generally substituted by Ala, Thr, Arg, Glu, and Tyr in thermophiles to enhance thermostability.

Amino acid coupling and neighboring patterns in thermophilic proteins

Amino acid composition analysis provides a useful but simplified picture of the relative importance of each individual amino acid type in the thermophilic proteins. However, such analysis overlooks the coupling effects between amino acid types on thermal stability of proteins. To analyze the amino acid coupling patterns in thermophilic protein, a statistical approach was developed by Liang et al. (2005) and this approach is proved to be able to identify and provide a more detailed description of sequence features in thermophilic proteins than the conventional composition analysis.

The amino acid coupling sequence pattern is defined as any 2 types of amino acids separated by 1 or more amino acids. In this pattern, [X*d*Z] denote the amino acid coupling pattern of amino acid types X and Z that are separated by *d* amino acids. The sign of d is determined by the relative positions of X and Z, *d* is defined as positive if X is closer to the N-terminal side and it is defined as negative if X is closer to the C-terminal side (Liang et al., 2005). The amino acid coupling sequence patterns for a data set comprising 74 mesophilic and 15 thermophilic genomes was analyzed and it was found that thermophiles and mesophiles exhibit significant bias in their amino acid coupling patterns as follows:

(1) Patterns [C*d*C], such as [C*3*C], [C*4*C] and [C*7*C], are increased in thermophiles, though Cys show lower frequency in thermophilic proteins. Structural analy-

sis showed that the increased stability of the cysteine clusters is probably due to their involvement in coordination of metal ions such as zinc, iron, or FeS groups, or in disulfide bonds (Rosato et al., 2002).

(2) [*xdE*] and [*xdV*] show similar patterns, specifically for patterns [*KdE*], [*RdE*], [*EdE*], [*DdE*] and [*DdV*] [*KdV*], [*NdV*], and [*YdV*]. Patterns [*KdE*], [*RdE*], [*EdE*] usually occur in helices when $d = 3$. The structural implications of [*xdV*] coupling patterns are not clear, though these patterns frequently occur in α-helices or β-sheets, and a higher proportion of secondary structures is known to be an important contributor to increased thermal stability (Chakravarty and Varadarajan, 2002).

(3) The coupling patterns of [x*d*P] are similar to that of [x*d*C]. Most instances of [x*d*P] are increased in thermophiles, especially for [CdP] and [PdP]. [PdP]s (or proline clusters) are often involved in the formation of the polyproline II helix. The helical conformation, together with the reduced conformational entropy, may contribute to protein stability (Liang et al., 2005).

(4) The coupling patterns involving polar amino acids are usually decreased in thermophiles, such as [xdQ], [xdT], and [xdH]. Though Glu is usually increased in thermophiles, the coupling pattern [E3T] is in fact decreased in thermophiles. It was interesting to find that [(charged residue)dH] is also significantly decreased in thermophilic proteins.

(5) Other coupling patterns: [xdL] does not show any significant bias toward thermophiles but for [CdL], which is decreased in thermophiles with statistical significance. [xdI] is increased in thermophiles. For the patterns involving aromatic amino acids, [xdF] and [xdW] are decreased in thermophilic proteins, but [xdY] is increased. For patterns involving charged amino acids, [xdE], [xdK], and [xdR] are increased in thermophilic proteins, but interestingly, [xdD] is decreased. For patterns involving polar amino acids, [xdS] and [xdN] are in general decreased in thermophilic proteins. The profile of [xdA] pattern is similar to that of [xdN] and is decreased in thermophiles, despite the fact that alanine and asparagine are 2 different types of amino acids.

The net thermal stability of proteins usually results from a multitude of different coupling patterns, and no single outstanding sequence or structural feature can adequately account for thermophilic proteins. The following sets are denoted which contains the following thermophilic amino acid coupling patterns: [C(2)P], [C1P], [C3C], [C4C], [C6C], [C7C], [K(7)E], [K(4)E], [K3E], [K4E], and [H(4)V], and the following mesophilic amino acid coupling patterns: [C(4)L], [C(3)L], [C(2)L], [C2L], [C3L], [D(5)T], [D(4)T], [E(8)T], [E(4)T], [E1Q], [E3T], [E4T], [G(3)Q], [K(4)T], [K2T], and [K3T] (Liang et al., 2005).

The pattern based on dipeptide composition patterns can also be used for discrimination of thermophilic and mesophilic proteins. Zhang and Fang (2006a, b) systematically analyzed the distribution of two neighboring amino acids in the sequences of thermophilic and mesophilic proteins. It was observed that the occurrence of EE, KK, RR, PP, KI, VV, VE, KE and VK, which were apt to located in the α-helix, were significantly higher in thermophilic proteins, while the occurrence of QQ, AA, EQ, LL, QA, QL, NN, KQ, QG, RQ, QT and AQ, which were apt to located in the β-sheet turn or the coil, were significantly lower. This maybe because α-helix might increase the rigidity of a protein and protect it against unfolding, while the loop-deletion would enhance the thermostability of a protein (Thompson and Eisenberg, 1999).

However, there is no single outstanding pattern which can adequately account for protein thermophily, applicable to distinguish between thermophiles and mesophiles in combination with the amino acids composition analysis and different amino acids coupling or neighboring patterns.

Amino acid preference in different thermophilic proteins

Compared to the amino acids differences among different thermophilic proteins, amino acid preference exists and this preference may vary from protein to protein within an organism or may be taxon specific (Kawashima et al., 2000; Trivedi et al., 2006). For example, Ile is preferred in *Methanococcus* (McDonald et al., 1999) and *Picrophilus torridus* (Futterer et al., 2004), whereas Val or Gly is preferred over Ile in others. A comparative study between hyperthermophilic Archaea *Methanococcus jannaschii* and thermophilic Eubacteria *Bacillus stearothermophilus* (with OGT 85 and 60 °C, respectively) shows that Lys and Tyr is preferred over Arg in *M. jannaschii* but Gly is preferred over Ile and Ala over Tyr in *B. stearothermophilus* (McDonald et al., 1999). Almost universal preference for Glu and Lys over Gln and His in hyperthermophiles is reported (De Farias and Bonato, 2002; Farias and Bonato, 2003). It is apparent that the variations in preference for other amino acids between mesophiles, thermophiles and hyperthermophiles are not only organism specific but are also protein specific within the organism (Trivedi et al., 2006).

Conclusions

Compared to mesophilic proteins, the change of amino acid composition in thermophilic proteins can be categorized into six distinct aspects: (1) an increase in non-polar amino acids, especially hydrophobic and Pro residues which contribute to the hydrophobic interactions; (2) an increase in charged amino acids, especially Arg and Glu residues which contribute to the ionic interactions; (3) an increase in aromatic amino acids, especially Tyr residue which contribute to the cation-pi interactions; (4) a decrease in Met and uncharged polar residues which are thermoliable amino acids; (5) certain amino acid coupling patterns; and (6) amino acid preference should be considered for thermophilic proteins from different thermophiles. These features of amino acid composition are related to some properties of amino acids, especially shape (s) and G_{hn} of amino acid.

However, amino acid composition difference is only the fundamental factor of the factors that support the protein thermostability. Other structural factors including additional hydrogen bonds, additional electrostatic interactions, additional hydrophobic interactions, additional disulfide bonds, more rigid, compact packing, conformational strain release, stability of α-helix, reduction of the entropy of unfolding, are important (Vieille and Zeikus, 2001). Post-translational modifications of amino acids could also affect protein stability through changing the characteristic of amino acid (Cacciatore et al., 2005; Gredicak et al., 2007). Glycosylation and lysyl methylation showed stabilizing effects as post-translational modifications (Olsen and Thomsen, 1991; Febbraio et al., 2004; Eichler and Adams, 2005; D'Auria et al., 2006). It is suggested that prior to translation there may be certain mechanisms prevailing in these organisms that would provide charges to uncharged amino acids or vice versa, but possibly such changes remain undetected in the in vitro analysis (Trivedi et al., 2006; Bartesaghi et al., 2007).

Still, the factors mentioned above show complicated correlations among each other, and it is also necessary to notice the inconsistency of findings from different studies. Additional information on differences of amino acids composition between thermophilic and mesophilic proteins and its mechanism surely await discovery as more investigations being done and more proteins sequences available for comparison (Schmidinger et al., 2006; Vercauteren et al., 2006; Kuric, 2007). This information will help us to engineer mesophilic proteins to thermophilic proteins using a more ''rational'' approach.

Acknowledgements

This review was partly supported by National Natural Science Foundation of China (NSFC, Project 30000118 and 30470021). We also acknowledge valuable help provided by our colleagues for their many helpful suggestions.

References

Argos P, Rossmann MG, Grau UM, Suborn H, Frank G, Tratschin JD (1979) Thermal stability and protein structure. Biochemistry 18: 5698–5703

Bartesaghi S, Ferrer-Sueta G, Peluffo G, Valez V, Zhang H, Kalyanaraman B, Radi R (2007) Protein tyrosine nitration in hydrophilic and hydrophobic environments. Amino Acids 32: 501–515

Beeby M, O'Connor BD, Ryttersgaard C, Boutz DR, Perry LJ, Yeates TO (2005) The genomics of disulfide bonding and protein stabilization in thermophiles. PLoS Biol 3: e309

Britton KL, Baker PJ, Borges KM, Engel PC, Pasquo A, Rice DW, Robb FT, Scandurra R, Satillman TJ, Yip KS (1995) Insights into thermalstability from a comparison of the glutamate dehydrogenases from Pyrococcus furiosus and Thermococcus litoralis. Eur J Biochem 229: 688–695

Cacciatore I, Cocco A, Costa M, Fontana M, Lucente G, Pecci L, Pinnen F (2005) Biochemical properties of new synthetic carnosine analogues containing the residue of 2,3-diaminopropionic acid: the effect of N-acetylation. Amino Acids 28: 77–83

Catanzano F, Graziano G, Capasso S, Barone G (1997) Thermodynamic analysis of the effect of selective monodeamidation at asparagine 67 in ribonuclease A. Protein Sci 6: 1682–1693

Chakravarty S, Varadarajan R (2000) Elucidation of determinants of protein stability through genome sequence analysis. FEBS 470: 65–69

Chakravarty S, Varadarajan R (2002) Elucidation of factors responsible for enhanced thermal stability of proteins: a structural genomics based study. Biochemistry 41: 8152–8161

Creighton TE (1997) Proteins: structures and molecular properties, 2nd ed. Freeman & Company, New York, pp 1–20

Das R, Gerstein M (2000) The stability of thermophilic proteins: a study based on comprehensive genome comparison. Funct Integr Genomics 1: 76–88

Das S, Paul S, Bag SK, Dutta C (2006) Analysis of nanoarchaeum equitans genome and proteome composition: indications for hyperthermophilic and parasitic adaptation. BMC Genomics 7: 186

D'Auria S, Ausili A, Marabotti A, Varriale A, Scognamiglio V, Staiano M, Bertoli E, Rossi M, Tanfani F (2006) Binding of glucose to the d-galactose/d-glucose-binding protein from Escherichia coli restores the native protein secondary structure and thermostability that are lost upon calcium depletion. J Biochem (Tokyo) 139: 213–221

De Farias ST, Bonato MC (2002) Preferred codons and amino acid couples in hyperthermophiles. Genome Biol 3: 1–18

Denisov VP, Venu K, Peters J, Hörlein HD, Halle B (1997) Orientational disorder and entropy of water in protein cavities. J Phys Chem 101: 9380–9389

Diao Y, Ma D, Wen Z, Yin J, Xiang J, Li M (2007) Using pseudo amino acid composition to predict transmembrane regions in protein: cellular automata and Lempel-Ziv complexity. Amino Acids [Epub ahead of print]

Dill KA (1990) Dominant forces in protein folding. Biochem 29: 7133–7155

Eichler J, Adams MWW (2005) Posttranslational protein modification in archaea. Microbiol Mol Biol Rev 69: 393–425

Estell DA, Graycar TP, Wells JA (1985) Engineering an enzyme by site-directed mutagenesis to be resistant to chemical oxidation. J Biol Chem 260: 6518–6521

Farias ST, Bonato MC (2003) Preferred amino acids and thermostability. Genet Mol Res 2: 383–393

Farias ST, Van Der Linden MG, Rego TG, Araujo DA, Bonato MC (2004) Thermo-search: lifestyle and thermostability analysis. In Silico Biol 4: 30

Febbraio F, Andolfo A, Tanfani F, Briante R, Gentile F, Formisano S, Vaccaro C, Scire A, Bertoli E, Pucci P, Nucci R (2004) Thermal stability and aggregation of sulfolobus solfataricus {beta}-glycosidase are dependent upon the N-{epsilon}-methylation of specific lysyl residues: critical role of in vivo post-translational modifications. J Biol Chem 279: 10185–10194

Fields PA (2001) Review: protein function at thermal extremes: balancing stability and flexibility. Comp Biochem Physiol A 129: 417–431

Fukuchi S, Yoshimune K, Wakayama M, Moriguchi M, Nishikawa K (2003) Unique amino acid composition of proteins in halophilic bacteria. J Mol Biol 327: 347–357

Futterer O, Angelov A, Liesegang H, Gottschalk G, Schleper C, Schepers B, Dock C, Antranikian G, Liebl W (2004) Genome sequence of Picrophilus torridus and its implications for life around pH 0. Proc Natl Acad Sci USA 101: 9091–9096

Gredicak M, Kolonic A, Jeric I (2007) Novel chloroenyne-modified amino acid derivatives. Amino Acids Jul 6 [Epub ahead of print]

Gromiha MM, Oobatake M, Sarai A (1999) Important amino acid properties for enhanced thermostability from mesophilic to thermophilic proteins. Biophys Chem 82: 51–67

Haney PJ, Badger JH, Buldak GL, Reich CI, Woese CR, Olsen GJ (1999) Thermal adaptation analyzed by comparison of protein sequences from mesophilic and extremely thermophilic Methanococcus species. Biochem 96: 3578–3583

Haslbeck M, Franzmann T, Weinfurtner D, Buchner J (2005) Some like it hot: the structure and function of small heat-shock proteins. Nat Struct Mol Biol 12: 842–846

Ikai A (1980) Thermostability and aliphatic index of globular proteins. J Biochem (Tokyo) 88: 1895–1898

Jaenicke R (1996) Stability and folding of ultrastable proteins: eye lens crystalline and enzymes from thermophiles. FASEB J 10: 84–92

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

Kannan N, Vishveshwara S (2000) Aromatic clusters: a determinant of thermal stability of thermophilic proteins. Protein Eng 13: 753–761

Kawashima T, Amano A, Koike A, Makino S, Higuchi S, Kawashima-Ohya Y, Watanabe K, Yamazaki M, Kanehori K, Kawamoto K, Nunoshiba T, Yamamoto Y, Aramaki H, Makino K, Suzuki M (2000) Archaeal adaptation to higher temperatures revealed by genomic sequence of thermoplasma canium. Proc Natl Acad Sci USA 97: 14257–14262

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

Kumar S, Tsai CJ, Nussinov R (2000) Factors enhancing protein thermostability. Protein Eng 13: 179–191

Kuric L (2007) The digital language of amino acids. Amino Acids 2007 Jan 26 [Epub ahead of print]

Ladbury JE, Wynn R, Thomson JA, Sturtevant JM (1995) Substitution of charged residues into the hydrophobic core of *Escherichia coli* thioredoxin result in a change in heat capacity of the native protein. Biochem 34: 2148–2152

Li WF, Zhou XX, Lu P (2005) Structural features of thermozymes. Biotechnol Adv 23: 271–281

Liang HK, Huang CM, Ko MT, Hwang JK (2005) The amino acid-coupling patterns in thermophilic proteins. Proteins Structure Function Bioinformatics 59: 58–63

Lin W, Chan M, Goh LL, Sim TS (2007) Molecular basis for thermal properties of Streptomyces thermovulgaris fumarase C hinge at hydrophilic amino acids R163, E170 and S347. Appl Microbiol Biotech 75: 329–335

Lopez-Camacho C, Salgado J, Lequerica JL, Madarro A, Ballestar E, Franco L, Polaina J (1996) Amino acid substitutions enhancing thermostability of Bacillus polymyxa beta-glucosidase A. Biochem J 314: 833–838

Lu B, Wang G, Huang P (1998) A comparison of amino acid composition of proteins from thermophiles and mesophiles. Wei Sheng Wu Xue Bao (Acta Microbiologica Sinica) 38: 20–25 (in Chinese)

Ma JC, Dougherty DA (1997) The cation-pi interaction. Chem Rev 97: 1303–1324

Mattos C (2002) Protein-interactions in a dynamic world. Trends Biochem Sci 27: 203–208

McDonald JH, Grasso AM, Rejto LK (1999) Patterns of temperature adaptation in proteins from Methanococcus and Bacillus. Mol Biol Evol 16: 1785–1790

Mrabet NT, Van den Broeck A, Van den Brande I, Stanssens P, Laroche Y, Lambeir AM, Matthijssens G, Jenkins J, Chiadmi M, van Tilbeurgh H, Rey F, Janin J, Quax WJ, Lasters I, De Maeyer M, Wodak SJ (1992) Arginine residues as stabilizing elements in proteins. Biochem 31: 2239–2253

Nagendra HG, Sukumar N, Vijayan M (1998) Role of water in plasticity, stability, and action of proteins: the crystal structures of lysozyme at very low levels of hydration. Proteins 32: 229–240

Nakashima H, Fukuchi S, Nishikawa K (2003) Compositional changes in RNA, DNA and proteins for bacterial adaptation to higher and lower temperatures. J Biochem 133: 507–513

Olsen O, Thomsen KK (1991) Improvement of bacterial β-glucanase thermostability by glycosylation. J Gen Microbiol 137: 579–585

Pack SP, Yoo YJ (2004) Protein thermostability: structure-based difference of amino acid between thermophilic and mesophilic proteins. J Biotechnol 111: 269–277

Pack SP, Yoo YJ (2005) Packing-based difference of structural features between thermophilic and mesophilic proteins. Int J Biol Macrom 35: 169–174

Panasik N, Brenchley JE, Farber GK (2000) Distributions of structural features contributing to thermostability in mesophilic and thermophilic alpha/beta barrel glycosyl hydrolases. Biochim Biophys Acta 1543: 189–201

Pantazaki AA, Tsolkas GP, Kyriakidis DA (2007) A DING phosphatase in Thermus thermophilus. Amino Acids [Epub ahead of print]

Parthasarathy S, Murthy MR (2000) Protein thermal stability: insights from atomic displacement parameters (B values). Protein Eng 13: 9–13

Perry LJ, Wetzel R (1986) Unpaired cysteine-54 interferes with the ability of an engineered disulfide to stabilize T4 lysozyme. Biochem 25: 733–739

Ponnuswamy PK, Muthusamy R, Manavalan P (1982) Amino acid composition and thermal stability of globular proteins. Int J Biol Macromol 4: 186–190

Rosato V, Pucello N, Giuliano G (2002) Evidence for cysteine clustering in thermophilic proteomes. Trends Genet 18: 178–281

Russell RJ, Ferguson JM, Hough DW, Danson MJ, Taylor GL (1997) The crystal structure of citrate synthase from the hyperthermophilic archaeon Pyrococcus furiosus at 1.9 Å resolution. Biochemistry 36: 9983–9994

Russell RJ, Hough DW, Danson MJ, Taylor GL (1994) The crystal structure of citrate synthase from the thermophilic archaeon thermoplasma acidophilum. Structure 2: 1157–1167

Sadeghi M, Naderi-Manesh H, Zarrabi M, Ranjbar B (2006) Effective factors in thermostability of thermophilic proteins. Biophys Chem 119: 256–270

Saunders NFW, Thomas T, Curmi PM, Mattick JS, Kuczek E, Slade R, Davis J, Franzmann PD, 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

Scandurra R, Consalvi V, Chiaraluce R, Politi L, Engel PC (1998) Protein thermostatbility in extremophiles. Biochimie 80: 933–941

Schmidinger H, Hermetter A, Birner-Gruenberger R (2006) Activity-based proteomics: enzymatic activity profiling in complex proteomes. Amino Acids 30: 333–350

Schumann J, Bohm G, Schumacher G, Rudolph R, Jaenicke R (1993) Stabilization of creatinase from Pseudomonas putida by random mutagenesis. Protein Sci 2: 1612–1620

Singer GAC, Hickey DA (2003) Thermophilic prokaryotes have characteristic patterns of codon usage, amino acid composition and nucleotide content. Gene 317: 39–47

Stetter KO (1996) Hyperthermophilic prokaryotes. FEMS Microbiol Rev 18: 149–158

Suhre K, Claverie JM (2003) Genomic correlates of hyperthermostability, an update. J Biol Chem 278: 17198–17202

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

Tanaka Y, Tsumoto K, Yasutake Y, Umetsu M, Yao M, Fukada H, Tanaka I, Kumagai I (2004) How oligomerization contributes to the thermostability of an archaeon protein. Protein l-isoaspartyl-O-methyltransferase from Sulfolobus tokodaii. J Biol Chem 279: 32957–32967

Tekaia F, Yeramian E, Dujon B (2002) Amino acid composition of genomes, lifestyles of organisms, and evolutionary trends: a global picture with correspondence analysis. Gene 297: 51–60

Thompson MJ, Eisenberg D (1999) Transproteomic evidence of a loop-deletion mechanism for enhancing protein thermostability. J Mol Biol 290: 595–604

Tomazic SJ, Klibanov AM (1988) Mechanisms of irreversible thermal inactivation of Bacillus alpha-amylases. J Biol Chem 263: 3086–3091

Trivedi S, Gehlot HS, Rao SR (2006) Protein thermostability in Archaea and Eubacteria. Genet Mol Res 5: 816–827

Van den Burg B, Vriend G, Veltman OR, Venema G, Eijsink VGH (1998) Engineering an enzyme to resist boiling. Proc Natl Acad Sci USA 95: 2056–2060

Veltman OR, Vriend G, Middelhoven PJ, Van Den Burg B, Venema G, Eijsink VGH (1996) Analysis of structural determinants of the stability of thermolysinlike proteases by molecular modelling and site-directed mutagenesis. Protein Eng 9: 1181–1189

Vercauteren FG, Arckens L, Quirion R (2006) Applications and current challenges of proteomic approaches, focusing on two-dimensional electrophoresis. Amino Acids Nov 30 [Epub ahead of print]

Vetriani C, Maeder DL, Tolliday N, Yip KS, Stillman TJ, Britton KL, Rice DW, Klump HH, Robb FT (1998) Protein thermostability above 100 degrees C: a key role for ionic interactions. Proc Natl Acad Sci USA 95: 12300–12305

Vieille C, Zeikus GJ (2001) Hyperthermophilic enzymes: sources, uses, and molecular mechanisms for thermostability. Microbiol Mol Biol Rev 65: 1–43

Vogt G, Woell S, Argos P (1997) Protein thermal stability, hydrogen bonds, and ion pairs. J Mol Biol 269: 631–643

Watanabe K, Hata Y, Kizaki H, Katsube Y, Suzuki Y (1997) The refined crystal structure of Bacillus cereus oligo-1,6-glucosidase at 2.0 Å resolution: structural characterization of proline-substitution sites for protein thermostabilization. J Mol Biol 269: 142–153

Xiao L, Honig B (1999) Electrostatic contributions to the stability of hyperthermophilic proteins. J Mol Biol 289: 1435–1444

Xu Z, Liu Y, Yang Y, Jiang W, Arnold E, Ding J (2003) Crystal structure of D-Hydantoinase from Burkholderia pickettii at a resolution of 2.7 Angstroms: insights into the molecular basis of enzyme thermostability. J Bacteriol 185: 4038–4049

Yokota K, Satou K, Ohki SY (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. Sci Tech Adv Mat 7: 255–262

Zhang G, Fang B (2006a) Application of amino acid distribution along the sequence for discriminating mesophilic and thermophilic proteins. Proc Biochem 41: 1792–1798

Zhang G, Fang B (2006b) Discrimination of thermophilic and mesophilic proteins via pattern recognition methods. Proc Biochem 41: 552–556

Authors' addresses: Yuan-Jiang Pan, Institute of Chemical Biology and Pharmaceutical Chemistry, Zhejiang University, Zhejiang University Road 38, Hangzhou 310027, China,
Fax: +86-571-8795-3000, E-mail: panyuanjiang@zju.edu.cn
Wei-Fen Li, Microbiology Division, College of Animal Science, Zhejiang University, Hangzhou 310029, China,
Fax: +86-571-8699-4963, E-mail: wfli@zju.edu.cn