

Amino acid contacts in proteins adapted to different temperatures: hydrophobic interactions and surface charges play a key role

Gisle Sælensminde · Øyvind Halskau Jr ·
Inge Jonassen

Received: 11 March 2008 / Accepted: 4 September 2008 / Published online: 30 September 2008
© The Author(s) 2008. This article is published with open access at Springerlink.com

Abstract Thermophiles, mesophiles, and psychrophiles have different amino acid frequencies in their proteins, probably because of the way the species adapt to very different temperatures in their environment. In this paper, we analyse how contacts between sidechains vary between homologous proteins from species that are adapted to different temperatures, but displaying relatively high sequence similarity. We investigate whether specific contacts between amino acids sidechains is a key factor in thermostabilisation in proteins. The dataset was divided into two subsets with optimal growth temperatures from 0–40 and 35–102°C. Comparison of homologues was made between low-temperature species and high-temperature species within each subset. We found that unspecific interactions like hydrophobic interactions in the core and solvent interactions and entropic effects at the surface, appear to be more important factors than specific contact types like salt bridges and aromatic clusters.

Keywords Thermophiles · Psychrophiles · Cold-adaptation · Stability · Amino acid frequency · Protein structure

Abbreviations

OGT Optimal growth temperature
SASA Solvent accessible surface area

Introduction

Life has a great ability to adapt to the most diverse habitats, and almost all places on earth have a variety of life forms. One of the most remarkable adaptations is the range of temperatures life has adapted to. Life can be found in arctic and antarctic waters holding temperatures below the ordinary freezing point of water, to superheated water in hot springs and in deep ocean smokers, where some organisms can survive temperatures up to 121°C (Kashefi and Lovley 2003). Despite this, the basic building blocks of DNA and proteins are the same, and a pair of homologous proteins from one psychrophilic and one thermophilic species may have almost identical structure. This is generally interpreted to mean that there must be more subtle differences in interactions and packing that make up most of the differences between them (Kumar et al. 2000).

Thermophilic proteins have been the most studied ones for several reasons. First, they are stable at temperatures where most proteins get denature, and have therefore been attractive to use in order to understand what makes proteins stable in general. Because they are easier to purify, and probably also to crystallise, they have been picked out among homologues when a representative protein for a family was to be picked, especially in structural genomics

Communicated by T. Matsunaga.

Electronic supplementary material The online version of this article (doi:10.1007/s00792-008-0192-4) contains supplementary material, which is available to authorized users.

G. Sælensminde (✉) · I. Jonassen
Computational Biology Unit (CBU), BCCS,
University of Bergen, Bergen, Norway
e-mail: gisle@cbu.uib.no

G. Sælensminde · I. Jonassen
Department of Informatics, University of Bergen,
Bergen, Norway

Ø. Halskau Jr
Department of Biomedicine, University of Bergen,
Bergen, Norway

studies (Jenney and Adams 2008). Proteins from psychrophiles are likewise thought to be more difficult to purify and crystallise, which likely is the reason why there are relatively few structures of cold adapted proteins, while there are more sequences of these than for thermophiles in the genomic databases.

Many studies of thermostabilisation have been done on the structure level. Of the earliest factors mentioned is salt bridges, and a number of groups have studied their stabilising effects (Zhou and Dong 2003; Thomas and Elcock 2004; Elcock 1998). Das and Gerstein (2000) found that thermophilic proteins contain considerably more charged residues, and likely also more salt bridges, than mesophiles. Cation- π bonds that form between the positively charged residues arginine and lysine and the aromatic residues, particularly Tyr and Trp, have been suggested to be of importance in thermostabilisation by Chakravarty and Varadarajan (2002) as well as Gromiha et al. (2002). Kannan and Vishveshwara (2000) suggested that aromatic clusters are important in thermostabilisation.

Disulfide bridges bind two cysteine residues in the peptide chain with a covalent bond. This reduces the freedom of movement and therefore stabilises the protein. Traditionally it has been thought that disulfide bridges only form in extracellular proteins, since cytosolic conditions would reduce the disulfide bonds (Branden and Tooze 1999). This view has recently been challenged by Beeby et al. (2005), who claim to have found a mechanism for forming and stabilization of disulfide bonds within some thermophilic organisms.

The amino acid frequency is also known to be different in psychrophiles, mesophiles and thermophiles. Often these can be approximated as a near linear function of temperature as shown by Nakashima et al. (2003), for all amino acids. Zeldovich et al. (2007) found that the frequency of the amino acids Ile, Val, Tyr, Trp, Arg, Glu and Leu are the strongest predictors of optimal growth temperature for bacterial species. The trends in amino acid distribution are very different, and often opposite, between the core and on the surface of proteins. Saelensminde et al. (2007) found that in proteins, in species adapted to higher temperatures, the core is getting more hydrophobic, while the surface is getting less hydrophobic. In particular, for thermophiles, the surfaces are more charged, and especially positively charged residues are more common.

For psychrophiles, there are fewer studies. Reasons for this include that there are fewer industrial applications so far, and also that fewer protein structures have been solved. Most of the studies have focused on single proteins and families. Although proteins may cold denature, the main challenge of cold-adapted enzymes is not stability, but efficiency. The Arrhenius equation tells that the turnover rate for enzyme-catalysed reactions reduces exponentially

with decreasing temperature. For the enzymes to retain efficiency sufficient to maintain life, they must be optimised to a higher turnover rate. For reviews of cold-adapted enzymes, see Feller and Gerday (2003), Siddiqui and Cavicchioli (2006) and Feller and Georges (2007).

Increased flexibility around the active site is very often stated as the key mechanism for the higher turnover rate at low temperature, and some studies into this have been performed. Olufsen et al. (2005) studied several cold-adapted proteins and mesophilic homologs by MD-simulations, and could see increased flexibility in the cold adapted proteins known to be more efficient at low temperatures. Less seems to be known about how this flexibility is achieved on a sequence or structural level.

In this paper, we study whether there are any differences in amino acid contacts when comparing proteins adapted to different temperatures. Different means of adaptation have been suggested, many of them involving or implying changes in the number of certain amino acid contacts. This paper reports a systematic study using a comprehensive dataset followed by a discussion of which of the hypotheses reported in the literature are supported by our analyses.

Methods

Sequence data and temperature data

All sequences were retrieved from the UniProt database in October 2006 (The UniProt Consortium 2008). We used only those sequences originating from species with an entry in the Prokaryotic growth temperature database (PGTDB) (Huang et al. 2004), database containing optimal growth temperature information for around 1,000 prokaryotic species. While PGTDB gives an interval of temperatures for each species, for simplicity we used the midpoint of this interval. PGTDB specifies one or both of optimal growth temperature and common laboratory growth conditions. In the case optimal growth temperature was known, we used this. Since both PGTDB and UniProt identifies species using the NCBI taxonomy ID (Benson et al. 2000; Wheeler et al. 2000), this ID was used to match sequences and species. This resulted in around 380,000 sequences each associated with a growth temperature T_{opt} .

For our analyses, we made two datasets. In the first dataset (the psychrophile–mesophile set), we included all sequences having T_{opt} below 40°. The other set (the mesophile–thermophile set) contained all sequences having T_{opt} above 35°. The sequences from each of these sets were assembled into clusters by the program cd-hit (Li and Godzik 2006), using a cutoff at 60% identity for the clustering. We got 47,067 clusters for the psychrophile set and

among these 963 with a known pdb structure (see below). For the thermophile set, we got 34,574 clusters, out of which 353 had at least one known structure. In each cluster, we picked the sequence pair with the biggest temperature difference, requiring the difference to be at least 20°. We also avoided pairs where one sequence in the pair was less than 70% of the length of the other, since pairs with very different lengths are more likely to be misaligned. When all this was done, we had 148 mesophile–thermophile pairs and 368 psychrophile–mesophile pairs, which formed the dataset analysed in this paper.

Protein structure and amino acid contacts

Since we wanted to find how amino acid contacts change with temperature, we studied only those clusters where at least one structure was known. To find structures we used the cross references to PDB (Berman et al. 2000) in UniProt. We then aligned all the protein sequences in each cluster using MUSCLE (Edgar 2004), including the sequence from the relevant chain in the PDB file. Subsequently, amino acid contacts were determined using a method similar to the half-sphere method of Hamelryck (2005), but based on the coordinates of the carbon-beta atoms rather than those of the carbon alpha. In this method, two amino acids are said to be in contact if the C_β of the second amino acid is inside a half-sphere with the C_β of the first amino acid in the center, pointing away from the C_α . We used a radius of 6Å, and defined contacts as mutual by saying that if one amino acid was found to be in contact with another, the reverse was also true.

We assume that if two amino acids are in contact in the structure analysed, the amino acids in the same columns in the multiple alignment are also in contact with each other. Since all proteins in each cluster are at least 60% identical to each other. That amino acid contact is preserved has been shown in Kleinjung et al. (2004).

Contact substitution matrices

We will use the term contact substitution matrix to denote a matrix used to keep track of differences between aligned sequences taking into account contact patterns. If a Leu is aligned with a Val at a residue position in contact with a Phe (no substitution at this site), then this will be contribute a count to the field (Leu:Phe, Val:Phe) in the contact substitution matrix. Here, Leu:Phe means a Leu in contact with a Phe, and (Leu:Phe, Val:Phe) means that a leucine in contact with a phenylalanine has been changed to a valine in contact with a Phenylalanine. This results in a 400×400 matrix (or alternatively a hypercube of 20^4 values) recording all the changes in amino acids contacts over the set of sequence pairs analyzed.

Temperature ranges and solvent exposure

The amino acid distribution and frequency changes due to temperature are quite different between the exposed and the buried residues of protein. In order to see which effects those are most important on the surface and in the core of the proteins, we studied the amino acids that were exposed more and less than 25% separately. The surface exposure for each amino acid was found using the DSSP program (Kabsch and Sander 1983) on the homologous structure to find the absolute surface area. To find the relative surface area, we divided the surface size on the exposed surface of a configuration of the amino acid in a tripeptide Ala-Xxx-Ala in a helix conformation.

Amino acid groups

The naturally occurring amino acids were divided into groups sharing similar biophysical properties (Table 1). The division into groups are inspired by the Venn diagram of Taylor (1986), and similar to that of Sælensminde et al. (2007). Since we studied amino acid contacts, we changed the grouping slightly by putting cysteine in a separate group since it can form disulphide bridges, thought to enhance thermostability. Histidine was also placed in a separate category. We also made a category of Tyr and Trp, since these amino acids are important for forming connective networks, as have been suggested by Kannan and Vishveshwara 2000 and Brinda and Vishveshwara et al. 2005.

Changes in amino acid contacts

The key question in this study is how the contacts differ between proteins adapted to different temperatures. Some changes are suggested by the changes in amino acid frequencies, but it is possible that some important mechanisms

Table 1 Amino acid groups by biophysical properties

Property	Amino acids
Nonpolar	Val, Ile, Leu, Met, Phe, Ala
Aromatic	Phe, His, Trp, Tyr
Trp + Tyr	Trp, Tyr
Positive	Lys, Arg
Negative	Asp, Glu
Polar	Ser, Thr, Asn, Gln, Cys
Small	Gly, Cys, Ser, Ala, Thr, Val
Long	Lys, Arg, Met, Glu, Gln
Proline	Pro
Glycine	Gly
Cysteine	Cys
Histidine	His

of adaptation are not visible by frequency change alone. It is believed that some pairs of amino acids are stabilising when in contact, for example by forming salt bridges. Other amino acid are believed to be stabilising or destabilising not because of contacts with other amino acids, but, e.g. their rigidity, or because they form favorable interactions with the solvent.

To see how the contacts differed between the proteins adapted to higher or lower temperatures, we counted the mutations where the amino acid belongs to one group in the coldest adapted sequence and not in the hottest adapted sequences and vice versa. If the number of contacts is not depending on temperature, the numbers should be approximately the same, and should thus be accordance with a binomial distribution with $P = 0.5$. We did this test on contacts between all pairs of groups, with two cutoffs of P value on 0.001 and 0.01, respectively.

This analysis does not take into the account that one would expect the number of contacts to change when the frequencies of the involved amino acids change. This is hard to analyse statistically, because the probability of a mutation is dependent of the environment around the amino acid, and it is an open issue how to test this in an unbiased way. However, this effect should be taken into account when interpreting the results.

Results

Figures 1, 2, 3, 4 show the result of the contact analyses. Each amino acid group is represented as a circle, where the circle size represents the frequency of the group. If there is a significant change in frequency of that group (P value < 0.001) in the sequences adapted to the highest

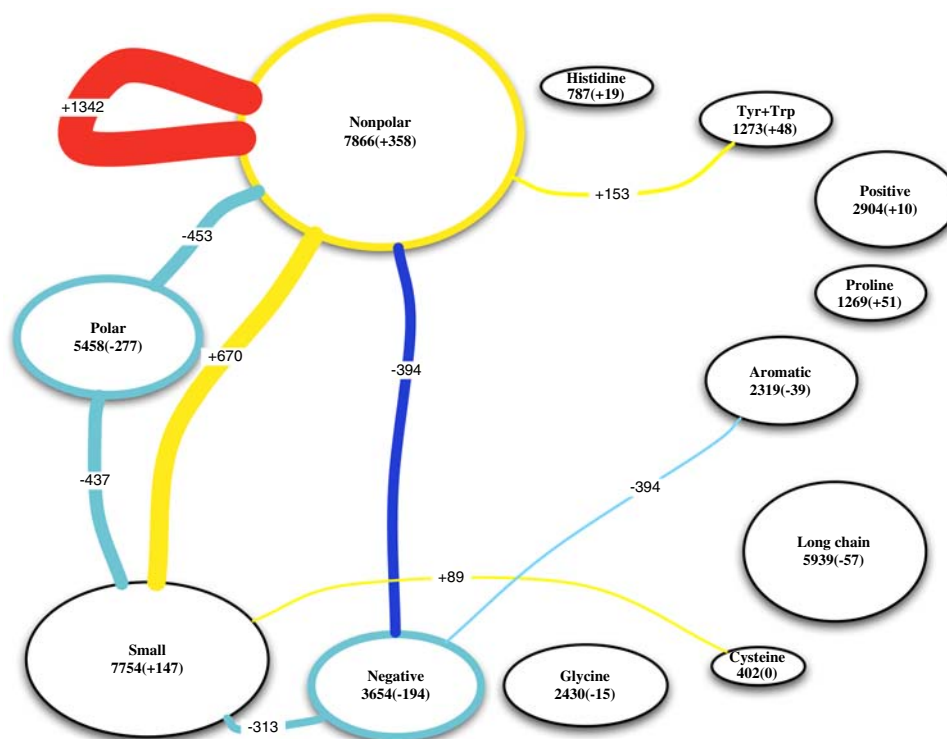


Fig. 1 Changes in contacts between classes of amino acids from psychrophiles to mesophiles, buried residues only. Each circle represents an amino acid class as described in Table 1. The size of the circle is bigger the more frequent the group of amino acids is. The numbers below the name of the amino acid class in the circles is the number of amino acids in this class in the coldest adapted sequences, while the number in parentheses is the net difference in the number of amino acids in homologous positions in the proteins adapted to the highest temperature. There is a line between two groups if there is a significant change in the number of contacts between two amino acid classes when going from a low to high temperature protein. The line is colored red if there is a significant increase in such contacts with a

P value of 0.001 or less, or yellow it increases with a P value of 0.01. In the case of a significant loss of contacts from the proteins adapted to the coldest to the hotter environment is marked with a blue line. If there is a less significant increase or decrease of contacts between a pair of groups ($P \leq 0.01$) the line is coloured light blue. The number of contacts that changed is printed on each line connecting two groups, and the magnitude of the difference is also indicated by the thickness of the line, so that thick lines indicates the biggest changes. Similarly, if there is a significant increase or decrease in the number of the amino of an amino acid group, the circles representing the group are coloured using the same color scheme as for the lines between the circles

Fig. 2 Changes in contacts between classes of amino acids from psychrophiles to mesophiles, exposed residues only. See caption of Fig. 1 for details

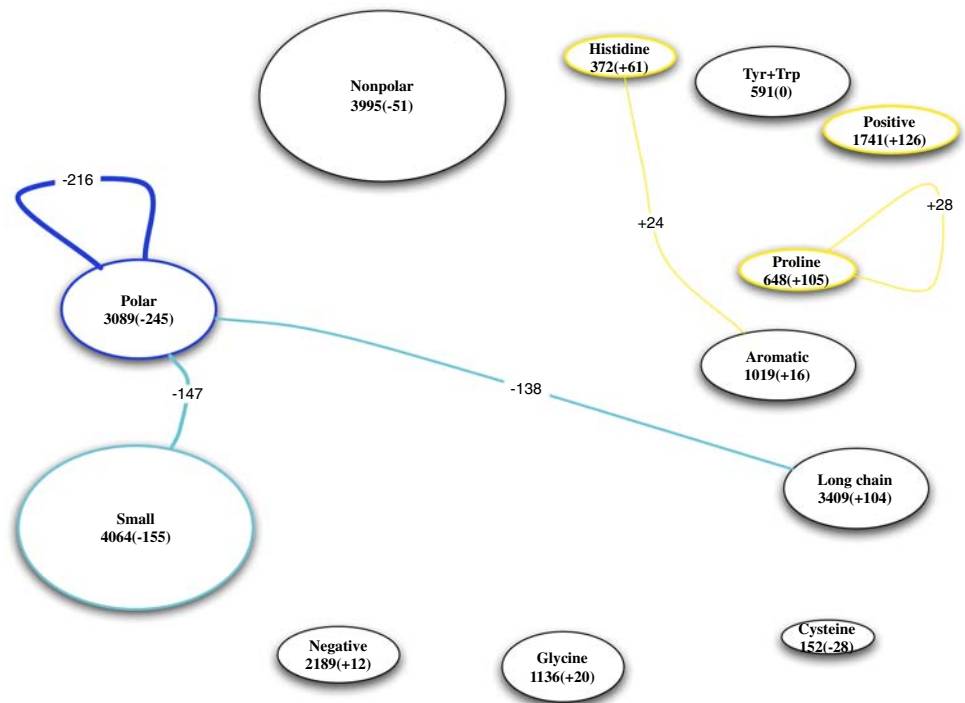
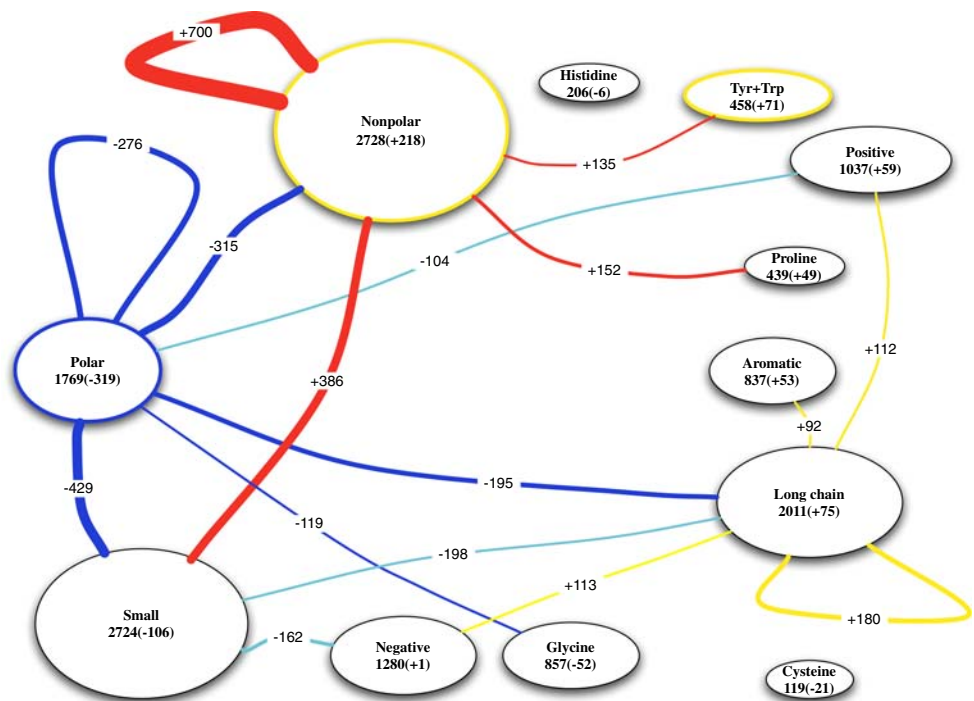


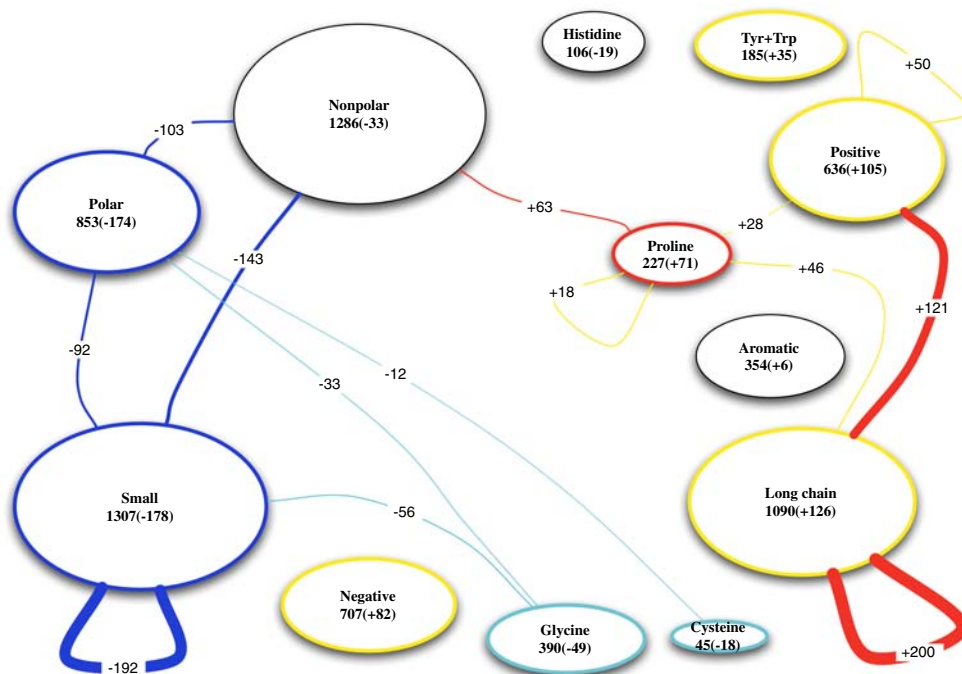
Fig. 3 Changes in contacts between classes of amino acids from mesophiles to thermophiles, buried residues only. See caption of Fig. 1 for details



temperature, the circle is coloured red. If the *P* value is above 0.01, it is coloured yellow. If there is a decrease in the group, it is coloured blue (*P* value <0.001) or cyan (*P* value <0.01). If there is a significant change in the number of contacts between two groups, there is a line between them—the thickness is according to the magnitude of the net change in contacts between the groups. The colour of

each line indicates the significance of the change, using the same scheme as for the groups. Figures 1 and 2 show the results for psychrophiles versus mesophiles (temperatures 0–40°), while Figs. 3 and 4 show the changes between mesophiles and thermophiles (35–102°). The odd numbered figures show results for buried residues, and the even numbered for exposed residues.

Fig. 4 Changes in contacts between classes of amino acids from mesophiles to thermophiles, exposed residues only. See caption of Fig. 1 for details



Supplementary Tables 1–4 contain the full results of the same data. For each pair of amino acid groups, the tables show the number of contacts in the sequences adapted to the coldest temperatures and then the change in the hottest sequence as well as the *P* value obtained using the binomial test as described in the “Methods” section.

The mesophiles versus thermophiles analyses show the greatest differences, and the changes in frequency are very different between the surface and the core of the proteins.

Discussion

Polar amino acids

In thermophiles, contacts between polar noncharged amino acids and almost all other groups of amino acids are less common at higher temperatures, both in the core and on the surface of the proteins. The frequency of polar amino acids also drops from psychrophiles to mesophiles, although less so. Since the overall frequency of polar amino acids drops when going from cold to warmer-adapted proteins, it is not surprising that contacts involving these residue types also become less common. This is also the case for polar-polar contacts that are normally thought to be favourable for stability because their sidechains can form hydrogen bonds with each other. For buried residues, one might think that a hydrophobic core is more favourable, and that there are fewer polar residues for that reason. This is unlikely to be the explanation for the fewer polar residues on the surface,

where a more likely explanation is that most polar residues are quite small and small residues seem to be unfavourable on the surface at higher temperatures (Saelensminde et al. 2007).

Nonpolar amino acids

It is well known that hydrophobic interactions in the protein core are important for the folding and stability for natural proteins (Branden and Tooze 1999). Thus, a more hydrophobic core could make the protein more stable at higher temperatures. This effect can be seen clearly in Fig. 3, where it appears as very strong and indicates that a more hydrophobic core with a high number of contacts indeed is an important factor in thermostabilisation of proteins. Notably in this context, nonpolar amino acids have fewer contacts with polar and negatively charged residues near the surface. In terms of stabilisation, contacts with these hydrophilic residues are generally thought to be unfavourable, and in fact, they seem to be eliminated from thermophiles. We also see more contacts between nonpolar amino acids and aromatic amino acids as well as proline, both types being quite hydrophobic, as well as more contacts with small amino acids.

Charged amino acids

Saltbridges, especially near the surface are often considered as an important factor in thermostabilisation (Kumar et al. 2000; Li et al. 2005). Even though the number of

positively and negatively charged residues increases in frequency on the surface of thermophilic proteins, we do not see a significant increase in the number of contacts between them. Nor do we see many new contacts with other amino acids, which make us think that the charged amino acids are mainly in contact with the solvent. In the core of mesophiles, there are less negatively charged residues than in psychrophiles, and the contacts that disappear are those with nonpolar residues. On the other hand, there are more contacts with longchained amino acids.

Proline and glycine

Proline is more common in proteins adapted to high temperatures. This is expected because proline is more rigid than other amino acids and is reducing the entropy of the main chain, making unfolding less likely at high temperatures. This should be independent of which residues it is in contact with, so the level of change in contacts should correspond to the changes in frequency. The extra rigidity would be more stabilising in the loops that typically can be found near the surface. On the other hand, proline is rather hydrophobic, so it could make contacts with other hydrophobic residues. In thermophiles, we see more contacts between proline and hydrophobic residues both in the core and on the surface. We also see more contacts with the other groups that increase in frequency on the surface, although weakly. From psychrophiles to mesophiles, proline increases in frequency, but except for a slight increase in proline–proline contacts in exposed residues in mesophiles, there is not much change.

Unlike proline, glycines make the mainchain more flexible rather than more rigid, since glycines lack a sidechain that restricts freedom of movement for the polypeptide chain. Since we study sidechain interactions, the result of this study can be said to not really apply to glycines. We can only see changes in the number of contacts with polar and small amino acids at the surface of thermophiles, where we see less such contacts. These amino acids show a sharp reduction in frequency anyway, so this is likely a result of the general lower frequency of these groups.

Cysteine

Cysteine is a quite small amino acid, and should behave like those, but it has also the ability to form disulfide bonds with other cysteins, which is considered to be important for thermostabilisation. We should therefore see an increase in cysteine–cystein contacts. The problem is that cysteine is both an infrequent amino acid, and also very conserved in its positions, so we simply do not have enough data to draw any conclusion on what role this amino acid may play in the dataset.

Long and short chained amino acids

Long chained amino acids are thought to be favourable for thermostabilisation, since they can form more van der Waal interactions with other amino acid sidechains. This would increase the total contact number for the amino acid that again could increase stability. Especially, long-range contacts between different parts of the protein chain are favourable. Although our method cannot see additional contacts (i.e. contacts not found analysing the structure as described in the pdb file), we do see more contacts involving longchained amino acids, especially on the surface of thermophiles, but also in the core. Between psychrophiles and mesophiles there seems to be little difference in the number of contacts involving amino acids with long sidechains.

Small amino acids generally decrease in frequency on the surface of thermophiles, and contacts with polar, negative and nonpolar amino acids decrease significantly. In the core, there is no overall change in frequency of small amino acids, but there is a significant increase in contacts between small and nonpolar amino acids, and less contacts with most other types, significantly so for other small and polar amino acids. The same effects can be seen between psychrophiles and mesophiles, but to a smaller degree. For exposed long chained amino acids, we do not see such an effect.

Aromatic amino acids

The aromatic amino acids are thought to be important for stabilising the fold of the proteins because of their heavy compact sidechains. Between psychrophiles and mesophiles, we do not see much of a difference. From mesophiles to thermophiles there is a significant increase in the frequency of Tyr and Trp, both for buried and exposed positions. For buried residues, we see a higher number of Tyr + Trp in contact with nonpolar residues, but we do not see more or less of contacts involving other aromatic amino acids. Nor can we see a significant increase in contacts with positively charged residues or contacts between aromatic residues. The aromatic amino acids have been theorised to play an important role in thermostabilisation, and many interaction types involving aromatic amino acids have been discussed. Kannan and Vishveshwara (2000) suggested that aromatic clusters could play an important role in thermostabilisation. If this were the case, one would expect our analysis to show an increase in aromatic–aromatic contacts. However, we cannot see this in our data. This may be due to limitations of our analysis (see below). Another kind of interaction thought to play a role is the cation– π bond (Chakravarty and Varadarajan 2002; Gromiha et al. 2002), thought to be stronger than salt bridges. Our data do not

give any evidence that this kind of bond is important for thermostabilisation, at least as a widely used adaptive strategy. What we can see is that there is an increase in contacts between Trp and Tyr and nonpolar amino acids. This can be because of the hydrophobicity of the aromatic amino acids, and the need to form tightly packed networks of interaction in the core for increased cooperativity in the folding of the protein (see below). On the other hand, the contacts with other aromatic amino acids do not increase significantly.

Concluding remarks

Salt bridges can potentially stabilize proteins, e.g. (Elcock 1998; Thomas and Elcock 2004). In the thermophilic proteins, there are a slightly higher number of salt bridges among exposed residues, but the change does not give a *P* value below our cutoff, and this is only a marginal effect compared to the increase and decrease of other types of contacts. Recently, Glyakina et al. (2007) found a higher number of salt bridges in thermophilic proteins, but although the change was statistically significant, the difference between mesophiles and thermophiles was not huge. It seems reasonable that salt bridges plays a positive but not decisive role in thermostabilisation, and cannot explain the higher number of charged residues on the surface or the fact that positively charged residues are increasing much more than negatively charged residues. In our previous paper (Saelensminde et al. 2007), we noted that the changes in charged amino acid frequencies was not what would be expected if the formation of a maximum number of salt bridges was the underlying driving force behind adaption to hotter environments. Other contact types that have been proposed to play an important role in thermostabilization are the cation- π bond and contacts between positively charged residues Arg and Lys with aromatic residues Trp or Tyr. These have been suggested to be stronger than salt bridges (Chakravarty and Varadarajan 2002; Gromiha et al. 2002). We do not see any evidence of an increased number of such contacts, even though all of these amino acids are more frequent in thermophiles. The most straightforward explanation for the trends in charged residues is, in our opinion, that solvent interactions with the charged residues at the surface provides a stability advantage in hotter environments. At higher temperatures, the entropic cost of ordering water around non-charged residues is greater than at lower temperatures. There is also a large enthalpic gain when solvent water interacts with charged groups. This solvent effect does not require specific residue-residue contacts, only a general increase of charged residues consistent with what is seen here and in other studies (Kumar et al. 2000).

Aromatic residues have been suggested to play an important role in increased thermostability. Kannan and Vishveshwara (2000) suggested that aromatic clusters play a key role in thermostabilization. We cannot see any of these trends in our data, even though Tyr, Arg and Lys are more common in protein from species adapted to the highest temperatures.

A number of amino acids are thought to be unstable at high temperatures. This is especially the case for Gln and Asn that can be deamidated (Das and Gerstein 2000), but can also be the case for the sulphur containing amino acids Met and Cys that can be oxidised (Jaenicke and Bohm 1998), as well as Trp. These effects are more likely to happen for exposed residues that can react with substances in the solvent. This is consistent with the lower number of polar residues on the surface of thermophiles we observe.

This study does not uncover many earlier unknown relationships that earlier frequency-based studies has not found, but it does weakens the evidence for some of the contact types that has been suggested to play an important role in thermostabilisation. Especially, factors like aromatic stacking and cation- π bonds seem not to play a major role. Salt bridges show a weak, but not significant increase. This is what we would expect, if salt bridges contribute to some families and only a few is needed. It does not support salt bridges as the major strategy for thermostabilisation.

While solvent interactions with exposed charged residues may be a sufficient explanation for our observations, there are other hypotheses to be considered as well. A high number of charged residues on the surface was also observed by Berezovsky et al. (2007), and they suggested that these charged residues were not forming salt bridges, but were part of what they called negative design. Negative design means that the proteins have been selected during evolution so that the most likely partially unfolded states get less favourable because charged residues of the same sign will come in close contact and repel each other. This will give a higher barrier of unfolding, so that the protein is less likely to unfold, even at extreme conditions. Along a different approach, the loss of long-range contacts has been reported to affect the cooperativity of the folding of hen egg white lysozyme (Dumoulin et al. 2005; Zhou et al. 2007). These and other studies indicate that there is a link between contacts and cooperativity in protein folding (Abkevich et al. 1995; Glyakina et al. 2007). For thermophiles that need to retain a well-defined structure at high temperatures, very cooperative folding with a high barrier separating the folded and unfolded state would be advantageous. A very recent work shows that homologous proteins with the same structural scaffold can have very different folding barriers, due to selective pressure on their charged residues (Halskau et al. 2008). It may thus be that

folding cooperativity and folding barriers can contribute to understanding thermostabilisation.

Regrettably, the method here used for finding contacts is not able to find new contacts. This is because the contacts are only inferred from homology, and uses a measure that is independent of the amino acid type. This also prevents us from detecting contacts that can exist between long side chains, as we need to use a relatively low distance threshold when defining contacts. Using a higher threshold than 6 Å, would result in inclusion of too many residue pairs and reflect the general environment around the amino acid rather than specific contacts. We are most likely to miss contacts between large amino acids like Arg, Lys and Trp, possibly explaining why we find little tendency for the aromatic clustering reported by Kannan and Vishveshwara (2000). While we may have missed some contacts, we think we have captured the major differences between the proteins in terms of changes in sidechain contacts. A more accurate study, allowing more complete capture of contacts including novel contacts would require structures for both homologous proteins, which would restrict the data sets available.

In general, it seems like unspecific contacts, like hydrophobic interactions and better packing of core residues are the most important factor in stabilising the protein core at high temperatures, and is more important than specific interactions like salt bridges, cation- π bonding or aromatic stacking that have been suggested to be important factors. On the surface, long-range interactions and solvent effects seem to dominate.

This is not to say that specific contacts are of no importance. For example, negative–positive contacts is important in protein globules, and these are more common than expected according to Singh and Thornton (1992), as are aromatic interactions, but our data do not indicate that an increase in these contact types is the dominant factor in thermostabilisation.

Based on our analysis, we see it as likely that proteins through evolution have become more thermostable (or gradually lost thermostability if the origin of life was thermophilic), by gradually improving the packing interaction to make the core better packed and more hydrophobic as well as improving solvent interactions, and that this gradual improvement has been more prevalent than evolving specific stabilising contacts. Even though it may still be possible to by design and modify proteins to become more thermostable by changing a few amino acids to improve the contacts in a predictable way, our data indicate that nature by evolution has obtained thermostability by more gradual and global changes.

With respect of cold adaptation, the patterns we see are similar from psychrophiles to mesophiles as they are from mesophiles to thermophiles, although weaker. It thus seems like there are similar ways of adaptation all along the

temperature scale. The signal is much stronger towards the higher temperatures, which likely means that there are tougher constraints on protein globules at high temperatures.

Acknowledgments Thanks to James McDonald at NIMR (UK) for the program to compute amino acid distances. This work was funded by the Norwegian functional genomics programme (FUGE) through the ENVISION project and the FUGE bioinformatics platform. It also received funding from the Norwegian Cancer Society.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

- Abkevich VI, Gutin AM et al (1995) Impact of local and non-local interactions on thermodynamics and kinetics of protein folding. *J Mol Biol* 252(4):460–471
- Beeby M, O'Connor BD et al (2005) The genomics of disulfide bonding and protein stabilization in thermophiles. *PLoS Biol* 3(9):e309
- Benson DA, Karsch-Mizrachi I et al (2000) GenBank. *Nucleic Acids Res* 28(1):15–18
- Berezovsky I, Zeldovich K et al (2007) Positive and negative design in stability and thermal adaptation of natural proteins. *PLoS Comput Biol* 3(3):e52
- Berman HM, Westbrook J et al (2000) The protein data bank. *Nucleic Acids Res* 28(1):235–242
- Branden C, Tooze J (1999) Introduction to protein structure, Garland, New York
- Chakravarty S, Varadarajan R (2002) Elucidation of factors responsible for enhanced thermal stability of proteins: a structural genomics based study. *Biochemistry* 41(25):8152–8161
- Das R, Gerstein M (2000) The stability of thermophilic proteins: a study based on comprehensive genome comparison. *Funct Integr Genomics* 1(1):76–88
- Dumoulin M, Canet D et al (2005) Reduced global cooperativity is a common feature underlying the amyloidogenicity of pathogenic lysozyme mutations. *J Mol Biol* 346(3):773–788
- Edgar RC (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 5:113
- Elcock AH (1998) The stability of salt bridges at high temperatures: implications for hyperthermophilic proteins. *J Mol Biol* 284(2):489–502
- Feller G (2007) Life at low temperatures: is disorder the driving force? *Extremophiles* 11(2):211–216
- Feller G, Gerday C (2003) Psychrophilic enzymes: hot topics in cold adaptation. *Nat Rev Microbiol* 1(3):200–208
- Glyakina AV, Garbuzynskiy SO et al (2007) Different packing of external residues can explain differences in the thermostability of proteins from thermophilic and mesophilic organisms. *Bioinformatics* 23(17):2231–2238
- Gromiha MM, Thomas S et al (2002) Role of cation- π interactions to the stability of thermophilic proteins. *Prep Biochem Biotechnol* 32(4):355–362
- Halskau O Jr, Perez-Jimenez R et al (2008) Large-scale modulation of thermodynamic protein folding barriers linked to electrostatics. *Proc Natl Acad Sci USA* 105(25):8625–8630

- Hamelryck T (2005) An amino acid has two sides: a new 2D measure provides a different view of solvent exposure. *Proteins* 59(1):38–48
- Huang SL, Wu LC et al (2004) PGTdb: a database providing growth temperatures of prokaryotes. *Bioinformatics* 20(2):276–278
- Jaenicke R, Bohm G (1998) The stability of proteins in extreme environments. *Curr Opin Struct Biol* 8(6):738–748
- Jenney FE, Adams MW (2008) The impact of extremophiles on structural genomics (and vice versa). *Extremophiles* 12(1):39–50
- Kabsch W, Sander C (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* 22(12):2577–2637
- Kannan N, Vishveshwara S (2000) Aromatic clusters: a determinant of thermal stability of thermophilic proteins. *Protein Eng* 13(11):753–761
- Kashefi K, Lovley DR (2003) Extending the upper temperature limit for life. *Science* 301(5635):934
- Kleinjung J, Romein J et al (2004) Contact-based sequence alignment. *Nucleic Acids Res* 32(8):2464–2473
- Kumar S, Ma B et al (2000) Electrostatic strengths of salt bridges in thermophilic and mesophilic glutamate dehydrogenase monomers. *Proteins* 38(4):368–383
- Li W, Godzik A (2006) Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22(13):1658–1659
- Li WF, Zhou XX et al (2005) Structural features of thermozymes. *Biotechnol Adv* 23(4):271–281
- Nakashima H, Fukuchi S et al (2003) Compositional changes in RNA, DNA and proteins for bacterial adaptation to higher and lower temperatures. *J Biochem* 133(4):507–513
- Olufsen M, Smalås AO et al (2005) Increased flexibility as a strategy for cold adaptation: a comparative molecular dynamics study of cold- and warm-active uracil DNA glycosylase. *J Biol Chem* 280(18):18042–18048
- Saelensminde G, Halskau Ø et al (2007) Structure-dependent relationships between growth temperature of prokaryotes and the amino acid frequency in their proteins. *Extremophiles* 11(4):585–596
- Siddiqui KS, Cavicchioli R (2006) Cold-adapted enzymes. *Annu Rev Biochem* 75:403–433
- Singh J, Thornton J (1992) Atlas of protein side-chain interactions. IRL Press, Oxford
- Taylor WR (1986) The classification of amino acid conservation. *J Theor Biol* 119(2):205–218
- The UniProt Consortium (2008) The Universal Protein Resource (UniProt). *Nucleic Acids Res* 36(Database issue):D190–D195
- Thomas AS, Elcock AH (2004) Molecular simulations suggest protein salt bridges are uniquely suited to life at high temperatures. *J Am Chem Soc* 126(7):2208–2214
- Wheeler DL, Chappey C et al (2000) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* 28(1):10–14
- Zeldovich K, Berezovsky I et al (2007) Protein and dna sequence determinants of thermophilic adaptation. *PLoS Comput Biol* 3(1):e5
- Zhou H-X, Dong F (2003) Electrostatic contributions to the stability of a thermophilic cold shock protein. *Biophys J* 84(4):2216–2222
- Zhou R, Eleftheriou M et al (2007) Destruction of long-range interactions by a single mutation in lysozyme. *Proc Natl Acad Sci USA* 104:5824–5829