Biochemical and Biophysical Research Communications 380 (2009) 581–585

Contents lists available at ScienceDirect

Biochemical and Biophysical Research Communications

journal homepage: www.elsevier.com/locate/ybbrc

Stability and solubility of proteins from extremophiles

Richard B. Greaves, Jim Warwicker *

Faculty of Life Sciences, Michael Smith Building, University of Manchester, Oxford Road, Manchester M13 9PT, UK

ARTICLE INFO

Article history:
Received 6 January 2009
Available online 29 January 2009

Keywords:
Extremophiles
Thermostability
Solubility
Protein structure
Charge interactions
Sugar binding

ABSTRACT

Charges are important for hyperthermophile protein structure and function. However, the number of charges and their predicted contributions to folded state stability are not correlated, implying that more charge does not imply greater stability. The charge properties that distinguish hyperthermophile proteins also differentiate psychrophile proteins from mesophile proteins, but in the opposite direction and to a smaller extent. We conclude that charge number relates to solubility, whereas protein stability is determined by charge location. Most other structural properties are poorly separated over the ambient temperature range, apart from the burial of certain amino acids. Of particular interest are large non-polar sidechains that tend to increased exposure in proteins evolved to function at higher temperatures. Looking at tryptophan in more detail, this increase is often located close to the termini of secondary structure elements, and is discussed in terms of a novel potential role in protein thermostabilisation.

Among the habitats colonised by specially adapted organisms are environments of extreme salinity and also of low temperature. Psychrophiles are adapted to life at low temperatures, typically defined as being below 15 °C [1]. Such organisms are generally found in the Arctic, the Antarctic and in the ocean depths as well as in alpine and glacial conditions [2]. Halophiles function in conditions of extreme salinity (salt concentrations up to 5 M) such as those found in salterns and hypersaline lakes [3,4]. Moderate halophiles grow under salt concentrations of between 2% and 20% salt (0.3–3.4 M) whereas extreme halophiles grow at greater than 15% (2.6 M) salt [5].

It has been argued that proteins from psychrophiles exhibit heightened catalytic efficiency and also greater thermolability at room temperature than proteins from mesophiles and thermophiles [1,6]. Study of psychrophile proteins should be aligned with mesophile and thermophile proteins to extend the ambient temperature range. In previous work [7], we found that whereas ionisable group (charge) differences were extensive between proteins from hyperthermophiles and mesophiles, variation was not so clear for features based on atomic packing. Several groups have investigated the nature and extent of stabilising interactions within proteins from halophiles and psychrophiles. Most simple is to directly compare sequences or structures of proteins from halophilic (or psychrophilic) organisms with those of mesophile-derived homologues [6,8]. However, such studies are typically based on rather limited numbers of structures. Other work removes the homologues pair restriction, employing computational methods to compare a larger number of sequences and structures [9,10], or to analyse entire genomes [4,5,11]. There has also been protein engineering aimed at investigating thermostability in psychrophile proteins [12].

Factors discriminating psychrophile proteins are, as expected, the reverse of those thought to account for the stability of hyperthermophile proteins [8–10]. The current work looks at discrimination between datasets of 20 protein structures from psychrophiles, 22 protein structures from halophiles, an expanded set (742) of proteins from non-extremophiles (mesophiles), 143 proteins from hyperthermophiles, and 147 proteins from moderate thermophiles. The use of structures and energy calculations allows us to investigate mechanisms of thermo-adaptation, adding to the observation of amino acid preferences [13,14]. Charge interactions and the substitution of charged for polar amino acids [15] are found to be key features, but it is evident that their number and predicted contribution to stability do not correlate. We investigate another intriguing finding, the relatively high exposure of some non-polar sidechains in proteins from organisms adapted to higher temperature. For tryptophan, this increased non-polar surface area is more evident close to the termini of secondary structure elements, consistent with the hypothesis that such effects could play a role in stability to unfolding.

Materials and methods

Datasets of extremophile and mesophile protein structures. We previously [7] obtained a list, culled at 25% sequence identity, of all monomeric, mesophile proteins in the Protein Data Bank (PDB) [16]. From this list, a '290 set' was made through pairing

* Corresponding author. Fax: +44 (0) 161 275 5082.
E-mail address: jim.warwicker@manchester.ac.uk (J. Warwicker).

0006-291X/$ - see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.bbrc.2009.01.145

with thermophile proteins [7]. The remainder of the monomer mesophile set (742 proteins) is now used to expand the mesophile data. Feature calculations gave very similar results for the 290 and 742 protein mesophile sets (e.g. Fig. 1). The original 290 set of mesophile proteins was used to establish a new set of 69 pairings with hyperthermophile protein homologues with the conditions of BLAST [17] *E*-value $<10^{-2}$ and chain length difference of $\leqslant$30 amino acids. Homologue pair sets were not made from the much smaller numbers of halophile and psychrophile proteins. A total of 22 halophile derived protein chains (excluding transmembrane proteins) were returned from a cull at 25% sequence identity [18] of a list of monomeric proteins from the PDB. A similar procedure returned a list of 20 protein chains from psychrophilic organisms. None of the psychrophiles in our dataset live at temperatures more than 20 °C away from a mesophile classification.

Calculated properties. The methodology follows previous work [7]. GminN is the predicted contribution of ionisable groups to folded state stability, normalised to a per amino acid value, and is estimated with Debye-Hückel electrostatics (uniform relative dielectric of 78.4 and ionic strength of 0.15 M). The percentage of ionisable groups in proteins includes those most commonly ionised at neutral pH (Asp, Glu, His, Lys, Arg, N-t, C-t). The number of contacts per atom was calculated using centre to centre contact distance of 6 Å, for non-hydrogen atoms. Solvent accessible surface area was calculated with a solvent probe radius of 1.4 Å. Statistical significance of differences between distributions of properties was calculated at the 5% level with *t*-tests.

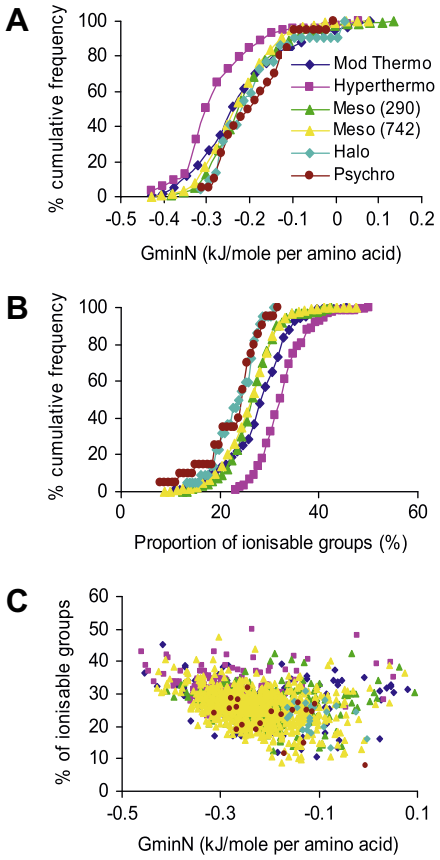

Fig. 1. Plots of cumulative distributions over proteins for charged group properties in mesophile and extremophile proteins. (A) Predicted ionisable group contribution to folded state stability per amino acid, GminN. (B) Proportion of ionisable groups. (C) Scatter plot of GminN versus proportion of ionisable groups.

Results

Datasets

Our analysis gives the following sets and numbers of proteins, which have been culled at 25% sequence identity within each set: original mesophile (290); remaining mesophile (742); moderate thermophile (147); hyperthermophile (143); psychrophile (20); halophile (22). The psychrophile, halophile and larger mesophile (742) protein sets are new to this work. The two mesophile protein sets do not overlap, so that consistency of results can be judged from variation between them.

Charged group features

Fig. 1A shows the predicted contribution of ionisable group interactions to folded state stability, GminN, for the various datasets. As in other features studied, the smaller and larger mesophile protein sets give essentially the same results. The large separation of hyperthermophile proteins is recapitulated from previous work [7]. Halophile proteins are not significantly different from mesophile proteins (see also Table 1) and psychrophile proteins appear to have the smallest magnitude of GminN, although this is not significant at the 5% *t*-test level (Table 1). In Fig. 1B, comparing the proportions of ionisable groups, both the halophile and psychrophile protein sets now separate significantly (Table 1) from mesophile proteins, in the opposite sense to thermophile proteins. Of particular interest (Fig. 1C) is confirmation that changes in ionisable group proportion are not correlated with the predicted contribution of these groups to folded state stability.

Fig. 2A and Table 1 demonstrate the significant separation of the various datasets using the ratio of accessible surface areas for charged and polar amino acids (C*v*P, [19]), including the psychrophile and halophile proteins. This ratio correlates with the proportion of ionisable groups (not shown). It should be noted that, whilst C*v*P is an excellent discriminator of the datasets and leads to discussion of the role of charged groups [20], our analysis of both the ionisable group proportion and predicted contribution to stability suggests that charges have more than one role. Solubility (charge number) as well as folded state stability (charge location) should be considered.

Packing features

In Fig. 2B, the average contacts per atoms show relatively small variation. The only significant separation at the 5% level is halophile from hyperthermophile proteins, with fewer contacts in the hyperthermophile set. These data support the view that packing differences in protein structures, across the ambient temperature range are not clear-cut [21].

Exposed non-polar surface area of amino acids

Composition and amino acid substitutions have been well-studied (recent examples include [13,14]). Preferences for our structural datasets (not shown) are consistent with the previous work. Of note are negative correlations between organism growth temperature (OGT) and content for Asn, Gln, Ser and Thr, consistent with the C*v*P indicator. Although we do not see a large Trp content rise with OGT, Trp shows interesting behaviour when non-polar exposed area is analysed. The average non-polar surface area of some large sidechains (notably Trp) is bigger in hyperthermophile proteins (Fig. 3). This is counter-intuitive, since one might expect more non-polar burial to increase stability at higher OGT. In contrast to the larger non-polar sidechains, Ala