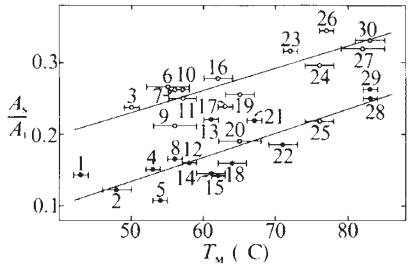

Fig. 1 Relationship of fractional polypeptide area accessible to solvent, A_s/A_t, and protein thermostability, T_m. A_s is the polypeptide surface area in Å^2 accessible to the solvent in a native globular protein, A_t the total accessible surface area of a denatured protein, and T_m the midpoint of the thermal denaturation detected by the $\Delta p\mathrm{H}/\Delta T$ procedure. The open circles indicate monomeric proteins and the filled circles polymeric proteins. The number associated with each experimental value indicates the identity of the protein as encoded in Table 1. The lines represent least-squares analyses of the values for the monomeric (except for proteins 20 and 25) and the polymeric proteins.

mined using the factors[5] 0.857 and 0.714, respectively, to account for the surface lost in the monomeric unit by polymerisation. As seen in Fig. 1, the values for the 30 proteins seem to cluster about two linear relationships with a common slope, one relationship being populated exclusively by monomeric proteins and the second principally by polymeric proteins.

The slopes of these lines suggest that increasing the internal area of these proteins at the expense of surface area decreases the thermostability in the temperature range considered. As hydrophobic index and percentage apolar atoms are independent of polypeptide chain length, and as polar atoms are preferentially located on protein surfaces, it follows that more polar atoms, particularly those of the peptide backbone, must be internalised as the fractional surface area is diminished. Such internalisation of polar atoms in contrast to apolar atoms presents a problem, for burial of a polar atom not complementarily hydrogen bonded in the internal apolar environment is energetically unfavourable. We propose that as more polar atoms need to be internalised to fold longer polypeptide chains into globular units, the probability of noncomplementary burial increases, leading to the decreased thermostability indicated in Fig. 1. In the case of burial of peptide backbone polar atoms, it should be noted that the amount of secondary structure, a common method to pair these atoms complementarily, is independent of chain size. The relationship exhibited by the monomeric proteins predicts that a polypeptide with a molecular weight >60,000 folded into a single globular domain would be unstable when present in an organism maintained at 37°C. Fisher[6] has suggested that large polypeptide chains would be stabilised by forming asymmetrical structures, thus increasing their surface to volume ratios.

Alternatively, the results shown in Fig. 1 suggest two alternative ways in which the thermostability of larger polypeptide chains can be enhanced. One involves using globular surfaces to contribute to conformational stability through association of globular subunits to form polymers. Analysis[7] of subunit interfaces suggests that polar complementarity provides the selectivity while apolar interactions provide the added stability for these inter-subunit interactions. The relationship exhibited by the polymeric proteins in Fig. 1 suggests that polypeptide chains with molecular weights up to about 160,000 would be stable at 37°C provided that such chains fold into globular units and that the globular units polymerise. This limitation would encompass the polypeptide chain lengths of all globular polymeric proteins[8]. The results shown in Fig. 1 also suggest a second way to increase the thermostability of larger polypeptide chains, namely, folding a single chain into multiple globular domains. A case in point is the apparent enhanced thermostability of phosphoglycerate kinase, protein 9. The crystallographic structure[9] of this protein demonstrates that its single polypeptide chain folds into two globular domains of nearly equal size, with minimal interdomain contact; this conformation would increase the fractional surface area relative to folding into a single globular domain and accordingly increase thermostability. Analysis[10] of the sequence of serum albumin suggests that this protein folds into three globular domains of closely related sequence, probably the result of fusion of duplicated structural genes. The position of the T_m for serum albumin, protein 20 in Fig. 1, suggests that its structure resembles that of a covalently linked polymeric protein. The position of the T_m for ovalbumin, protein 25, in Fig. 1, suggests that its relatively large polypeptide chain is also folded into more than one globular domain. Unfortunately, no sequence or crystallographic information is yet available for ovalbumin to validate this suggestion.

In presenting our analysis of thermostability in terms of the size of globular domains and polymerisation of domains, we recognise that other considerations such as the presence of disulphide bonds, ligation of strongly bound metal ions, haem moities and other cofactors, packing densities, and intra-surface interactions will also influence thermostability. Indeed, such considerations may be partly responsible for the scatter of the values about the linear relationships considered in Fig. 1. We also recognise that the inverse A_s/A_t dependence on T_m for globular proteins produces an absurd extrapolation, namely, that the most thermostable protein would have no buried area. Clearly, other considerations must limit this relationship. However, it seems to extend to include pancreatic trypsin inhibitor, which has a A_s/A_t value corresponding to a T_m of 111°C, as no $\Delta p\mathrm{H}/\Delta T$ discontinuity could be detected for this protein at temperatures up to 98°C. Although, as more information becomes available, our analysis of thermostability may suffer the same fate as the average residue volume correlation, a cursory review of thermal inactivation measurements for a variety of enzymes suggests that size of globular domains and polymerisation are critical to protein thermostability.

This investigation was supported by USPHS research grant GM-13215 from the Institute of General Medical Sciences.

EARLE STELLWAGEN
HARVEY WILGUS

Department of Biochemistry,
University of Iowa,
Iowa City, Iowa 52242

Received 30 May; accepted 31 July 1978.

1. Bull, H. B. & Breese, K. *Arch. Biochem. Biophys.* **156,** 604–612 (1973).
2. Bull, H. B. & Breese, K. *Arch. Biochem. Biophys.* **158,** 681–686 (1973).
3. Goodno, C. C., Swenson, C. A. & Bull, H. B. *Analyt. Biochem.* **67,** 220–225 (1975).
4. Chothia, C. *Nature* **254,** 304–308 (1975).
5. Teller, D. C. *Nature* **260,** 729–731 (1976).
6. Fisher, H. F. *Proc. natn. Acad. Sci. U.S.A.* **51,** 1285–1291 (1964).
7. Chothia, C. & Janin, J. *Nature* **256,** 705–708 (1974).
8. Klotz, I. M., Darnall, D. W. & Langerman, N. R. in *The Proteins*, Vol. 1, 3rd edn (eds Neurath, H. & Hill, R. L.) 294–412 (1975).
9. Bryant, T. N., Watson, H. C. & Wendell, P. L. *Nature* **247,** 14–17 (1974).
10. McLachlan, A. D. & Walker, J. E. *J. molec. Biol.* **112,** 543–558 (1977).
11. Bigelow, C. C. *J. theor. Biol.* **16,** 187–193 (1967).

Errata

In the letter 'The units of calcium conduction in *Helix* neurones' by N. Akaike *et al.*, *Nature* **274,** 379–382, ref. 6 should read: Kostyuk, P. G., Krishtal, O. A. & Pidoplichko, V. I. *Nature* **257,** 691–693 (1975). (This replaces the reference to a 1977 paper by the same authors.)

In the letter 'Single crystal analysis of the structure of stishovite' by W. Sinclair and A. E. Ringwood, *Nature* **272,** 714, in paragraph 4 line 2 the space group should read: $P\,4_2/m2_1/n2/m$.

In the letter 'Origin of pregalactic microwave background' by M. J. Rees, *Nature* **275,** 35–37, in paragraph 1 lines 5, 6 should read: $(\delta\rho/\rho) \propto M^{-\alpha} (\frac{1}{3} \leqslant \alpha \leqslant \frac{2}{3})$ (refs 12–15). Also, in paragraph 6 lines 1, 2 should read: Nucleosynthesis could have supplied heavy elements by $t \simeq 10^7$ yr.