The second species (designated 'α' in Fig. 1*b*), which is seen more clearly at lower microwave power, is poorly defined and in low abundance, but it occurs on the low field side of $g = 2.0023$ (the free-spin value), with a maximum value of g of ~2.018. This is therefore also a sulphur species, the spread in g-values being close to those reported for the radical anion, $(RS\dot{-}SR)^-$. These anions cannot be formed by bond homolysis or in any direct manner. It is possible that they are formed by a 'triboelectric' effect or by a partial deprotonation of radicals produced in the reaction shown in equation (2). This would account for the constant presence of both species, and is reasonable because the pK_a value of these radicals is thought to be close to the neutral point[9].

We conclude that cutting α-keratin results in radical formation, the products, as judged by ESR spectroscopy, being remarkably specific. These results should be considered in measurements of free radical production resulting from radiation damage.

This work was supported by the National Radiation Protection Board, Harwell.

Received 27 May; accepted 7 July 1987.

1. Hadley, J. H. & Gordy, W. *Proc. natn. Acad. Sci. U.S.A.* **72**, 3486-3492 (1975).
2. Hadley, J. H. & Gordy, W. *Proc. natn. Acad. Sci. U.S.A.* **71**, 3106-3112 (1974).
3. Hadley, J. H. & Gordy, W. *Proc. natn. Acad. Sci. U.S.A.* **74**, 216-223 (1977).
4. Henriksen, T. *J. chem. Phys.* **37**, 2189-2198 (1962).
5. Nelson, D. J., Petersen, R. L. & Symons, M. C. R. *J. chem. Soc. Perkin Trans.* **II**, 2005-2015 (1977).
6. Nelson, D. J., Petersen, R. L. & Symons, M. C. R. *J. chem. Soc. Perkin Trans.* **II**, 225-231 (1978).
7. Symons, M. C. R. *J chem. Soc. Perkin Trans.* **II**, 1618-1622 (1974).
8. Symons, M. C. R. in *Oxygenated Sulphur Radicals in Chemistry and Medicine* (eds Breccia, A., Rodgers, M. A. J. & Semerano, G.) 71-88 (Lo Scarabeo, Bologna, 1986).
9. Hoffman, M. Z. & Hayon, E. *J. Am. chem. Soc.* **94**, 7950-7963 (1972).

The accessible surface area and stability of oligomeric proteins

Susan Miller*, Arthur M. Lesk†‡, Joël Janin§ & Cyrus Chothia*†

* Christopher Ingold Laboratories, University College London, 20 Gordon Street, London WC1H 0AJ, UK
† Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK
‡ Biocomputing Programme, European Molecular Biology Laboratory, Meyerhofstrasse 1, 6900 Heidelberg, FRG
§ Laboratoire de Biologie Physicochimique, Université Paris-Sud, Bât. 433, 91405 Orsay, France

Protein structures are stabilized by hydrophobic and van der Waals forces, and by hydrogen bonds. The relation between these thermodynamic quantities and the actual three-dimensional structure of proteins can not be calculated precisely. However, certain empirical relations have been discovered. Hydrophobic energy is gained by the reduction of surface in contact with water[1]. For monomeric proteins, the area of the surface accessible to solvent, and of that buried in the interior, is a simple function of molecular weight. Proteins with different shapes and secondary structures, but of the same molecular weight, have the same accessible surface area[2–5]. It has been argued that there is no similar relationship for large oligomeric proteins[6]. In this paper we show that the surface areas of oligomeric proteins, and the areas of the surface buried within them, are directly related to relative molecular mass. Although oligomers of the same molecular weight bury the same amounts of surface, the proportions buried within and between subunits vary. This has important implications for the role of subunit interfaces in the stability and activity of oligomeric proteins.

Calculations were carried out on the 23 oligomers listed in Table 1. All these structures are well determined as judged by the resolution and R factor of the structure determination, the torsion angles of the polypeptide chain and the hydrogen bonding of buried polar groups. Atomic coordinates for these proteins were obtained from the Protein Data Bank[7] or were given to us by their authors. Table 1 gives the references to the crystallographic analyses that determined the structures.

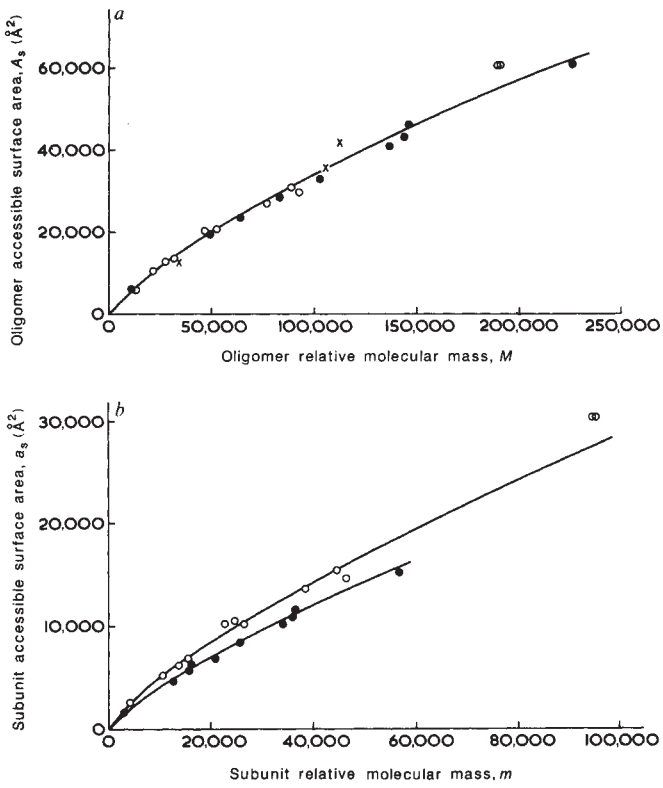

Fig. 1 *a*, The accessible surface areas of the oligomeric proteins (A_S) in Table 1 plotted against relative molecular mass (M). The data for their subunits are given in Table 2. The values for dimeric proteins are indicated by ○, those for tetramers by ●, and those for the hexamers and octomer by ×. The curve in the drawing is given by the equation $A_S = 5.3M^{0.76}$ (see text). *b*, Accessible surface area of the subunits (a_S) of dimers (○) and tetramers (●) are plotted against their relative molecular mass (m). The two curves in the drawing are given by the equations $a_S = 4.5m^{0.76}$ and $a_S = 3.8m^{0.76}$ (see text).

The accessible surface area[8] of individual protein atoms were calculated using the Shrake and Rupley[9] algorithm in the manner described previously[5]. For each protein, we determined the accessible surface area of the monomer in isolation and of the monomer as part of the oligomer (Table 2). The area of the monomer surface buried in the subunit interfaces is the difference between these two numbers.

In Fig. 1*a* the accessible surface area of the oligomers is plotted against their molecular mass. This graph shows a direct relationship between these two quantities. Oligomers with different shapes, secondary structures and number of subunits, but with the same molecular mass (M), have very similar accessible surface area (A_S). If only the high molecular weight proteins are considered there is a linear relation between A_S and M. Over the whole range of molecular weights, the equation

$$A_S = 5.3M^{0.76} \quad (1)$$

gives A_S values for 17 of the proteins that deviate by no more than 5% from the observed values. For the other five proteins the deviations are 8–13%.

It might be thought that, because equations of the form $A_S = kM^{2/3}$ would apply to any set of solid bodies having the same shapes and densities, equation (1) is simply the geometrical consequence of polypeptide chains folding to form globular structures. This, however, is not the origin of equation (1). The

Table 1 Protein structures used in this work

Protein	Residues in monomer	Data bank file number* and structure reference	
Dimers			
Avian pancreatic polypeptide	36	1PPT	14
Subtilisin inhibitor	107	2SSI	15
Cytochrome *c'*	128	2CCY	16
Superoxide dismutase	152	2SOD	17
Fab KOL	216/229	1FB4	18
Triose phosphate isomerase	247	1TIM	19
Alcohol dehydrogenase	374	—	20
Aspartate aminotransferase	401	—	21
Citrate synthase	432	3CTS	22
Phosphorylase *a*	842	—	23
Phosphorylase *b*	842	—	†
Tetramers			
Melittin	26	1MLT	24
Prealbumin	127	2PAB	25
Haemoglobin	141/146	3HHB	26
Glutathione peroxidase	198	1GP1	27
Concanavalin A	237	2CNA	28
Phosphofructokinase	319	—	29
Lactate dehydrogenase	329	—	30
Glyceraldehyde-3-phosphate dehydrogenase	334	—	31
Catalase	506	7CAT	32
Hexamers			
Insulin	51	1INS	33
Phycocyanin	162/172	—	34
Octomer			
Haemerythrin	113	1HMQ	35

* Atomic coordinates were taken from the named file of the protein structure databank[7]; except for those proteins marked (—), coordinates for these structures were given to us by their authors.

† The 1.9 Å resolution atomic structure of phosphorylase *b* was determined by K. R. Acharya, D. I. Stuart and L. N. Johnson (personal communication).

proteins studied here differ greatly in shape: some of the dimers and tetramers are 'ellipsoids' or 'dumb-bells'; in other tetramers the four subunits extend in tetrahedral directions away from the interfaces (see the references in Table 1). Thus in these proteins the polypeptide chains fold into shapes that can be very different but whose surface areas are related to their molecular masses by equation (1).

A recent analysis of 37 well determined monomeric proteins showed that the expression:

$$A_S = 6.3M^{0.73} \quad (2)$$

gives A_S values that on average are within 4% of those observed[5]. (The parameters in this equation differ slightly from those previously obtained with less well determined structures[3,4].) The molecular mass range of the monomeric proteins used to derive equation (2) is 4,000 to 35,000 (ref. 5). In this range the A_S values given by equation (2) are 7% to 13% lower than those given by equation (1). The significance of this small difference is unclear.

Equation (1) implies that subunits of the same molecular weight have lower accessible surface areas in tetramers than they have in dimers. If m is the molecular weight of a subunit, the accessible surface area, a_S, that we would expect it to have if it is part of a dimer is:

$$a_S = \frac{A_S}{2} = \frac{5.3}{2}(2m)^{0.76}$$

$$a_S = 4.5\ m^{0.76}$$

If the subunit is part of a tetramer we would expect:

$$a_S = \frac{A_S}{4} = \frac{5.3}{4}(4m)^{0.76}$$

$$a_S = 3.8\ m^{0.76}$$

Figure 1*b* shows that the subunits in the dimers and tetramers have observed accessible surface areas close to the expected values.

The total accessible surface area of unfolded proteins (A_T) can be estimated by calculating the accessible surface area of the polypeptide in an extended conformation. The expression:

$$A_T = 1.48M \quad (3)$$

fits calculated values to within 1% on average[5]. Because the surface area that remains accessible when oligomers fold is a function of molecular weight, the surface buried within oligomers, $A_T - A_S$, is also a function of molecular weight and its value can be calculated from equations (1) and (3):

$$A_T - A_S = 1.48M - 5.3M^{0.76}$$

The surface buried within an oligomer is the sum of that buried within the subunits and of that buried in the interfaces between them. For the oligomers studied here the size of the surface buried between the subunits varies widely, from 1,360 Å^2 in superoxide dismutase to 42,300 Å^2 in catalase (Table 2). Of particular interest is the variation in the relative contributions the interfaces make to the total buried surface. The area buried in the interface represents between 3% and 40% of the total buried surface $A_T - A_S$. Thus, the contribution of the surfaces buried within and between subunits varies greatly, even though the total buried surface is essentially the same for proteins of a given molecular mass.

In oligomers with small interfaces the isolated subunits have accessible surface areas similar to those found in monomeric proteins of the same molecular mass. For example, the values for the isolated subunits of superoxide dismutase, prealbumin, haemoglobin, glutathione peroxidase, insulin and haemerythrin (Table 2) are within 4% of the values given by equation (2). Conversely, in the oligomers with large interfaces, the isolated

Table 2 Accessible surface area of the monomers in oligomeric proteins (Å^2)

Protein	Subunit molecular mass	Accessible surface area of monomer in oligomer	isolation	interface
Dimers				
Avian pancreatic polypeptide	4,250	2,630	3,320	690
ROP protein	6,650	3,050	4,390	1,340
Subtilisin inhibitor	10,910	5,470	6,220	750
Cytochrome *c'*	13,910	6,340	7,150	810
Superoxide dismutase	15,550	6,760	7,440	680
Fab KOL VL-CL	22,880	10,220	12,050	1,830
VH-CH1	24,300	10,500	12,280	1,780
Triose phosphate isomerase	26,510	10,130	11,720	1,580
Alcohol dehydrogenase	40,590	13,610	15,240	1,630
Aspartate aminotransferase	45,150	15,440	18,590	3,150
Citrate synthase	47,930	14,540	19,420	4,880
Phosphorylase *a**	95,300	30,370	33,820	3,450
Phosphorylase *b**	95,350	30,380	32,690	2,310
Tetramers				
Melittin	2,830	1,560	2,630	1,070
Prealbumin	12,450	4,890	6,400	1,510
Haemoglobin α	15,780	5,700	7,420	1,720
β	16,390	6,340	7,840	1,500
Glutathione peroxidase	20,790	7,090	8,680	1,590
Concanavalin A	25,540	8,300	10,880	2,580
Phosphofructokinase	35,310	10,150	13,750	3,600
Lactate dehydrogenase	36,490	11,700	17,240	5,540
Glyceraldehyde-3-phosphate dehydrogenase	36,560	10,810	14,610	3,800
Catalase	57,940	15,220	25,780	10,570
Hexamers				
Insulin subunit 1	5,740	2,080	3,510	1,430
subunit 2	5,740	2,250	3,680	1,430
Phycocyanin α	18,010	6,600	8,800	2,200
β	19,380	7,200	9,600	2,400
Octomer				
Haemerythrin	13,300	4,480	6,190	1,7120

* In phosphorylase *a* and *b* different regions of the protein are disordered: residues 314 to 324 and 835 to 840 in *a* and residues 1 to 18 in *b*. These residues are not included in the accessible surface areas or in the subunit molecular weights given above.

† The subunit molecular weights given in this table do not include any parts of the protein for which the structure was not determined. They do include any co-factor present in the structure.

subunits have accessible surface areas much larger than those of monomeric proteins of the same size. For example, the values for the isolated subunits of melittin and lactate dehydrogenase are 26–28% larger than those given by equation 2. In these structures we would expect the quaternary interactions to make an important contribution to the conformational stability of the subunits, as is also suggested by experimental studies of their folding[10].

The structural features of oligomeric proteins reported here can be related to their thermodynamic and functional properties. Free energies of protein folding are small differences between large terms favouring either the unfolded or folded state[11,12], particularly conformational entropy of the polypeptide chain and hydrophobic free energy. Conformational entropy favours the unfolded state and depends upon the length of the polypeptide chain, while hydrophobicity depends upon contacts with the solvent and favours the folded state[1,13]. The balance between these two terms implies that solvent accessible surface should be correlated with molecular weight. The correlation has been established in monomeric proteins[2–5]. Here it is established for oligomeric proteins which have the added complexity of subunit interfaces.

We thank John Cresswell for the figure; Drs H. Ekland, P. R. Evans, R. J. Fletterick, R. Huber, J. N. Jansonius, L. Johnson, T. A. Jones, M. G. Rossmann, T. Schirmer and A. J. Wonacott for the atomic coordinates of protein structures; and the Royal Society, the National Institute of General Medical Sciences and the National Science Foundation for support.

Received 22 April; accepted 13 July 1987.

1. Kauzmann, W. *Adv. Protein Chem.* **14,** 1–63 (1959).
2. Chothia, C. *Nature* **254,** 304–308 (1975).
3. Janin, J. *J. molec. Biol.* **105,** 13–14 (1976).
4. Teller, D. *Nature* **260,** 729–731 (1976).
5. Miller, S., Janin, J., Lesk, A. M. & Chothia, C. *J. molec. Biol.* **196,** 641–656 (1987).
6. Sprang, S., Yang, D. & Fletterick, R. J. *Nature* **280,** 333–335 (1979).
7. Bernstein, F. C. *et al. J. molec. Biol.* **112,** 535–542 (1977).
8. Lee, B. & Richards, F. M. *J. molec. Biol.* **55,** 379–400 (1971).
9. Shrake, A. & Rupley, J. A. *J. molec. Biol.* **79,** 351–371 (1973).
10. Jaenicke, R. *Angew. Chem.* **23,** 395–413 (1984).
11. Privalov, P. L. *Adv. Protein Chem.* **33,** 167–241 (1979).
12. Creighton, T. *Proteins* (Freeman, New York, 1983).
13. Baldwin, R. L. *Proc. natn. Acad. Sci. U.S.A.* **83,** 8069–8072 (1986).
14. Blundell, T. L., Pitts, J. E., Tickle, I. J., Wood, S. P. & Wu, C. W. *Proc. natn. Acad. Sci. U.S.A.* **78,** 4175–4179 (1981).
15. Mitsui, Y., Satow, Y., Watanabe, Y. & Iitaka, Y. *J. molec. Biol.* **131,** 697–724 (1979).
16. Finzel, B. C., Weber, P. C., Hardman, K. D. & Salemme, F. R. *J. molec. Biol.* **186,** 627–643 (1985).
17. Tainer, J. A., Getzoff, E. D., Beem, K. M., Richardson, D. C. & Richardson, J. S. *J. molec. Biol.* **160,** 181–217 (1982).
18. Marquart, M., Deisenhofer, J. & Huber, R. *J. molec. Biol.* **141,** 369–391 (1980).
19. Banner, D. W. *et al. Nature* **255,** 609–614 (1975).
20. Ekland, H. *et al. J. molec. Biol.* **146,** 561–587 (1981).
21. Ford, G. C., Eichele, G. & Jansonius, J. N. *Proc. natn. Acad. Sci. U.S.A.* **77,** 2559–2563 (1980).
22. Remington, S., Wiegand, G. & Huber, R. *J. molec. Biol.* **158,** 111–152 (1982).
23. Sprang, S. & Fletterick, R. J. *J. molec. Biol.* **131,** 523–551 (1979).
24. Terwilliger, T. C. & Eisenberg, D. *J. biol. Chem.* **257,** 6016–6022 (1982).
25. Blake, C. C. F., Geisow, M. J., Oatley, S. J., Rerat, B. & Rerat, C. *J. molec. Biol.* **121,** 339–356 (1978).
26. Fermi, G., Perutz, M. F., Shaanan, B. & Fourme, R. *J. molec. Biol.* **175,** 159–174 (1984).
27. Epp, O., Ladenstein, R. & Wendel, A. *Eur. J. Biochem.* **133,** 51–69 (1983).
28. Reeke, G. N., Becker, J. W. & Edelman, G. M. *J. biol. Chem.* **250,** 1525–1547 (1975).
29. Evans, P. R. & Hudson, P. J. *Nature* **279,** 500–504 (1979).
30. White, J. L. *et al. J. molec. Biol.* **102,** 759–779 (1976).
31. Biesecker, G., Harris, J. I., Thierry, J.-C. & Wonacott, A. J. *Nature* **266,** 328–333 (1977).
32. Murthy, M. R. N., Reid, T. J., Sicignano, A., Tanaka, N. & Rossmann, M. G. *J. molec. Biol.* **152,** 465–499 (1981).
33. Blundell, T. L., Dodson, G. G., Hodgkin, D. C. & Mercola, D. *Adv. Protein Chem.* **26,** 279–402 (1972).
34. Schirmer, T., Bode, W., Huber, R., Sidler, W. & Zuber, H. *J. molec. Biol.* **184,** 257–277 (1985).
35. Stenkamp, R. E., Sieker, L. C. & Jensen, L. A. *Acta crystallogr.* B**39,** 697–703 (1983).