wave height map on the cover by the light-coloured region close to Antarctica.

While climatological averages emphasize geographical differences within each field, they give no indication of the range of conditions likely to be encountered in a particular region at different times. The rather striking zonal banding of each of the fields suggests that latitudinal distributions can provide some initial insight into water vapour, wind speed and wave height variability during the Seasat mission. The probability distributions of each of the three fields for the three 5° lat. bands centred at 52.5° N, 2.5° N and 52.5° S are shown in Fig. 3.

The features evident in the water vapour distributions in Fig. 3*a* are not surprising. They clearly show the higher integrated water vapour levels found in the tropics. Only 2% of the observations in the equatorial band were $<3\ g\ cm^{-2}$ while 96% of the observations in the high southern latitude band were $<2\ g\ cm^{-2}$. This reflects the lower water vapour saturation level of the cold Southern Hemisphere winter air. The cool high northern latitude band shows intermediate water vapour values with 75% of the observations between 1 and $3\ g\ cm^{-2}$. A more detailed study of the water vapour estimates might be useful in climatological studies of the global time variability of the latent heat of vapourization transferred from the ocean to the atmosphere.

The wind speed distributions in Fig. 3*b* show some unexpected features. As the high southern latitude band is in the winter storm region, the higher wind speeds in this band are not surprising. Large variations might also be expected due to the passage of individual storms. However, the winds in this band seem to be relatively steady: 68% of the observations were between 8 and $12\ m\ s^{-1}$. In comparison, the trade winds of the equatorial band, which are noted for their steadiness, show a peak at much lower wind speeds but a broader distribution: 76% of the observations were between 4 and $10\ m\ s^{-1}$. Another curious feature is the probability distribution for the high northern latitude band: the range of wind speeds is broader than either of the other two bands but there are two pronounced peaks, a primary peak at $9\ m\ s^{-1}$ and a secondary peak at $5\ m\ s^{-1}$. Present studies are aimed at determining whether the double peaks arise from zonal averaging of an inhomogeneous field or whether they reflect a seasonal change in the wind speed distribution during the 3.5-month period.

The wave height distributions are shown in Fig. 3*c*. The equatorial band has the narrowest range of wave conditions: 70% of the waves were <2 m and virtually none were >4 m. In comparison, only 50% of the waves in the high northern latitude band were <2 m and 5% were >5 m. In contrast to both the equatorial and high northern latitude bands, the peak in the wave height distribution in the stormy high southern latitude band is at a much higher value (3.5 m as opposed to 1.5 m) and the range of wave conditions is much wider. Only 2% of the waves observed were <2 m and over 40% were >5 m. The secondary peak at 6.5 m is rather puzzling; it is difficult to explain in terms of wind forcing as there is no similar double-peaked structure in the wind speed distribution in this latitudinal band. The double peaks probably result from zonal averaging of the large wave heights south and west of Australia with the smaller wave heights elsewhere in this 5° latitudinal band (see cover).

There is no way of knowing at present whether the 3.5-month average wind speed and wave heights presented here are typical for this time of year. However, these geophysical maps demonstrate the unique ability of satellites to provide global measurements of wind and wave conditions. This will be especially useful in data-sparse regions such as the Southern Hemisphere (see Fig. 1*b*) where present forecasts of weather and sea states are extremely unreliable. Even in the Northern Hemisphere, forecasting skill is limited in many regions by sparse data coverage over the oceans. An operational altimetric mission could improve worldwide sea state forecasting.

This research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under contract with NASA. L. Fedor and G. Brown made their results (Fig. 2) available before publication of their own work.

Received 15 July; accepted 12 October 1981.

1. *Science* **204**, 1405–1424 (1979).
2. Walsh, J., Uliana, E. A. & Yaplee, B. S. *Bound. Layer Met.* **13**, 263–276 (1978).
3. Fedor, L. S. & Brown, G. S. *J. geophys. Res.* (in the press).
4. Brown, G. S. *J. geophys. Res.* **84(B8)**, 3974–3978 (1979).
5. Tapley, B. D., Lundberg, J. B. & Born, G. H. *J. geophys. Res.* (in the press).
6. Nelson, C. S. *NOAA Tech. Rep. NMFS SSRF-714* (1977).

Dipoles of the α-helix and β-sheet: their role in protein folding

Wim G. J. Hol*, Louis M. Halie* & Christian Sander†

* Laboratory of Chemical Physics, Department of Chemistry, University of Groningen, Nijenborgh 16, 9747 AG Groningen, The Netherlands
† Biophysics Department, Max Planck Institute of Medical Research, Jahnstrasse 29, Heidelberg, FRG

As a result of the regular arrangement of peptide dipoles in secondary structure segments and the low effective dielectric constant in hydrophobic cores, the electrostatic energy of a protein is very sensitive to the relative orientation of the segments. We provide here evidence that the alignment of secondary structure dipoles is significant in determining the three-dimensional structure of globular proteins.

IT has long been known that in the α-helix the alignment of the peptide dipoles parallel to the helix axis gives rise to a macrodipole of considerable strength (for review see ref. 1). With a dipole moment for each peptide unit of ~3.5 D, a helix of, for example, 10 residues has a dipole moment of 34 D, as 97% of the peptide dipole points in the direction of the helix axis. It has been shown[2] for points near the helix termini that the effect of the helix dipole is equivalent to the effect of half a positive unit charge at the N-terminus of the helix and half a negative charge at the C-terminus. The same authors provided evidence that the considerable electric field due to the helix dipole moment is used by proteins in: (1) binding negatively-charged groups, such as phosphate groups; (2) rendering protein side chains located near the N-terminus more nucleophilic; and, (3) stabilizing charged transition states or intermediates along the catalytic pathway. So far ~20 helices have been observed which bind negatively-charged groups near their N-termini. Another 10 proteins have their active site close to the N-terminus of a helix (W.G.J.H., unpublished results). For papain, an extensive quantum mechanical calculation[3] showed that the field generated by the helix dipole is a major factor in the stabilization of the sulphhydryl–imidazole ion pair essential in

0028–0836/81/500532—05$01.00

catalytic mechanism of this enzyme[4]. Secondary structure formation also seems to be influenced by side chain–main chain electrostatic interactions[5,6].

Here we extend the role of the helix dipole to the area of protein folding. In this we are motivated by the idea that the directionally sensitive and relatively long-range dipole–dipole interaction is an important factor in ordering the secondary structure in the hydrophobic interior of folded globular proteins.

Model calculations of helix–helix interaction

Figure 1 shows, for a simple case, the result of model calculations of the electrostatic energy between the backbone dipoles of two helices as a function of their distance and relative orientation. The strongest variation in energy occurs when one helix axis is turned by 180° relative to the other (Fig. 1*b*). The energy difference between a parallel helix pair and an antiparallel pair is considerable (Fig. 1*b*,*c*). These calculations also show that the interaction between two helix dipoles reaches an upper limit at a helix length of ~35 residues (Fig. 1*d*). We consider first the extent to which proteins containing helices and no β-sheets reflect these simple electrostatic model considerations. Proteins containing β-structures will be discussed later.

All-helical proteins

In an elegant analysis of protein structures, Levitt and Chotia[7] showed schematically three major types of all-helical structures. By assigning 'plus' signs to the N-termini and 'minus' signs to the C-termini (indicating the effective charge) in their two-dimensional drawings, the alternating pattern of plus and minus signs tentatively confirms the trend of our model calculations.

Estimates of the actual three-dimensional interaction energies between the helices in the all-helical and predominantly helical proteins tobacco mosaic virus (TMV)[8,9], myoglobin[10] and parvalbumin[11] are given in Table 1. These calculations show

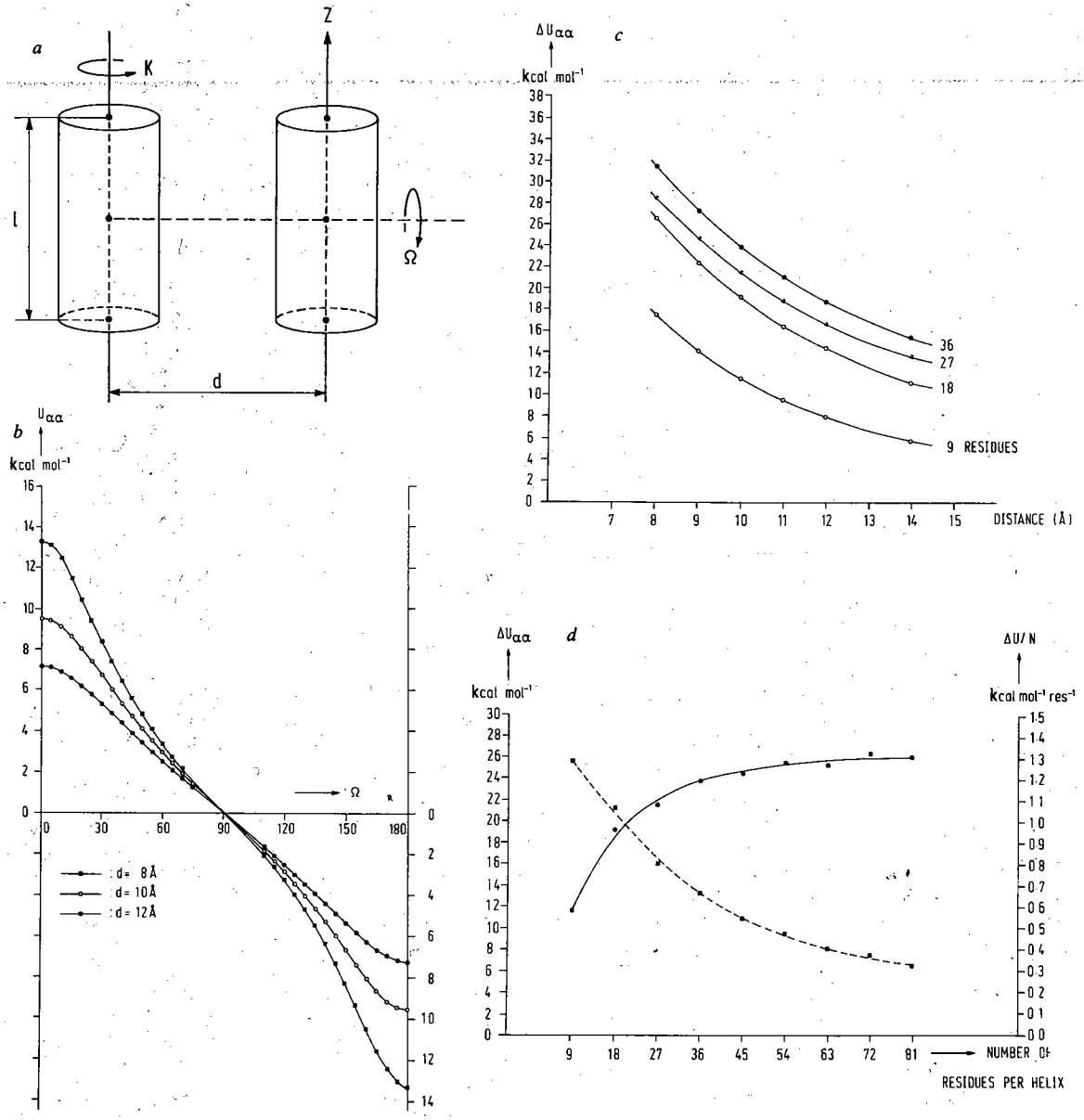

Fig. 1 *a*, The relative position and orientation of two helices in the simple case where the helix axes are perpendicular to the line connecting the centres. A helix can be rotated about its own axis by an angle K (with negligible variation in electrostatic energy) and about the inter-helical axis by an angle Ω (with considerable variation in electrostatic energy). *b*, The strong variation in electrostatic energy between two model α-helices as a function of twist angle Ω. The length of both helices is five turns, that is, 18 residues or 27 Å. *c*, The difference in electrostatic energy between a pair of parallel (Ω = 0) and a pair of antiparallel (Ω = 180) helices falls off gradually with increasing distance, d. *d*, The difference in electrostatic energy between a pair of parallel (Ω = 0) and a pair of antiparallel (Ω = 180) helices at a mutual distance of 10 Å as a function of helix (solid line). The figure illustrates that the difference becomes less important on a per residue basis for longer helices (broken line). For the model calculations, we used polyalanine in a helical geometry, generated by the program of Hermans *et al.*[75]. The only partial charges are those of the peptide group[2]: C, 0.42; O, −0.42; N, −0.20 and H, 0.20 unit charge. No cutoff radius was used and dielectric screening was neglected.

Table 1 Calculated electrostatic interaction energies between the backbones of α-helices and β-strands in various proteins

Protein	Type of protein	No. of α-helices	No. of β-strands	$U_{\alpha\alpha}$ (kcal mol^{-1})	$U_{\alpha\beta}$ (kcal mol^{-1})	Ref.
Met-myoglobin	α_a	9	–	−22.7	–	10
TMV	α_a	4	–	−14.3	–	8, 9
Parvalbumin	α_a	6	–	−11.8	–	11
Cytochrome b_5	α_a	6	–	−4.7	–	22
Phospholipase A2	$\alpha_a\beta_a$	4	2	−8.3	−0.8	41
Cytochrome c	$\alpha_a\beta_a$	4	3	−3.4	−1.4	42
Hen egg-white lysozyme	$\alpha_a\beta_a$	6	5	−12.8	−1.6	19
Phage T4 lysozyme	$\alpha_a\beta_a$	10	3	−23.5	−0.7	20
Thermolysin	$\alpha_a\beta$	7	14	−3.6	−5.4	21
Lactate dehydrogenase	$\alpha_p\beta_p$	9	15	+14.0	−14.0	43
Alcohol dehydrogenase	$\alpha_p\beta_p$	10	20	+14.2	−15.3	44
Glyceraldehyde phosphate dehydrogenase	$\alpha_p\beta_p$	7	18	+5.8	−5.7	45
Adenylate kinase	$\alpha_p\beta_p$	10	5	+4.9	−10.5	46
Rhodanese	$\alpha_p\beta_p$	10	10	+13.5	−10.3	32
Subtilisin	$\alpha_p\beta_p$	8	9	+5.2	−15.0	47
Dihydrofolate reductase	$\alpha_p\beta_p$	4	8	+3.2	−3.8	48
Flavodoxin	$\alpha_p\beta_p$	5	7	+13.3	−4.9	36
Triose phosphate isomerase	$\alpha_p\beta_p$	8	8	+6.6	−7.0	49
Carboxypeptidase	$\alpha_p\beta_p$	9	12	+11.4	−11.0	50
Papain/actinidin	$\alpha\beta+\beta_a$	6	8	+13.3	−6.6	51, 52

The table demonstrates the favourable alignment of α-helices in all-α proteins and the compensation for unfavourable helix interactions by favourable helix–strand interactions in other proteins. $U_{\alpha\alpha}$ is the sum of all main-chain dipole helix–helix interactions, $U_{\alpha\beta}$ the sum of all main-chain dipole helix–strand interactions. $U=\Sigma(q_iq_j/r_{ij})$, where r_{ij} is the distance between two charges q_i and q_j, with no cut-off in r and with the partial charges[2]: O = −0.42, C = +0.42, N = −0.20, H = +0.20, corresponding to a peptide dipole moment of 3.5 D in the direction of the N–H and O–C bonds. Indices i and j label the residues in the segments; i and j are not in the same segment. A dielectric constant was not used because of the breakdown on the atomic scale of continuum theories on which the concept of dielectric constant is based. This breakdown is illustrated quantitatively by molecular dynamics simulations[53]. The dielectric response, however, can be estimated and is physically due to the electronic polarizability of the protein atoms and the orientational polarizability of solvent water. The effects of the former can be separated into two contributions which occur in opposite directions. (1) Recent theoretical estimates of an effective dielectric constant for the non-polar atoms in the protein interior are $\varepsilon=2$ (for example, derived from atomic polarizabilities[54]). This decreases U by a factor of ~2. (2) Cooperatively aligned backbone dipoles in both α-helix and β-sheet mutually polarize each other, increasing the O← C and N→ H dipoles by a factor of ~1.4 (from 3.5 to 4.8–5.0 D; see refs 1, 55) and thus increasing the interaction energies U by a factor of $1.4\times1.4\approx2$. The dielectric screening due to water is difficult to calculate. Kirkwood and Westheimer[56] and Ehrenson[34] have estimated the effective screening of charge–charge and charge–dipole interactions in the interior of a sphere of low dielectric surrounded by water: the screening is only effective near the surface and dies off quickly in the interior. As helices are usually nearer the protein surface than parallel β-sheets, this means that our $U_{\alpha\alpha}$ values are an overestimate relative to the $U_{\alpha\beta}$ values. Overall, these estimates of the dielectric response indicate that our energy estimates of $U_{\alpha\alpha}$ and $U_{\alpha\beta}$ are incorrect by no more than a factor of 2. The coordinates are from the Protein Data bank[57], except for dihydrofolate reductase, TMV and phospholipase, the coordinates of which were provided by J. Kraut, A. Bloomer and J. Drenth, respectively. The position of the peptide hydrogen was generated according to standard geometry. The lengths of the helices and strands were taken from Levitt and Greer[58], except in the following cases: TMV[8], phospholipase[41], lactate dehydrogenase[43] and actinidin[52] were taken from the literature and dihydrofolate reductase[48] was obtained by inspection on an interactive display. Exceptionally favourable interactions between secondary structure segments next to each other in the sequence were removed by deleting residues from the calculations so that no atoms from different segments were closer than 3.6 Å. Parts of kinked helices were merged into one helix. The types of proteins are grouped according to the predominant arrangement of helices and strands. The plus sign in the second column indicates that different domains belong to different types of arrangement. For papain[51] and the closely related protease actinidin[52] there is no obvious cancellation of the parallel helix dipoles by either strands or by many exposed charged residues. There is, however, an unusual placement of charged interior residues and a network of interior water molecules[52]. Therefore, the helical domain of these proteins is intriguing from the electrostatic as well as from the hydrophobic viewpoint.

very favourable overall electrostatic interactions between the helices, in complete qualitative agreement with our model calculations. Although these energies are difficult to calculate precisely (see Table 1 legend), we do not expect any change in the qualitative conclusions when more accurate calculations are available.

The all-helical proteins myohaemerythrin[12], ferritin[13], cytochrome b_{562} (ref. 14) and cytochrome c' (ref. 15) strongly resemble the optimally antiparallel TMV structure[16]. The all-helical protein uteroglobin contains only antiparallel helices[17]. The larger protein cytochrome c peroxidase[18] contains eight helices of which five form an antiparallel arrangement and none are aligned in parallel. Our model is also supported by the structure of 6-phosphogluconate dehydrogenase (M. Adams and S. White, in preparation): four pairs of antiparallel helices are wrapped perpendicularly around two long antiparallel helices which form the core of this protein. In addition, helices in the helical domains of hen egg-white lysozyme[19], phage T4 lysozyme[20], thermolysin[21] and cytochrome b_5 (ref. 22) have a definite tendency to run antiparallel. This is reflected in the negative interaction energies $U_{\alpha\alpha}$ calculated for these proteins (see Table 1).

Proteins with alternating α and β structures

From an electrostatic viewpoint, it is surprising that in proteins with alternating α-helices and β-strands, the helices are roughly parallel, that is, with seemingly unfavourable helix dipole alignment. In studying this phenomenon electrostatically, we examined the arrangement of peptide units in parallel β-strands. From schematic drawings of hydrogen bonding patterns in parallel β-sheets (Fig. 2), it is obvious that the N→ H as well as the C← O dipoles point backwards with respect to the N- to C-terminal direction of the strands. This suggests that the parallel β-sheet has a significant overall dipole moment, with the N-terminal end of the strands corresponding to the positive end of the dipole. Quantitatively, about one-third of the peptide dipole is parallel to the strand direction. This allows an approximate description of a parallel β-strand as having 1/15 of a positive charge near the C-terminus (Fig. 2a). Electrostatically, then, the α–β dipolar interaction is favourable when helices and strands are antiparallel.

A second aspect of the electrostatic interaction between helices and sheets concerns the twist of the sheet[23]. A consequence of this twist is that the 'parallel' helices in the α/β proteins make considerable angles (typically −40°; ref. 24) with each other, thus decreasing appreciably the unfavourable interaction between their dipoles (Fig. 1b).

We have calculated the actual α–α and α–β dipole interactions in various α/β proteins and the simple considerations above seem to be valid generally (Table 1). The negative values for $U_{\alpha\beta}$ qualitatively confirm our hypothesis of a favourable strand–helix interaction. This interaction compensates fully or partly for the unfavourable helix–helix interaction (Table 1). In our energy estimates, we disregarded the difference in electrostatic screening near the surface and in the interior. As the parallel β-sheets are sandwiched by the helices, we have overestimated $U_{\alpha\alpha}$ relative to $U_{\alpha\beta}$ and so the actual total energies are probably more favourable. Thus, domain structures with parallel central β-sheets antiparallel to surrounding helices may be significantly stabilized by electrostatic interactions.

All-β proteins

In globular proteins containing only β-sheets, there is an overwhelming tendency for the strands to align in an antiparallel manner (ref. 7 and Table 2). The same seems to be true for fibrous β-sheet proteins[25,26]. Electrostatically, an energy

difference between the two kinds of pleated sheet may be related to the fact that in the antiparallel sheet all peptide dipole moments seem to cancel each other out: there is no residual moment in either direction (Fig. 2*b*). In the parallel arrangement the components of the dipole moment parallel to the strand direction interact unfavourably with each other. A simple model calculation, assuming only a dipole moment in the parallel strand and no dipole moment in the antiparallel strand, shows that this effect may be ~0.4 kcal mol^{-1} for a pair of short β-strands containing four residues. This gives a difference in electrostatic energy of ~0.8 kcal mol^{-1} between a parallel and antiparallel arrangement of three strands, which is much smaller than the energy difference for α-helices, but it may nevertheless be significant.

The $\alpha_a\beta_a$ and larger proteins

A small number of proteins do not fall into the all-α, α/β or all-β categories, for example the C-terminal fragment of the ribosomal protein L7/L12 (ref. 27), in which the three helices and three β-strands are optimally antiparallel to each other. The L7/L12 protein is very stable[27]—probably due not only to the close packing of nonpolar residues in its unusually hydrophobic core, but also to the optimally antiparallel alignment of secondary structure backbones embedded within that core. Proteins like L7/L12 can be described as $\alpha_a\beta_a$, whereas the 'classical' α/β proteins may be coined $\alpha_p\beta_p$ (a, antiparallel; p parallel). The all-helical proteins may then be described as α_a, and the all-β proteins as β_a.

In larger proteins a mixture of folding patterns occurs, for example, in *p*-hydroxybenzoate hydroxylase[28]. This protein consists of three domains which can be described as $\alpha_p\beta_p$, β_a and α_a.

In non-compact interfaces between subunits in oligomeric proteins, dielectric screening by the solvent will be much greater than in hydrophobic cores. Hence, preferred orientations for α- and β-dipoles from different subunits are expected to be less pronounced.

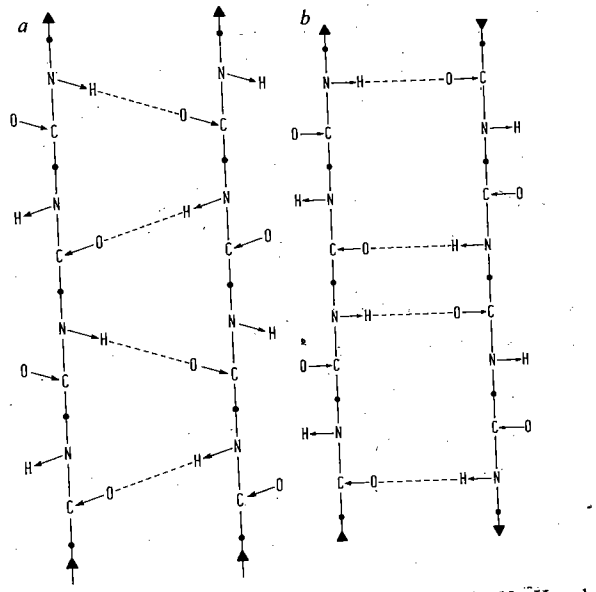

Fig. 2 Schematic representation of the arrangement of the N→H and C←O dipoles in β-sheets. *a*, Pair of parallel β-strands; *b*, pair of antiparallel β-strands. In parallel and antiparallel strands the N→H and O→C bonds have different directions relative to the strand direction. In the antiparallel case the bonds are fairly perpendicular to the strand direction, with little or no resultant dipole moment. In the parallel case, these bonds make angles of ~20° with the N to C strand direction, thus about one-third of the peptide dipole moment runs parallel to the strand. Hence the component along the strand direction is $\mu = 1/3\ \mu(\text{peptide}) = 1/3 \times 3.5\ \text{D} = 1.15\ \text{D} = 0.23\ \text{eÅ}$. As the repeat distance along a β-strand is ~3.5 Å [refs 25, 76], a parallel β-strand can be considered as a dipole with a charge of +1/15 at its N-terminus and −1/15 at its C-terminus. Similar reasoning shows that the dipole of the α-helix can be approximated by a charge couple of +1/2 and −1/2, respectively, at the N and C ends[2]. Thus on a per length basis, the β-parallel strand dipole is about a factor of five weaker than the helix dipole.

Table 2 The predominance of antiparallel over parallel strand pairs in all-β proteins and all-β domains

Protein	Strand pairs β_a	Strand pairs β_p	Ref.
Elastase	11	0	59
Papain (domain II)	6	0	51
Ribonuclease S	3	0	60
Concanavalin A	12	0	61, 62
Rubredoxin	2	0	63
High-potential iron protein	2	0	64
IgG (Fab')	14	1	65
Prealbumin	5	1	66
Neurotoxin	4	0	67, 68
Superoxide dismutase	6	0	69
Bacteriochlorophyll protein	16	1	70
Subtilisin inhibitor	4	0	71
Soybean trypsin inhibitor	3	0	72
Tomato bushy stunt virus	14	0	73
Acid proteases	7	2	74
Total	109	5	

All-β proteins here includes those with a negligible number of helices.

Charged side chains

In addition to the interactions between secondary structure dipoles in the protein interior, electrostatic interactions of charged side chains with each other and with the α-helix dipoles are another component of the total electrostatic energy of a protein molecule. This component will, in most cases, be less important than the interactions between the α and β dipoles. Experimental evidence for this is provided by chemical modification studies in which the charge of exposed residues is removed or even reversed while the catalytic activity of the proteins is little affected[29–31]. In addition, comparison of different amino acid sequences which form strikingly similar structures shows little similarity in the position of residues of similar charge (for example, rhodanese[32] and the serine proteases[33]). A likely explanation for this is effective electrostatic screening of charged residues due to counterions and solvent dipoles[34,35].

Flavodoxin probably has a significant electrostatic interaction between charged residues and α-helix dipoles[36]. In this protein, the charged residues are distributed non-uniformly with respect to the helices: ~10 negative charges occur near the N-termini of the helices without any oppositely charged side-chain partner nearby. The helices in this α/β protein interact quite unfavourably (Table 1), and in this case the large number of favourable charge–helix dipole interactions compensates, together with the favourable α–β dipole interactions, for the unfavourable helix dipole interactions.

The tropomyosin dimer, which is not a globular protein but forms a coil of two very long parallel helices of 284 residues[37], shows the predominance of a large number of salt bridges over the interaction between the helix dipoles. As McLachlan and Stewart observed[37], the charged residues in this protein are situated in such a way that like charges clash when forming antiparallel helices, but as many as 30 salt bridges can be formed in the parallel arrangement. As the helix dipole interaction does not increase after ~30 residues (Fig. 1*d*), the charge–charge interaction is now the major factor in determining the direction of the helices with respect to each other. However, α-helical homopolypeptides, which lack such charge pairs (for example, polyalanine) actually seem to prefer an antiparallel arrangement, both in fibres and in lamellar single crystals[26,38], which is in complete agreement with the results of our model calculations (Fig. 1).

Conclusion

If the values of the electrostatic energies given in Table 1 are compared with the experimentally observed values of 12±5 kcal mol^{-1} for the free energy difference between folded and

unfolded protein conformations[39], it seems that the interaction between the main chain dipole moments of secondary structure elements has a significant role in determining the conformation of a protein molecule.

Our electrostatic considerations are also in accord with the frequent occurrence of $\alpha\alpha$, $\beta\beta$ and $\beta\alpha\beta$ units in proteins[7]. The picture which emerges is that the interior of globular proteins is a medium with low dielectric screening in which the interactions between secondary structure dipoles lead to preferred folding patterns. Kauzman[40] suggested that hydrogen bonds between peptide links and hydrophobic interactions are the most important factors in determining the overall configuration of globular protein molecules. It seems reasonable to suggest that relatively long-range electrostatic interactions in the hydrophobic protein interior should be added to this list.

We thank P. van Duijnen and J. Drenth for stimulating and critical discussions, E. Samulski for his contributions in the early stages of this work, the Groningen University Computer Centre for their cooperation, and C. Moser at CECAM, Orsay, for the organization of workshops, which brought together minds with similar interests. This research was carried out, in part, under the auspices of the Dutch Organisation of Chemical Research (SON), and supported in part by a grant from the Deutsche Forschungsgemeinschaft.

Received 15 May; accepted 5 October 1981.

1. Wada, A. *Adv. Biophys.* **9,** 1–63 (1976).
2. Hol, W. G. J., Van Duijnen, P. Th. & Berendsen, H. J. C. *Nature* **273,** 443–446 (1978).
3. Van Duijnen, P. Th., Thole, B. T. & Hol, W. G. J. *Biophys. Chem.* **9,** 273–280 (1979).
4. Drenth, J., Swen, H. M., Hoogenstraaten, W. & Sluyterman, L. A. Æ. *Proc. K. ned. Akad. Wet.* **C78,** 104–110 (1975).
5. Jernigan, R. L., Miyazawa, S. & Szu, S. C. *Macromolecules* **13,** 518–525 (1980).
6. Blagdon, D. E. & Goodman, M. *Biopolymers* **14,** 241–245 (1975).
7. Levitt, M. & Chothia, C. *Nature* **261,** 552–558 (1976).
8. Bloomer, A. C., Champness, J. N., Bricogne, G., Staden, R. & Klug, R. *Nature* **276,** 362–368 (1978).
9. Stubbs, G., Warren, S. & Holmes, K. *Nature* **267,** 216–221 (1977).
10. Kendrew, J. C. *et al. Nature* **185,** 422–427 (1960).
11. Kretsinger, R. H. & Nockolds, C. E. *J. biol. Chem.* **248,** 3313–3326 (1973).
12. Hendrickson, W. A. & Ward, K. B. *J. biol. Chem.* **252,** 3012–3018 (1977).
13. Banyard, S. H., Stammers, D. K. & Harrison, P. M. *Nature* **271,** 282–284 (1978).
14. Mathews, F. S., Bethge, P. H. & Czerwinski, E. W. *J. biol. Chem.* **254,** 1699–1706 (1979).
15. Weber, P. C. *et al. Nature* **286,** 302–304 (1980).
16. Weber, P. C. & Salemme, R. *Nature* **287,** 82–84 (1980).
17. Mornon, J. P., Fridlansky, F., Bally, R. & Milgrom, E. *J. molec. Biol.* **137,** 415–429 (1980).
18. Poulos, Th. L. *et al. J. biol. Chem.* **255,** 575–580 (1980).
19. Imoto, T., Johnson, L. N., North, A. C. T., Phillips, D. C. & Rupley, J. A. in *The Enzymes* Vol. 7 (ed. Boyer, P. D.) 666–868 (Academic, New York, 1972).
20. Remington, S. J. *et al. J. molec. Biol.* **118,** 81–98 (1978).
21. Colman, P. M., Jansonius, J. N. & Matthews, B. W. *J. molec. Biol.* **70,** 701–724 (1972).
22. Mathews, F. S., Levine, M. & Argos, P. *J. molec. Biol.* **64,** 449–464 (1972).
23. Chothia, C. *J. Molec. Biol.* **75,** 295–302 (1973).
24. Janin, J. & Chothia, C. *J. molec. Biol.* **143,** 95–128 (1980).
25. Arnott, S., Dover, S. D. & Elliot, A. *J. molec. Biol.* **30,** 201–208 (1967).
26. Fraser, R. D. B. & MacRae, T. P. *Conformation in Fibrous Proteins* (Academic, New York, 1973).
27. Leijonmarck, M., Eriksson, S. & Liljas, A. *Nature* **286,** 824–826 (1980).
28. Wierenga, R. K., De Jong, R. J., Kalk, K. H., Hol, W. G. J. & Drenth, J. *J. molec. Biol.* **131,** 55–73 (1979).
29. Cohen, L. A. in *The Enzymes* Vol. 1 (ed. Boyer, P. D.) 147–211 (Academic, New York, 1970).
30. Ottesen, M., Johansen, J. T. & Swendsen, I. in *Structure–Function Relationships of Proteolytic Enzymes* (eds Desnuelle, P. Neurath, H. & Ottesen, M.) 175–187 (Munksgaard, Copenhagen, 1970).
31. Glazer, A. N. in *The Proteins* Vol. 2 (eds Neurath, H. & Hill, R. L.) 1–103 (1976).
32. Ploegman, J. H., Drent, G., Kalk, K. H. & Hol, W. G. J. *J. molec. Biol.* **123,** 557–594 (1978).
33. Delbaere, L. T. J., Hatcheon, W. L. B., James, M. N. G. & Thiessen, W. E. *Nature* **257,** 758–763 (1975).
34. Ehrenson, S. *J. Am. chem. Soc.* **98,** 7510–7514 (1976).
35. Rees, D. C. *J. molec. Biol.* **141,** 323–326 (1980).
36. Burnett, R. M. *et al. J. biol. Chem.* **249,** 4383–4392 (1974).
37. McLachlan, A. D. & Stewart, M. *J. molec. Biol.* **103,** 271–298 (1976).
38. Arnott, S. & Wonacott, A. J. *J. molec. Biol.* **21,** 371–383 (1966).
39. Privalov, P. L. *Adv. Protein Chem.* **33,** 167–241 (1979).
40. Kauzmann, W. *Adv. Protein Chem.* **14,** 1–63 (1959).
41. Dijkstra, B. W., Drenth, J., Kalk, K. H. & Vandermaelen, P. J. *J. molec. Biol.* **124,** 53–60 (1978).
42. Takano, T. *et al. J. biol. Chem.* **252,** 776–785 (1977).
43. Holbrook, J. J., Liljas, A., Steindel, S. J. & Rossmann, M. G. in *The Enzymes* Vol. 11 (ed. Boyer, P. D.) 191–292 (Academic, New York, 1975).
44. Brändén, C.-I., Jörnvall, H., Eklund, H. & Furugren, B. in *The Enzymes* Vol. 11 (ed. Boyer, P. D.) 103–190 (Academic, New York, 1975).
45. Buehner, M., Ford, G. C., Moras, D., Olsen, K. W. & Rossmann, M. G. *J. molec. Biol.* **82,** 563–585 (1974).
46. Schulz, G. E., Elzinga, M., Marx, F. & Schirmer, R. H. *Nature* **250,** 120–123 (1974).
47. Wright, C. S., Alden, R. A. & Kraut, J. *Nature* **221,** 235–242 (1969).
48. Matthews, D. *et al. J. biol. Chem.* **253,** 6946–6954 (1978).
49. Banner, D. W. *et al. Nature* **255,** 609–614 (1975).
50. Hartsuck, J. A. & Lipscomb, W. N. in *The Enzymes* Vol. 3 (ed. Boyer, P. D.) 1–56 (Academic, New York, 1971).
51. Drenth, J., Jansonius, J. N., Koekoek, R. & Wolthers, B. G. in *The Enzymes* Vol. 3 (ed. Boyer, P. D.) 485–499 (Academic, New York, 1971).
52. Baker, E. N. *J. molec. Biol.* **141,** 441–484 (1980).
53. Pollock, E. L., Alder, B. J. & Pratt, L. R. *Proc. natn. Acad. Sci. U.S.A.* **77,** 49–51 (1980).
54. Pethig, T. *Dielectric and Electronic Properties of Biological Materials* (Wiley, New York, 1979).
55. Applequist, J. & Mahr, T. G. *J. Am. chem. Soc.* **88,** 5419 (1966).
56. Kirkwood, J. G. & Westheimer, F. H. *J. chem. Phys.* **6,** 5419–5429 (1938); **7,** 437 (1939).
57. Bernstein, F. C. *et al. J. molec. Biol.* **112,** 535–542 (1977).
58. Levitt, M. & Greer, J. *J. molec. Biol.* **114,** 181–239 (1977).
59. Sawyer, L. *et al. J. molec. Biol.* **118,** 137–208 (1978).
60. Wyckoff, H. W. *et al. J. biol. Chem.* **245,** 305–328 (1970).
61. Hardman, K. D. & Ainsworth, C. F. *Biochemistry* **11,** 4910–4919 (1972).
62. Reeke, G. N., Becker, J. W. & Edelman, G. M. *J. biol. Chem.* **250,** 1525–1547 (1975).
63. Watenpaugh, K. D., Sieker, L. C., Herriott, J. C. & Jensen, L. H. *Acta crystallogr.* **B29,** 943–956 (1973).
64. Carter, C. W. *et al. J. biol. Chem.* **249,** 4212–4225 (1974).
65. Poljak, R. J., Amzel, L. M., Chen, B. L., Phizackerley, R. P. & Saul, F. *Proc. natn. Acad. Sci. U.S.A.* **71,** 3440–3444 (1974).
66. Blake, C. C. F., Geisow, M., Oatley, S. J., Rérat, B. & Rérat, C. *J. molec. Biol.* **121,** 339–356 (1978).
67. Tsernoglou, D. & Petsko, G. A. *Proc. natn. Acad. Sci. U.S.A.* **74,** 971–974 (1977).
68. Low, B. W. *et al. Proc. natn. Acad. Sci. U.S.A.* **73,** 2991–2994 (1976).
69. Richardson, J. S., Thomas, K. A., Rubin, B. H. & Richardson, D. C. *Proc. natn. Acad. Sci. U.S.A.* **72,** 1349–1353 (1975).
70. Matthews, B. W., Fenna, R. A., Bolognesi, M. C., Schmid, M. F. & Olson, J. M. *J. molec. Biol.* **131,** 259–285 (1979).
71. Mitsui, Y., Satow, Y., Watanabe, Y. & Iitaka, Y. *J. molec. Biol.* **131,** 697–724 (1979).
72. Sweet, R. M., Wright, H. T., Janin, J., Chothia, C. H. & Blow, D. M. *Biochemistry* **13,** 4212–4228 (1974).
73. Harrison, S. C., Olson, A. J., Schutt, C. E. & Winkler, F. K. *Nature* **276,** 368–378 (1978).
74. Tang, J., James, M. N. G., Hsu, I. N., Jenkins, J. A. & Blundell, T. L. *Nature* **271,** 618–621 (1978).
75. Hermans, J. & McQueen, J. E. *Acta crystallogr.* **A30,** 730–739 (1974).
76. Pauling, L. & Corey, R. B. *Proc. natn. Acad. Sci. U.S.A.* **39,** 253–256 (1953).

Clustered arrangement of immunoglobulin λ constant region genes in man

Philip A. Hieter*, Gregory F. Hollis*, Stanley J. Korsmeyer†, Thomas A. Waldmann† & Philip Leder*

* Laboratory of Molecular Genetics, National Institute of Child Health and Human Development, and † Metabolism Branch, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20205, USA

The immunoglobulin λ light chain locus of man contains six λ-like genes arranged tandemly on a 50-kilobase segment of chromosomal DNA. The sequences of three of these genes correspond to three known non-allelic λ chain isotypes: Mcg, Ke^-Oz^- and Ke^-Oz^+. They surround a highly polymorphic and evidently unstable region that is repeatedly deleted when cloned in Escherichia coli. *Three additional, but as yet unlinked, λ-like sequences have also been cloned, suggesting that the λ genes form an unexpectedly large family within the human genome.*

THE formation of an active immunoglobulin light chain gene (κ or λ) involves the covalent joining of two distantly encoded segments of germ-line DNA[1–5]. In the case of the κ variable region of the mouse, V region diversity is accounted for largely by the availability of several hundred germ-line variable genes[6,7] that can be joined to one of four active J region segments[8,9]. The