Backbone Dipoles Generate Positive Potentials in all Proteins: Origins and Implications of the Effect

M. R.Gunner, Mohammad A. Saleh, Elizabeth Cross, Asif ud-Doula, and Michael Wise
Physics Department, City College of New York, New York 10031

ABSTRACT Asymmetry in packing the peptide amide dipole results in larger positive than negative regions in proteins of all folding motifs. The average side chain potential in 305 proteins is 109 ± 30 mV (2.5 ± 0.7 kcal/mol/e). Because the backbone has zero net charge, the non-zero potential is unexpected. The larger oxygen at the negative and smaller proton at the positive end of the amide dipole yield positive potentials because: 1) at allowed phi and psi angles residues come off the backbone into the positive end of their own amide dipole, avoiding the large oxygen; and 2) amide dipoles with their carbonyl oxygen surface exposed and amine proton buried make the protein interior more positive. Twice as many amides have their oxygens exposed than their amine protons. The distribution of acidic and basic residues shows the importance of the bias toward positive backbone potentials. Thirty percent of the Asp, Glu, Lys, and Arg are buried. Sixty percent of buried residues are acids, only 40% bases. The positive backbone potential stabilizes ionization of 20% of the acids by >3 pH units (−4.1 kcal/mol). Only 6.5% of the bases are equivalently stabilized by negative regions. The backbone stabilizes bound anions such as phosphates and rarely stabilizes bound cations.

INTRODUCTION

The amide group of the protein backbone is the most prevalent polar group in any protein, and it plays several well established roles in determining protein structure and function. Thus, when a protein folds the backbone NH and C═O groups in the protein interior find hydrogen bonds to replace those made to water in the unfolded polypeptide (Yang and Honig, 1995a, b). The pattern of regular intra-backbone hydrogen bonds yields the protein secondary structures that have been the subject of research going back to the early work of Pauling. Amides in specific motifs have been shown to be important for the stabilization of buried charges. Interactions of charges with the backbone have been identified both by using geometric rules that identify hydrogen bonds (Baker and Hubbard, 1984; Rashin and Honig, 1984; Stickle et al., 1992; McDonald and Thorton, 1994; Gandini et al., 1996) and by calculation of the intra-protein electrostatic potential (Spassov et al., 1997). Interaction of charges with the α-helix dipole (Wada, 1976; Hol et al., 1978; Hol, 1985) have been implicated in increased protein stability (Nicholson et al., 1988; Sali et al., 1988) and in pK_a shifts of acidic and basic residues (Aqvist et al., 1991; Sancho et al., 1992; Sitkoff et al., 1994). Amides in loops also make hydrogen bonds to stabilize charges. The backbone is important in calcium, (Strydnaka and James, 1989), phosphate, and sulfate (Hol, 1985; Quiocho et al., 1987; Jacobson and Quiocho, 1988; Luecke and Quiocho, 1990; He and Quiocho, 1993; Yao et al., 1996) binding sites, and in ion binding in the potassium channel (Doyle et al., 1998). Amides also play important roles in enzyme reactions such as in the oxyanion hole of the serine proteases, where they stabilize the negative charge on the substrate carbonyl in the transition state (James et al., 1980).

Cations are stabilized in regions of negative potential and anions in positive regions. Because the amide group is a dipole, if it is properly oriented it can interact favorably with either charge. However, there is growing evidence that the backbone stabilizes anions more often than cations. For example, there are more bound anions such as phosphate and acidic amino acids at helix N-termini than cations at the C-termini (Hol et al., 1981; Richardson and Richardson, 1988; Gandini et al., 1996). A large positive potential is found at the redox center in iron-sulfur proteins (Langen et al., 1992; Swartz et al., 1996), at the phosphate binding site in α/β barrel proteins (Raychaudhuri et al., 1997), and at a cluster of buried acids in the bacterial photosynthetic reaction centers (Beroza et al., 1995; Lancaster et al., 1996). Calculations show that charges on acidic side chains are better stabilized than bases by the backbone dipoles in aspartate transcarbamylase (Oberoi et al., 1996). The backbone is found to produce a generally positive potential near the protein surface (Spassov et al., 1997). However, there has been no investigation of whether there is a general principle that the potential from the backbone is, on aver-

Received for publication 30 July 1999 and in final form 22 December 1999.

Address reprint requests to Marilyn Gunner, Physics Department, City College of New York, 138th St. and Convent Ave., New York, N.Y. 10031. Tel.: 212-650-5557; Fax: 212-650-6940; E-mail: gunner@sci.ccny.cuny.edu.

Abbreviations used: V_P, the potential averaged over all heavy atoms (C, N, O, and S) on side chains in a protein; V_S, the potential averaged over all heavy atoms in a given side chain; ΔG_{bkn}, the free energy from the electrostatic interaction between a charged or polar side chain or ligand with the backbone amide dipoles. This is calculated with Eq. 1 or 5; ΔG_{rxn}, the change in free energy of a charge from the loss in reaction field energy when a side chain is moved from water and into its location in the protein. This is calculated with Eq. 2.

Inter-conversion of energy units: Electrostatic potential, 1 kcal/mol/e = 42.5 mV; Free energy, 1.36 kcal/mol = 59 meV, will shift a pK_a by 1 pH unit.

0006-3495/00/03/1126/19 $2.00

age, positive, or of how the neutral amide dipoles could produce this result.

While the secondary structure motifs are the most obvious consequence of proteins having an amide linkage, this paper will show that the amide group imposes additional, inescapable consequences for protein structure and function. Most simply, the shape of the amide is dominated by the oxygen of the carbonyl (C═O) being substantially larger than the amine HN hydrogen (Fig. 1). One consequence of this is that to avoid a steric clash the peptide R group is *trans* to the C═O, closer to the HN, at favored phi and psi angles. Moreover, the curvature of a protein's surface favors placing the larger carbonyl oxygen out toward the solvent, while the smaller HN is more likely to be packed in the protein interior. Thus, the asymmetry of the amide group itself imposes an asymmetric packing of the amides within proteins. Electrostatic interactions are the most long-range in proteins. Asymmetry in the orientation of a collection of dipoles, even those that are involved in hydrogen bonds, will generate a significant, non-zero electrostatic potential. This can influence the disposition and energy of the charged groups within proteins.

This paper will describe the analysis of many protein structures to show that the neutral backbone dipoles make the electrostatic potential more positive within proteins of all motifs. It will then be shown how the structure of the amide dipole itself, negative toward the carbonyl oxygen and positive toward the amide proton, produces a non-zero potential in all proteins. Lastly, an analysis of the distribution of acidic and basic side chains and ionized substrates and cofactors in many proteins will show a bias toward burying anions rather than cations, not unexpected if the backbone dipoles make the protein interior more positive. Each protein represents a balance of many forces such as the hydrophobic effect favoring non-polar residues inside a protein and the solvation energy stabilizing charged residues on the surface. The basic geometry of the amide dipole by producing more positive potentials within all proteins adds another term to the forces that influence each protein's folding, structure, and function.

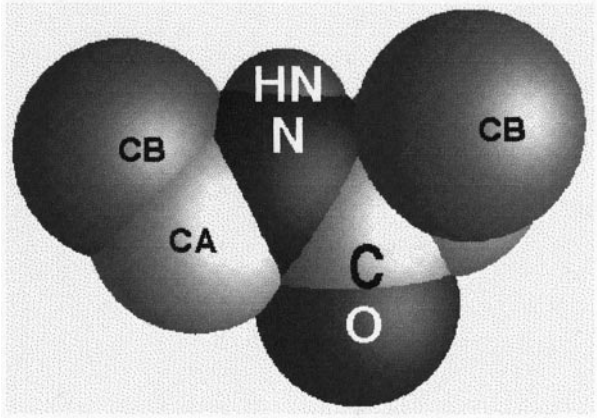

FIGURE 1 Space filling representation of an amide group. The amine HN (r = 1.0 Å) is substantially smaller than the carbonyl oxygen (r = 1.6 Å). The first atom of the two side chains (CB) adjacent to the amide are oriented as they would be in an α-helix.

MATERIALS AND METHODS

Protein structures

Proteins were selected from the Brookhaven data bank (Bernstein et al., 1977) to contain examples of many of folds in the SCOP classification system (Murzin et al., 1995). SCOP classes are α, all α-helix; β, all β-sheet; α/β, mainly parallel β-sheets (β-alpha-β units); $\alpha + \beta$, mainly antiparallel β-sheets (segregated α and β regions); small, usually dominated by metal ligand, heme, and/or disulfide bridges; multi, multi-domain (α and β); membrane, membrane and cell surface proteins and peptides. SCOP classifies domains independently, so proteins can belong to several motifs. When domains in one protein are in different SCOP classes the protein is designated mixed-motif, a group that includes all SCOP multi-domain proteins.

The following 305 proteins were used. The 141 proteins with resolution of ≤1.8 Å are underlined. The 30 structures with resolution ≥2.6 Å are in italics.

α-helix: *1aep,* 1ala, *1bbh*, *1bgc*, 1bgd, 1cce, *1ccr, 1clm, 1cmb, 1cpc*, 1cpt, *1csh*, 1dcc, *1eco*, 1fia, 1gmf, *1hdd, 1hrs,* 1huw, *1hyp*, 1lis, *1lmb*, 1lpe, *1mba, 1mbc*, *1mdy, 1oct,* 1omd, *1par, 1phc*, 1r69, 1rhg, 1rib, *1rop, 1utg, 256b*, 2abk, 2asr, *2ccy*, 2cep, 2cnd, 2cro, 2cts, *2cyp, 2hhb, 2hmq, 2mhr, 2pal, 2wrp*, 2ycc, *351c, 3c2c*, 3gly, 3icb, *4bp2, 5cpv, 5cyt*.

β-sheet: *1aac*, 1acx, *1arb*, *1avd,* 1bbp, *1bcx, 1bgh*, 1cau, 1ctm, 1f3g, 1gcs, *1gct*, *1gof*, 1hbp, *1hcb, 1hlc, 1hmr*, 1hne, 1hoe, 1hvj, *1icm, 1ifc*, 1igm, *1mdc*, 1mjc, 1mup, *1nsc*, *1paz, 1plc, 1pmy*, 1png, *1ppl*, 1pts, *1r1a,* 1rbp, *1scs, 1sgt*, 1shf, *1shg, 1snc*, *1stp, 1ten*, 1tie, *1tld*, 1tnf, *1ton, 1ttb*, 1vmo, *2alp, 2apr, 2ayh, 2aza,* 2ca2, 2cab, *2cpl, 2er7*, 2fb4, *2ltn, 2mcm*, *2mev, 2pab, 2pcy*, 2pec, *2plv, 2psg, 2rhe*, 2rsp, 2sam, *2sga, 2sil*, *2snv,* 2sod, 2stv, *3est, 4fgf, 4gcr, 4pep, 4sbv,* 6nn9.

α/β: *1aba*, 1aco, *1ads*, 1alk, *1amp*, 1bnh, 1cde, *1cus*, 1gdh, 1gpb, 1hmy, 1lct, *1nar*, 1nba, *1nip, 1ofv*, *1omp*, *1rpa,* 1rve, *1s01*, *1sto, 1thg, 1tml, 1tpf*, 1trk, *1 ulb,* 1wht, 2ak3, 2dkb, *2dri*, 2had, *2prk, 2rn2, 2trx, 3chy, 3cla, 3dfr*, 3eca, 3hsc, *4fxn, 5p21*, 7aat, *8abp*.

$\alpha + \beta$: 1aak, 1ahc, *1alc*, 1apa, *1ast*, 1aya, *1brn*, 1cew, *1ctf*, 1dtp, 1fdd, *1fkf*, *1frd*, *1fus*, *1fxd*, 1fxi, *1gmp*, 1iag, *1igd*, 1lba, 1mat, *1mol*, *1npk*, *1pkp, 1ppn*, 1ris, 1rms, *1sha*, *1tbp, 1ubq*, 1yat, 2acg, *2act*, *2bop*, 2chs, 2ci2, 2dnj, 2fxb, 2hpr, *2lzm*, *2ms2, 2msb*, 2pol, *2ssi, 2uce, 3b5c*, 3il8, 4tms, *7rsa*, *9rnt*.

small: *1aap*, *1cbn*, *1fas*, *1isu*, *1nxb*, *1rdg*, *2cdv*, *2ovo*, *2sn3*, *4ins*, *4pti*, *4rxn*, *9wga*.

multi-motif: *1ezm*, 1isb, 1sry, *2tmn*, 3sdp, 1bia, 1chm, *1cse*, 1emd, 1gal, *1glv,* 1lvl, 1pca, *1pda*, 1phh, 1rbl, 2glt, 2npx, 2reb, *2sic*, *3cox*, *3grs*, 4enl, *4gpd,* 4mdh, *5rub*, 9ldt, 2 cmd, 2pia, 8atc, *1dlh,* 1tss, 2aai, *2mha,* 1ddt, 1esl, 1dsb, *1glq*, 1gne, 1hna, *2gst*, 2pgd, 4ts1, 1gia, *1fc2,* 1lla, 1prc, *2bpf,* 3mdd, 1cdg, 1cdo, 1eft, 1hpl, 2aaa, 8adh, *1gla,* 1dlc, *1tnr, 2bbk*, *2por*, 1rpl, *1gma*, *1ppt*.

Crystallographic waters, SO_4, and PO_4 with >10% of their surface exposed to solvent were deleted. The surface exposure was determined with the program SURFV (Sridharan et al., 1992). Protons were added to the proteins with a 1.0 Å bond length and standard geometry.

Calculation of the electrostatic free energy terms for acidic and basic residues

Electrostatic free energy terms were calculated for the ionized form of the acidic residues Asp and Glu and the bases Arg and Lys. DelPhi calculations were run for each residue with charges only on the atoms of this one side

chain. All other atoms in the protein had zero charge. Focusing was used (Gilson et al., 1987) so that the minimum resolution for mapping the atoms and surface to the grid for the finite difference solution of the Poisson equation was 0.83 Å/grid. The dielectric constant for the protein (ϵ_{prot}) was 4, while that of the surrounding solvent (ϵ_{solv}) was 80. For each ionized side chain the same calculation provides the pairwise interactions of the residue with the backbone and its reaction field energy.

Pairwise interactions between the backbone and ionized side chains

The potential was determined at all atoms in the backbone in a protein where a single acidic or basic residue has charge. The free energy of the pairwise interaction between the backbone and side chain i (ΔG_{bkbn}) is:

$$\Delta G^{i}_{bkbn} = \sum_{j=1}^{R} \sum_{bj=1}^{bn} \Psi^{si}_{bj} q_{bj} \quad (1)$$

where Ψ^{si}_{bj} is the potential at atom b in the backbone of the *j*th residue from charges on the *i*th side chain. This pairwise interaction was obtained for the bn atoms of the backbone that bear partial charge (q_a) (Table 1). The interaction was then summed for all R backbone amides in the protein.

Reaction field energy

The reaction field energy (also referred to as the self, solvation, or Born energy) measures the difference in energy of an ion or dipole when it is transferred between media with different abilities to reorganize around charges. Electronic polarization and rearrangement of atomic dipoles both contribute. Using continuum electrostatic theory, the response of the media is encapsulated in the dielectric constant. The reaction field energy is calculated here using an algorithm in DelPhi, which determines the interaction energy between the charges on the protein atoms and charges induced at the protein-water dielectric boundary (Nicholls and Honig, 1991; Sridharan et al., 1992).

The penalty for placing a charge at its location in the protein is the difference between the reaction field energy of the residue in situ and the reaction field energy of the same residue isolated from the protein:

$$\Delta G_{rxn} = \Delta G_{rxn\ in\ protein} - \Delta G_{rxn\ in\ soln} \quad (2)$$

$\Delta G_{rxn\ in\ protein}$ and $\Delta G_{rxn\ in\ soln}$ are both negative, favorable terms. ΔG_{rxn} is always a positive, unfavorable energy term because the absolute value of $\Delta G_{rxn\ in\ protein}$ is always less than $\Delta G_{rxn\ in\ soln}$. The reaction field energy for side chains in solution were obtained for isolated coordinates of each side chain in the protein data bank file 1PRC (Table 2). There is very little variation between different conformers of any side chain, so one reference value is used for each type of residue.

TABLE 1 The charges used on the atoms of the backbone amides

	CHARMM	EQ	Carbonyl	Amine
C	0.55	0.55	0.55	—
O	−0.55	−0.55	−0.55	—
HN	0.25	0.35	—	0.25
N	−0.35	−0.35	—	−0.35
CA	0.10	—	—	0.10
Proline				
C	0.55	0.55	0.55	—
O	−0.55	−0.55	−0.55	—
N	−0.20	−0.20	—	−0.20
CD	0.10	0.20	—	0.10
CA	0.10	0.00	—	0.10

All calculations use CHARMM charges unless otherwise noted.

Calculation of interactions between the backbone and all side chains and bound ligands

Average potential in the protein

The potential was calculated by placing partial charges on all backbone amides. A DelPhi calculation was carried out with a 129^3 grid. This provides a grid spacing of >1.0 Å/grid for all but 30 proteins. The potential (Ψ^{bkbn}_a) from the backbone at each of the *m* non-backbone heavy atoms (a) was averaged to determine V_P. The potential at waters and other non-protein atoms was not included in the sum.

$$V_p = \frac{1}{m} \sum_{a=1}^{m} \Psi^{bkbn}_a \quad (3a)$$

In a group of *N* proteins the average of V_P is:

$$AvV_p = \frac{1}{N} \sum^{N} V_p \quad (3b)$$

The average potential (V_S) from the backbone at a residue was obtained from:

$$V_S = \frac{1}{n} \sum_{a=1}^{n} \Psi^{bkbn}_a \quad (4a)$$

where there are *n* non-backbone heavy atoms (a) in the side chain

In a group of R residues the average of V_S is:

$$AvV_S = \frac{1}{R} \sum^{R} V_S \quad (4b)$$

The free energy of interaction of the *j*th side chain or ligand with the backbone is:

$$\Delta G^{j}_{bkbn} = \sum_{a=1}^{n} \Psi^{bkbn}_{aj} q_a \quad (5)$$

where q_a is the charge on atom a in an appropriate partial charge set. The free energy of interaction of a side chain or ligand with the backbone (ΔG_{bkbn}) can be calculated with either Eq. 1 or 5. For Eq. 1 the side chain is charged and the potential is collected at all the atoms of the backbone.

TABLE 2 Reaction field energy in solution for acidic and basic side chains

	Kcal/mol	ΔpH units	Number of residues
Asp	−17.6 ± 0.1	−12.9 ± 0.0	47
Glu	−17.5 ± 0.1	−12.8 ± 0.1	51
Arg	−16.0 ± 0.1	−11.8 ± 0.0	66
Lys	−19.3 ± 0.1	−14.2 ± 0.1	34

Side chain coordinates from the file 1PRC were isolated from the rest of the protein. CHARMM charges were placed on the atoms. The net charge is −1 for Asp and Glu and +1 for Arg and Lys. The dielectric constants were $\varepsilon_{atoms} = 4$; $\varepsilon_{solvent} = 80$.

For Eq. 5, the backbone is charged and the potential is collected at the side chain atoms.

Unless otherwise noted, calculations of V_P, V_S, and ΔG_{bkn} use CHARMM partial atomic charges for backbone (Table 1) and side chains (Brooks et al., 1983); ϵ_{prot} is 4 and ϵ_{solv} is 80. The atomic radii were for each atom type H 1.2 Å, C 1.8 Å, N 1.5 Å, O 1.6 Å, S 1.9 Å, P 1.2 Å.

Interaction between side chains and specific amide groups

The interaction of each side chain with each amide was calculated in 51 proteins. Each DelPhi calculation had partial charges on only one amide group. Thus, R calculations were made for a protein with R residues. The grid resolution was >0.83 Å/grid for each protein. Where necessary the focusing technique was used centered on the amide that carried the partial charges (Gilson et al., 1987). The net charge was 0 in each run, resulting from ± 0.9 charge for a standard amide and ± 0.75 for Pro. Equations 3 and 4 were used to calculate the average potential from each amide within the protein or at specific side chains; Eq. 5 provided the free energy of interaction between specific side chains and individual amides.

Potential at CB from amide(n) and amide(c) as a function of the phi and psi angle

All non-terminal amino acids in a protein lie between an amide toward the N-terminal (amide(n)) and one toward the C-terminal (amide(c)) (Fig. 2). Two series of 36 Ala tripeptide coordinates were constructed. In one set the phi angle was changed in increments of 10°, in the other the psi angle was varied. For the series with different phi angles, all atoms toward the N-terminal were rotated holding the central CA and CB and all atoms toward the C-terminal rigid. The series with different psi angles were constructed holding the N-terminal and the central CA and CB fixed and rotating atoms toward the C-terminal.

The potential at the central CB was obtained using Coulomb's law assuming a uniform dielectric constant of 4. Calculations with the tripeptides surrounded by water (ϵ_{prot} = 4; ϵ_{solv} = 80) were calculated with DelPhi. In this case the positions of all atoms in the tripeptide modify the dielectric boundary, and so effect the results. The variation of phi was carried out in tripeptides where psi is −60°, while the psi rotation was carried out in peptides where phi is 120°.

Comparing the surface exposure of the carbonyl O and amine HN for each amide

In the standard protein, the N to HN distance is 1.0 Å and the H radius is 1.2 Å. In contrast the average C to O bond length is 1.23 Å and the O radius is 1.6 Å. This geometry ensures that the O will have more surface to expose to solvent than the HN does. Protein coordinates were prepared where the HN to N distance was 1.23 Å and the HN radius was 1.6 Å. The surface exposure of the O and the modified HN to a 1.4 Å probe were calculated with the program SURFV (Sridharan et al., 1992).

The in situ pK_a of acidic and basic residues

The pK_a of acids or bases in proteins can be different from that found in solution because interactions in the protein shift the relative energy of residue or ligand charged and neutral state (Churg and Warshel, 1986; Bashford and Karplus, 1990; Gunner and Honig, 1991; Yang et al., 1993; Antosiewicz et al., 1994; Gunner et al., 1997). The complete calculation of residue ionization states is beyond the scope of this paper. However, other interactions in the protein will modify the expected effects of ΔG_{rxn} and ΔG_{bkn}. Thus, if the charge state of all other R residues were fixed, the pK_a of residue i would be shifted from its value in solution ($pK_{soln,i}$) in the following way:

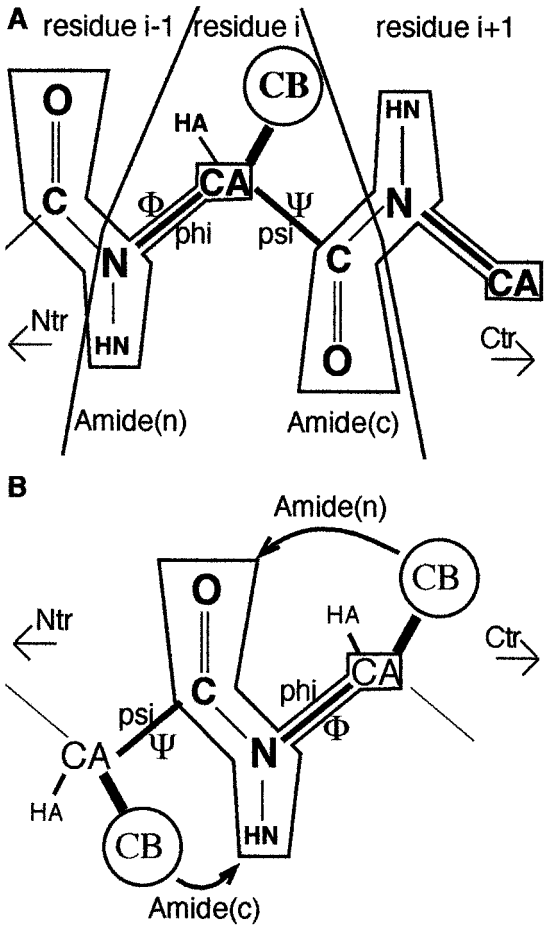

FIGURE 2 Each non-terminal side chain lies between 2 amides, one toward the N-terminal and the other toward the C-terminal. (*A*) The amides toward the N-terminal (amide(n)) and C-terminal (amide(c)) of the side chain of residue i. (*B*) One amide is amide(c) for one side chain (i) and is amide(n) for the next side chain (i + 1) in the protein.

$$pK_{prot,i} = pK_{soln,i} + \Delta G^{crg}_{rxn,i} + \Delta G^{crg}_{bkn,i} + \Delta G^{crg}_{other,i} - \Delta G^{neu}_{rxn,i} - \Delta G^{neu}_{bkn,i} - \Delta G^{neu}_{other,i} \sum_{j=1}^{R} (\Delta G^{crg}_{sdchn(j),i} - \Delta G^{neu}_{sdchn(j),i}) \quad (6)$$

The terms $\Delta G^{crg}_{bjkn,i}$ and $\Delta G^{crg}_{rxn,i}$, the charged residue's interaction with the backbone and its reaction field energy, are calculated with Eqs. 1 and 2, respectively, and will be described in detail here. The interactions of the neutral forms of a residue ($\Delta G^{neu}_{bkn,i}$ and $\Delta G^{neu}_{rxn,i}$) are often small. The final sum represents the difference in the pairwise interactions of the j other polar and charged side chains with residue i in its charged and neutral form. This is the most significant omitted term. Other terms can arise from intra-protein motions that are coupled to the ionization of the residue (ΔG_{other}). Within the protein the charge state of all residues are interdependent (see Bashford and Karplus, 1990; Yang et al., 1993; Antosiewicz et al., 1994; Alexov and Gunner, 1997 for a more complete description).

RESULTS

The potential from the backbone within proteins

Backbone potential within four representative proteins

The degree to which the backbone amides make protein interiors more positive is shown graphically for four proteins with the basic folding motifs: α, β, α/β, and $\alpha + \beta$. The potential at a representative slice through each protein with only backbone dipoles assigned partial charges is visualized with the program GRASP (Nicholls et al., 1991) (Fig. 3). Although the net charge on each protein is zero, the interior is predominately positive. At least a quarter of the total volume of each protein is at a potential above 120 mV, while <10% is below −120 mV (Table 3).

TABLE 3 Electrostatic potentials within four proteins with different folds

	PDB File			
	2HMQ	1HMR	1TPF	7RSA
	Protein Motif			
	α	β	α/β	$\alpha + \beta$
Percent protein at >120 mV	24%	24%	31%	29%
Percent protein at <−120 mV	6.0%	6.6%	9.7%	8.1%
Total volume (Å^3)	15432	18256	32626	15684
V_P (mv)	85	85	131	89

The volume of the protein and the volume within an isopotential contour at ±120 mV (±2.78 kcal/e) was calculated with GRASP (Nicholls et al., 1991). V_P was calculated with Eq. 3a. The proteins are described in the legend to Fig. 3.

Average potential from the amide backbone inside all proteins

The potential from the backbone (V_P) was determined in 305 proteins chosen to include representatives of many folding motifs (Eq. 3a). V_P determines the potential at non-polar, polar, and ionizable side chains. V_P is always positive, ranging from 57 to 244 mV (1.3–5.6 kcal/mol/e). The average V_P is 110 ± 30 mV (2.54 ± 0.70 kcal/mol/e) (Eq. 3b, Fig. 4).

The average potential from the backbone is positive for all protein motifs. Helical proteins have on average the smallest potentials (95 ± 23 mV) and α/β proteins the

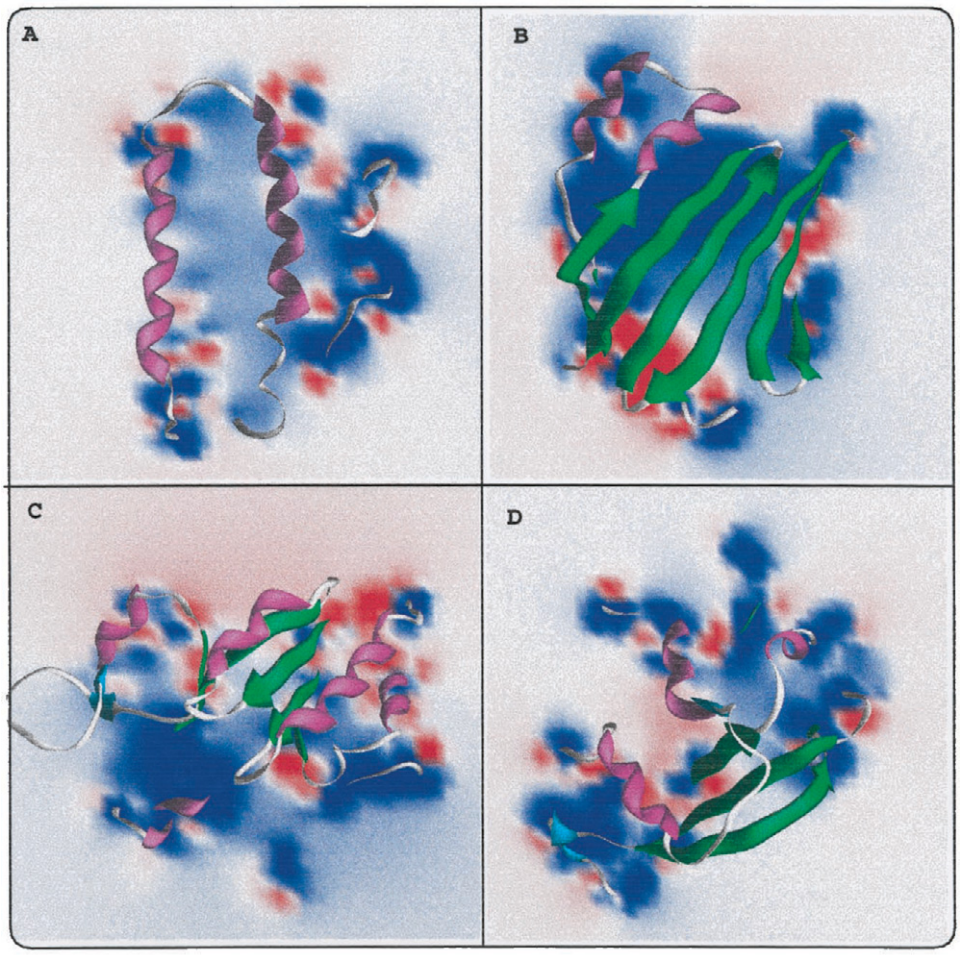

FIGURE 3 Electrostatic potential at a slice through four proteins with different folds. Potentials calculated and displayed with the program GRASP (Nicholls et al., 1991). Blue regions are at positive and red at negative potential; CHARMM charges, $\epsilon_{protein} = 4$; $\epsilon_{solvent} = 80$. (*A*) α motif: Met-hemerythrin from sipunculid worm (*Themiste dyscrita*) (2HMQ chain A) (Holmes and Stenkamp, 1991). A 104 residue iron-binding protein in a four-helical up-and-down bundle with a left-handed twist (Motif descriptions from the SCOP data base (Murzin et al., 1995)). (*B*) β motif: human lipid binding protein (1HMR) (Zanotti et al., 1992). A 129 residue 10-stranded meander β-sheet folded upon itself. (*C*) α/β motif: triose phosphate isomerase from *Trypanosoma brucie brucei* (1TPF) (Kishan et al., 1994); a 247 residue α/β barrel which has 8 alternating α and β segments forming an internal, parallel β-sheet barrel; and (*D*) $\alpha + \beta$ motif: bovine ribonuclease A (7RSA) (Wlodawer et al., 1988); A 124 residue protein with a long curved β-sheet and 3 α-helices.

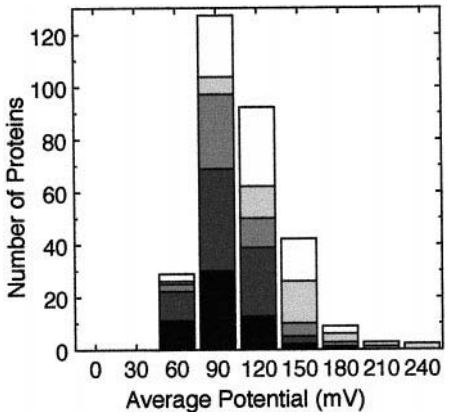

FIGURE 4 The number of proteins with different values of the average electrostatic potential at the side chain heavy atoms (V_P). V_P was calculated with Eq. 3a for 305 proteins. The patterns for different SCOP protein motifs: α, black; β, horizontal; $\alpha + \beta$, diagonal; α/β, cross-hatch; others, white.

largest (136 ± 36 mV) (Table 4). There are more small or pure α or β proteins among the least positive proteins, and more α/β or mixed motif proteins among the most positive. However, all folds are represented in both the most and least positive proteins studied except for the small proteins.

Importance of specific parameters used in the calculations

The dielectric constants for protein and solvent were varied to determine whether the bias toward the backbone potentials being positive is due to the specific parameters used (Table 4). If the calculations use a uniform dielectric constant of 4, rather than having an ϵ_{solv} of 80, the average potential of the proteins tested is 137 ± 62 mV. Thus the result does not depend on the high dielectric constant of the solvent. Raising the interior dielectric constant diminishes V_P without changing its sign (data not shown). The charge distribution can also be varied. For example, moving the 0.1 charge placed on CA in the CHARMM charge set to the HN (EQ charge in Table 1) also yields a positive average potential (93 ± 34 mV).

It is possible to determine the relative importance of the atoms that make up the backbone dipoles in determining V_P. Each amide can be viewed as two smaller dipoles with zero net charge: a unit made of the carbonyl (C and O) and one of the amine (HN, N, and CA) (Table 1). For each protein ~77% of the average potential is a result of the C—O dipole while 22% results from the HN-N-CA charges (Fig. 5). The same relative importance can be found in the contribution of each mini-dipole to the dipole moment of the amide. Thus, an amide with CHARMM charges has a dipole moment of 4.2 D. The carbonyl mini-dipole moment is 3.2 D, representing 76% of the total, while it is 1.0 D for the amine.

TABLE 4 The average potential at all non-hydrogen, side chain atoms from the backbone dipoles inside 305 proteins

Protein Motif	Average V_P (mV)	A*	B†	C‡	D§	E¶
$\epsilon_{protein}$ = 4, $\epsilon_{solvent}$ = 80, and CHARMM charges						
α	95 ± 23	12	3	57	34	71
β	99 ± 21	11	1	79	30	52
α/β	136 ± 36	1	12	43	35	64
$\alpha + \beta$	108 ± 28	3	4	50	42	79
Small	101 ± 28	3	0	13	9	38
Multi-motif	121 ± 27	0	10	63	—	—
All proteins	109 ± 30	—	—	305	154	327
Proteins with resolution ≤1.8 Å	109 ± 29	—	—	143	—	—
$\epsilon_{protein}$ = 4, $\epsilon_{solvent}$ = 80, charge distributions from Table 1						
All proteins "carbonyl" charges	97 ± 24					
All proteins "amine" charges	13 ± 8					
All proteins "EQ" charge	93 ± 34					
$\epsilon_{protein}$ = $\epsilon_{solvent}$ = 4 and CHARMM charges						
All proteins	137 ± 62					

Av V_P is calculated with Eq. 3b.
*A How many of the least positive 30 proteins (V_P: 57–74) are in each class.
†B How many of the most positive 30 proteins (V_P: 150–244 mV) are in each class.
‡C Number of proteins analyzed in each class.
§D Number of folds studied in each class.
¶E Number of folds in each class in the SCOP classification system (Feb. 1997).

Average potential at different types of side chains

The average potential was determined at each side chain (V_S) (Table 5). Only 2.0% of the residues are at potentials below −60 mV, while 75.6% are more positive than +60 mV. The average of V_S is always positive for all types of residues, ranging from 228 mV for Ala to 32 mV for Arg. The average side chain potential is most positive for small groups such as Ala, Cys, and Ser, and decreases as the side chain becomes larger. This results in the average V_S for all

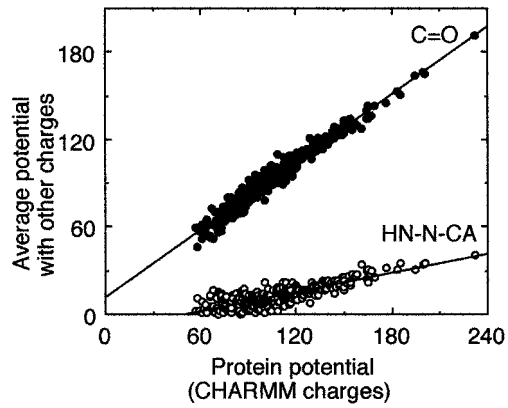

FIGURE 5 Comparison of the average potential at side chain heavy atoms (V_P) for proteins with different charges on the backbone. V_P was calculated with Eq. 3a. Charges from Table 1: (○), amine (HN, N, CA) charges; (●), carbonyl (C, O) charges. The straight lines are described by: 11.91 + 0.77x (r^2 = 0.96) and −11.2 + 0.22x (r^2 = 0.71)

TABLE 5 The distribution of side chains at different potentials from the backbone amide dipoles

	Percentage of the Side Chains within Given Range for V_S (mV)										
	<-300	-300 to -180	-180 to -60	-60 to 60	60 to 180	180 to 300	>300			Average ΔG_{bkn}	
kcal/mole/e	<-6.8	-6.8 to -4.1	-4.1 to -1.36	-1.36 to 1.36	1.36 to 4.1	4.1 to 6.8	>6.8	Atoms/ Residue	Average V_S (mV)	(meV)	(kcal/mol)	Number of Residues
ALA	0.01	0.05	0.29	2.89	32.56	44.94	19.26	1	228	—	—	7544
PRO	—	0.02	0.31	1.40	31.12	50.52	16.62		225			4145
CYS	—	—	0.89	9.58	42.13	29.07	18.33	2	203	18	0.41	1462
SER	0.03	0.20	0.91	8.10	47.43	25.79	17.54	2	193	0	−0.01	5941
ASP	**0.02**	**0.08**	**0.44**	**16.44**	**48.27**	**21.14**	**13.61**	**4**	**169**	**−148**	**−3.40**	**5188**
THR	0.04	0.30	1.04	10.23	54.82	21.56	12.02	3	164	2	0.05	5659
VAL	—	0.03	0.60	9.46	53.64	27.07	9.19	3	163	—	—	6298
ILE	0.02	0.12	0.75	14.27	55.25	22.91	6.67	4	145	—	—	4827
LEU	0.04	0.12	1.30	19.87	54.94	18.55	5.17	4	126	—	—	7459
ASN	0.19	0.26	2.16	28.77	43.48	16.68	8.45	4	124	−18	−0.41	4167
MET	—	0.12	2.60	24.59	50.29	16.03	6.37	4	121	1	0.01	1728
GLU	**—**	**0.06**	**0.39**	**34.28**	**49.63**	**10.31**	**5.34**	**5**	**108**	**−68**	**−1.58**	**5190**
PHE	—	0.37	1.93	29.57	51.49	13.13	3.52	7	103	—	—	3527
TRP	—	0.53	3.17	32.28	47.29	13.27	3.47	10	96	−12	−0.28	1326
TYR	0.03	0.24	1.96	34.49	48.89	10.03	4.36	8	95	−6	−0.13	3369
HIS	0.05	0.68	4.85	43.07	34.46	9.91	6.99	6	91	47	1.08	1918
GLN	—	0.12	2.88	44.60	40.43	8.38	3.59	5	84	−14	−0.31	3233
LYS	**0.27**	**0.59**	**3.20**	**53.39**	**39.33**	**2.65**	**0.57**	**5**	**52**	**−15**	**−0.35**	**5093**
ARG	**0.97**	**1.71**	**8.37**	**57.06**	**27.11**	**3.39**	**1.40**	**7**	**32**	**−29**	**−0.67**	**3929**
All	0.09	0.26	1.66	22.39	45.50	20.93	9.16		140			82003

The rows are placed with descending values of the average of V_S.

V_S for each side chain is calculated with Eq. 4a. $\varepsilon_{protein} = 4$; $\varepsilon_{solvent} = 80$; CHARMM charges are used. The potential range with the largest fraction of side chains is underlined. The residues that are likely to be ionized are in boldface. The Average V_S is calculated with Eq. 4b.

The number of atoms/residue counts the heavy atoms in each side chain. The ΔG_{bkn} is calculated for each side chain with Eq. 5 using CHARMM charges on the side chains.

side chains being more positive than the average V_P for all proteins. The smaller, more positive side chains contribute as much as a large side chain to the average of V_S, but not V_P.

Potential at small molecules, cofactors, and substrates bound to proteins

There are many ligands bound to the proteins analyzed here. The potential from the backbone was investigated at several types of bound molecules (see Table 6).

The average potential at buried waters is positive, with twice as many waters at potentials $>+60$ mV than at <-60 mV. Thus, these neutral dipoles are likely to be found at positive potential.

Metals are the only bound cations that are present in any abundance in proteins. Many of the divalent cations cadmium, cobalt, copper, non-heme iron, manganese, magnesium, ytterbium, and zinc are at potentials from the backbone >300 mV. Only Ca^{2+} and Na^{+} are ever found at potentials from the backbone more negative than -70 mV. The importance of specialized backbone motifs for coordinating Ca^{2+} is well established (Strydnaka and James, 1989; McPhalen et al., 1991). Thus, the bias toward the backbone being positive inside proteins extends even toward the binding sites for positive ions. With the exception of calcium and sodium, the backbone substantially destabilizes cation binding. These must be bound by protein side chains or anionic ligands.

The positive potential from the backbone at iron sulfur clusters has been previously described (Langen et al., 1992; Swartz et al., 1996). The very positive potential strongly favors the reduced over the oxidized form of these redox sites.

Many enzyme substrates such as ATP or GTP are nucleotides, while many cofactors such as flavins and nicotinamides are derived from nucleotides. Each has negatively charged phosphate groups. The average potential at the phosphates is 435 mV, which will substantially stabilize binding. Small anions such as phosphate or sulfate are also always bound in regions of positive potential from the backbone.

Structure of the amide group yields the imbalance between positive and negative regions generated by the protein backbone

Role of the neighboring amides in generating the bias toward positive potentials in proteins

The potential from each amide at each side chain was determined for 51 proteins that sample several folds and

TABLE 6 The distribution of ligands and cofactors at different potentials from the backbone amide dipoles

	Fraction of Ligands within Given Range for V_S (mV)								
kcal/mole/e	<−300 <−6.8	−300 to −180 −6.8 to −4.1	−180 to −60 −4.1 to −1.36	−60 to 60 −1.36 to 1.36	60 to 180 1.36 to 4.1	180 to 300 4.1 to 6.8	>300 >6.8	Average V_S (mV)	Number
HOH	3.7	6.0	14.4	**30.1**	18.8	11.9	15.1	76	7489
Cations									
Ca	**37.1**	4.3	11.4	25.7	14.3	2.9	4.3	−196	70
Na	**57.1**	—	—	—	—	—	42.9	−159	7
Mn	—	—	—	**60.0**	20.0	20.0	—	98	5
Cu	—	—	7.7	**38.5**	30.8	7.7	15.4	133	13
Zn	—	—	—	**34.6**	15.4	15.4	34.6	278	26
Fe	—	—	—	—	**62.5**	—	37.5	320	8
Mg	—	—	—	—	11.1	22.2	**66.7**	426	9
Anions									
Cl	—	—	—	33.3	0.0	**66.7**	—	168	3
PO_4	—	—	—	22.2	11.1	22.2	**44.4**	264	9
SO_4	—	—	—	22.2	—	33.3	**44.4**	282	9
MO_4	—	—	—	—	—	—	**100.0**	503	2
Cofactors									
Heme	—	—	5.9	**61.8**	23.5	5.9	2.9	47	34
FeS	—	—	—	—	—	8.3	**91.7**	701	12
P*	—	—	—	13.33	7.78	14.4	**64.4**	435	90

Only waters, PO_4, and SO_4 buried in the protein with less than 10% of their surface exposed to a 1.4-Å probe were considered.
*Phosphate bound to cofactors and substrates such as nucleotides, nicotinamides, and flavins.

include the most and least positive V_p values in each structural class (Table 7). This group of proteins is slightly more positive than the 305 proteins, yielding the small differences among Tables 5–7.

Each non-terminal side chain lies between two neighboring amides, one toward the N-terminal, the other toward the C-terminal (Fig. 2). All other amides in the protein are distal to this side chain. Phi and psi angles define the neighboring amide orientation, secondary and tertiary structures produce the arrangement of the distal amides. Analysis of the potential from neighboring and distal amides shows: 1) the potential from the neighboring amides is always positive; 2) the standard deviation of this potential increases as the flexibility of the side chain increases; 3) the potential from the distal amides is very variable, as seen in the large standard deviation of this value for each type of residue; 4) on average, the distal amides also raise the potential at all residues except at the bases Arg and Lys; and 5) the average potential for Cys from the distal amides is very positive. This is largely due to the very positive values at the Cys that are ligands in iron-sulfur clusters (Table 6) which are overrepresented in the group of proteins.

The potential at a side chain (V_S) is the sum of the potential from the neighboring and the distal amides (Fig. 6). The neighboring amides contribute 122 ± 68 mV to the average. The relative constancy of this value shows that, independent of protein motif, the potential from the backbone starts with a bias of ~110 mV within all proteins. Proteins with average potentials less than this have contributions from each group's distal amides that are on average negative. The average potential from the distal amides in the different proteins ranges from −40 to 120 meV, extending to higher positive than negative values.

Why the potential from the neighboring amides is always positive

The potential from the neighboring amides at CB in a medium of uniform dielectric constant is solely determined by the phi angle (for amide(n)) and the psi angle (amide(c)) (Fig. 2). Under these simplified conditions it becomes clear why the potential from the neighboring amides at any residue is almost always positive. The impact of the surrounding solvent and extended side chains on the potential and resulting ΔG_{bkn} will be described below.

The potential is shown visually for an amide group along with the CBs for which this is amide(n) and amide(c) (Fig. 2 and 7). The polypeptide chains are arranged with phi and psi angles found in α-helices or β-sheets. In each case the CBs toward the N- or the C-terminal are in the region of positive potential from the amide.

The potential was determined as a function of the phi and psi angles at the middle CB in an Ala-tripeptide (Fig. 8). The potential from amide(n) is less than zero only for phi values between 40° and 180°, a region that is unfavorable for any residue but Gly because of steric hindrance between CB (of residue i) and the amide(n) (residue i-1) carbonyl oxygen (Ramachandran et al., 1974). Thus, the side chain is

TABLE 7 The contribution of the neighboring and distal amides to the potential at different amino acids

	Total	Distal	Neighbor	Amide (n)	Amide (c)	Number of Residues
ALA	239 ± 144	13 ± 135	227 ± 67	124 ± 44	103 ± 40	1137
PRO	228 ± 120	28 ± 102	200 ± 55	77 ± 37	123 ± 34	588
CYS	225 ± 198	71 ± 181	155 ± 61	78 ± 42	76 ± 39	178
SER	206 ± 178	45 ± 166	162 ± 55	90 ± 40	71 ± 34	957
ASP	**190 ± 183**	**89 ± 172**	**101 ± 46**	**52 ± 32**	**49 ± 29**	**794**
THR	172 ± 150	35 ± 135	137 ± 52	75 ± 35	61 ± 35	880
VAL	163 ± 102	17 ± 100	147 ± 42	77 ± 31	69 ± 31	906
ILE	150 ± 130	27 ± 124	123 ± 42	62 ± 28	62 ± 28	644
LEU	130 ± 103	17 ± 104	113 ± 41	56 ± 27	57 ± 28	1025
ASN	133 ± 136	39 ± 127	94 ± 47	47 ± 32	47 ± 29	636
MET	123 ± 124	27 ± 123	96 ± 40	49 ± 29	47 ± 29	254
GLU	**123 ± 132**	**39 ± 117**	**84 ± 37**	**43 ± 22**	**41 ± 25**	**656**
PHE	110 ± 105	37 ± 101	73 ± 36	35 ± 29	38 ± 26	524
TRP	98 ± 95	35 ± 96	62 ± 31	32 ± 24	31 ± 25	177
TYR	105 ± 114	41 ± 109	64 ± 32	31 ± 25	34 ± 22	469
HIS	108 ± 164	30 ± 157	78 ± 40	40 ± 28	39 ± 28	277
GLN	90 ± 120	8 ± 113	82 ± 34	41 ± 23	41 ± 22	537
LYS	**48 ± 66**	**−27 ± 66**	**75 ± 30**	**39 ± 20**	**36 ± 19**	**719**
ARG	**35 ± 112**	**−23 ± 108**	**58 ± 29**	**30 ± 19**	**28 ± 20**	**522**
ALL	149 ± 145	27 ± 128	122 ± 68	63 ± 41	60 ± 39	11880

The potential was determined placing partial charges on one amide at a time in 51 proteins. The neighboring amides, amide (n) and amide (c), are defined in Fig. 2. All other amides are distal to a side chain.

constrained to come off the backbone into the positive rather than the negative end of amide(n) because the carbonyl oxygen has a van der Waals radius that is much larger than the HN. The phi angles in α-helices lie close to the maximum value of the potential, while β-sheets rotate the side chain into regions of lower potential from amide(n).

The potential from amide(c) is always positive, in part because the carbonyl C is always closer than the O to the CB. The region of maximum potential is at values for psi that are disallowed. The potential in helical regions is slightly larger than for β-sheets.

The potential at CB from the neighboring amides is influenced by the dielectric properties of the surrounding solvent. Thus, the isopotential contours from an amide group are smaller when the amide is immersed in solvent (Fig. 7). However, the pattern of the variation of the potential with phi and psi is independent of solvent (Fig. 9).

As the side chains become longer the potential from the neighboring amides decreases (Fig. 8). A decrease in the positive potential along individual side chains was noted previously by Spassov (Spassov et al., 1997). In addition, longer side chains have more allowable rotomers with atoms in different positions relative to the amide dipole, which increases the deviation from the average potential (Table 7).

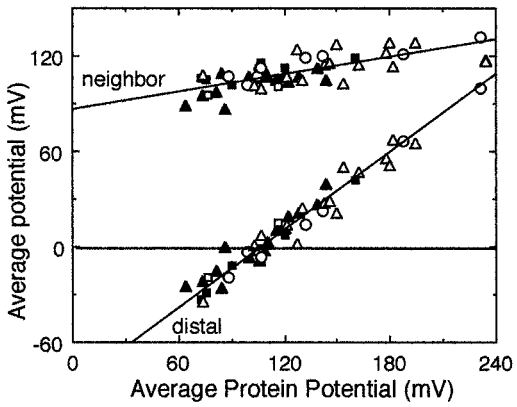

FIGURE 6 Comparison of the contribution of the neighbor and distal amides to the average potential for 51 proteins. Each residue is charged in turn in each protein and the potential collected at the two neighboring side chains and at the distal side chains. Different protein motifs: α, ■; β, □; $\alpha + \beta$, ▲; α/β, ○; others, △. The straight lines are described by neighboring amides, 86.91 + 0.18x ($r^2 = 0.56$); and distal amides, −86.9 + 0.82x ($r^2 = 0.96$)

The amide orientation relative to the protein surface affects the intra-protein potential

Modified protein structures were prepared where the HN to N bond in the amide amine was lengthened to be as long as the O to C bond in the carbonyl and the HN radius was increased to the size of the O. The surface accessibility of O and HN in these modified structures provides a simple, rough estimate of whether each amide points its carbonyl or amine out toward the solvent. With few exceptions, if an amide O is more surface-exposed than its HN, this amide raises the potential in the protein (*top right quadrant* of Fig.

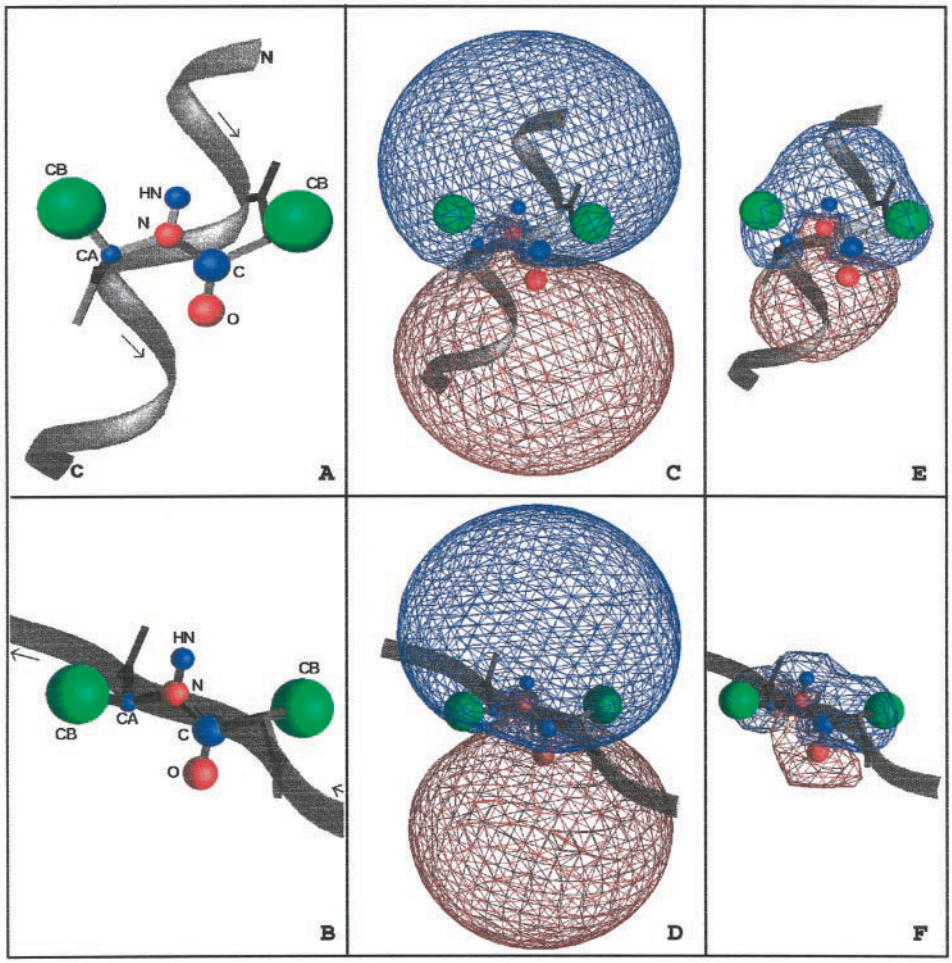

FIGURE 7 Each amide forms the junction between two residues (Fig. 2 *B*). One amide is amide(c) for residue (i), with an orientation between side chain and amide determined by the psi angle. The same amide is amide(n) for the next side chain (i + 1) and their orientation is described by the phi angle. GRASP (Nicholls et al., 1991) pictures showing the two CBs (*green*) neighboring one amide in (*A*) α-helix ($\phi = -52$, $\psi = -53$); (*B*) β-strand ($\phi = -123$, $\psi = 143$). The five atoms assigned charge are labeled, colored red (negative) or blue (positive), and given a radius that is proportional to the partial charge. The isopotential contours at +0.85 kcal/e (*blue*) and −0.85 kcal/e (*red*) calculated with (*C*, *D*) $\epsilon_{peptide} = \epsilon_{solv} = 4$; and (*E*, *F*) $\epsilon_{peptide} = 4$, $\epsilon_{solv} = 80$.

10). If the O is more buried the amide lowers the potential (*bottom left quadrant*). The same pattern is found for α-helical, β-sheet, and random coil regions of all protein folds.

The total contribution to the potential from amides with HN more exposed, O more exposed, or with little difference between their exposure were compared (Table 8). The residues that have little differential exposure contribute only a small amount to the average potential within the protein. For each protein the contribution per amide for those with the O or the HN more exposed are of similar magnitude, but opposite sign. However, there are always more amides where the O surface exposure exceeds that of the HN than those with the opposite orientation. Overall 38 ± 6% of the O's in the 305 proteins studied here have at least 10% of their surface exposed, while only 17 ± 6% of the HNs are this exposed. The preponderance of surface-exposed carbonyl oxygens is another reason why the interior of all proteins is at positive potential. This provides a mechanism for raising the potential at buried ligands that lack the interactions with neighboring amides that raise the potential at side chains.

How the positive potential from the backbone contributes to the free energy of ionized side chains in proteins

The free energy of interaction between side chains and the backbone

The potential is positive at the non-polar residues such as Val (average V_S is 163 mV), Ile (145 mV), and Leu (126 mV) (Table 5). Moving from a potential of 0 into a potential of 163 mV would stabilize a negative charge by −3.75 kcal/mol or destabilize a positive one by an equivalent amount. However, despite the significant potential (Ψ_I) these neutral, non-polar residues contribute little to the free energy of side chain interaction with the backbone (ΔG_{bkn}), because the net atomic partial charge (q_I) is near zero (Eq. 5). The large positive potential at non-polar residues supports the picture that forces other than favorable electrostatic interactions between side chain and amide dipoles are responsible for the predominately positive protein interior. However, the average of V_S at the acidic residues Asp and Glu is 45 and 24 mV, respectively, more positive than at

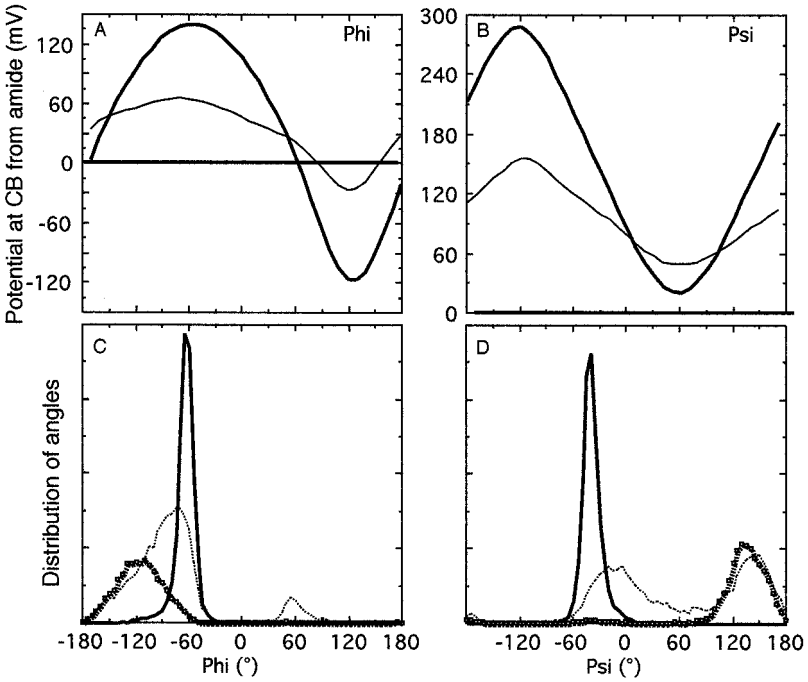

FIGURE 8 The potential at the middle CB in an Ala tripeptide from (*A*) amide(n) as a function of the phi angle and (*B*) amide(c) as a function of the psi angle (see Fig. 7 *A*). The potential was calculated with (*bold line*) $\epsilon_{peptide} = \epsilon_{solv} = 4$; (*flatter, light line*) $\epsilon_{peptide} = 4$, $\epsilon_{solv} = 80$. The relative occurrence of residues with different phi (*C*) and psi (*D*) angles in the 305 proteins considered in this study were determined with the program DSSP (Kabsch and Sander, 1983): Solid line, α-helix; heavy dotted line, β-sheet; light line, other.

their polar analogs Asn and Gln. The bases Arg and Lys do have the least positive average V_S. Thus, electrostatic interactions between backbone and side chains do contribute somewhat to the amide orientation that determines the potential.

V_S considers all side chain heavy atoms equally (Eq. 4a). In contrast, ΔG_{bkn} considers the partial charge on each atom and the potential (Eq. 5). ΔG_{bkn} is favorable at the basic residues despite the average side chain potential being positive. Thus, the atoms with positive charge must be in regions that are more negative than the average for the residue as a whole. In contrast, the average Glu V_S is 108 mV while the average ΔG_{bkn} is only -68 meV (-1.6 kcal/mol). Thus, the potential must be more positive at atoms that cannot add to the favorable ΔG_{bkn} because they have little charge.

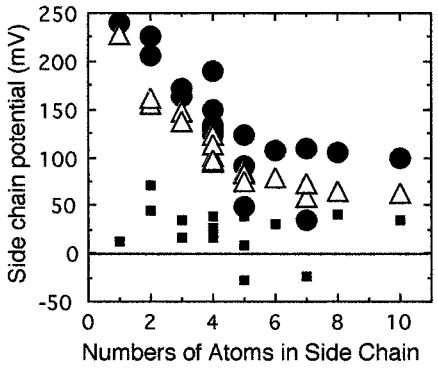

FIGURE 9 The dependence of the average potential at the side chain (V_S) on the length of the side chain. The average potential (V_S), ●; the contribution from the neighboring amides, △; the contribution of the distal amides, ■. Data from Table 8.

Loss of reaction field energy of ionized amino acids in proteins

The loss of reaction field energy (ΔG_{rxn}) (Eq. 2) provides a quantitative measure of the distribution of buried charges in proteins. The interactions with the potential created by the backbone will be most important for buried, charged residues. ΔG_{rxn} was calculated for the acids Asp and Glu, and bases Lys and Arg (Figs. 11 and 12; Table 9). Seventy percent have lost <4.1 kcal/mol of the reaction field energy they would have if free in water, shifting the residue pK_a by <3 pH units (Eq. 6). Thus, as expected, most of these ionizable residues are near the surface. However, 30% (5501) have $\Delta G_{rxn} > 4.1$ kcal/mol. Half of these have lost sufficient reaction field energy to shift their pK_a values by 5 pH units (6.8 kcal/mol). A 5 pH unit shift destabilizes an ionized Asp, moving its pK_a from 4 to 9. The same ΔG_{xn} shifts the pK_a of an Arg from 12.5 to 7.5. Burial in the protein can also be assessed by the exposure of the side chain to the surface. The fraction of residues that have lost >6.8 kcal/mol ΔG_{rxn} is comparable to the fraction of residues that have $<10\%$ of the side chain atoms with significant charge exposed to the solvent (Table 9).

Different propensities are found for burying each type of side chain. There are more buried Asp, similar numbers of buried Arg and Glu, and fewer buried Lys. Overall there are more buried acids than bases (Fig. 11, Table 9). This disparity becomes more significant as ΔG_{rxn} increases. For

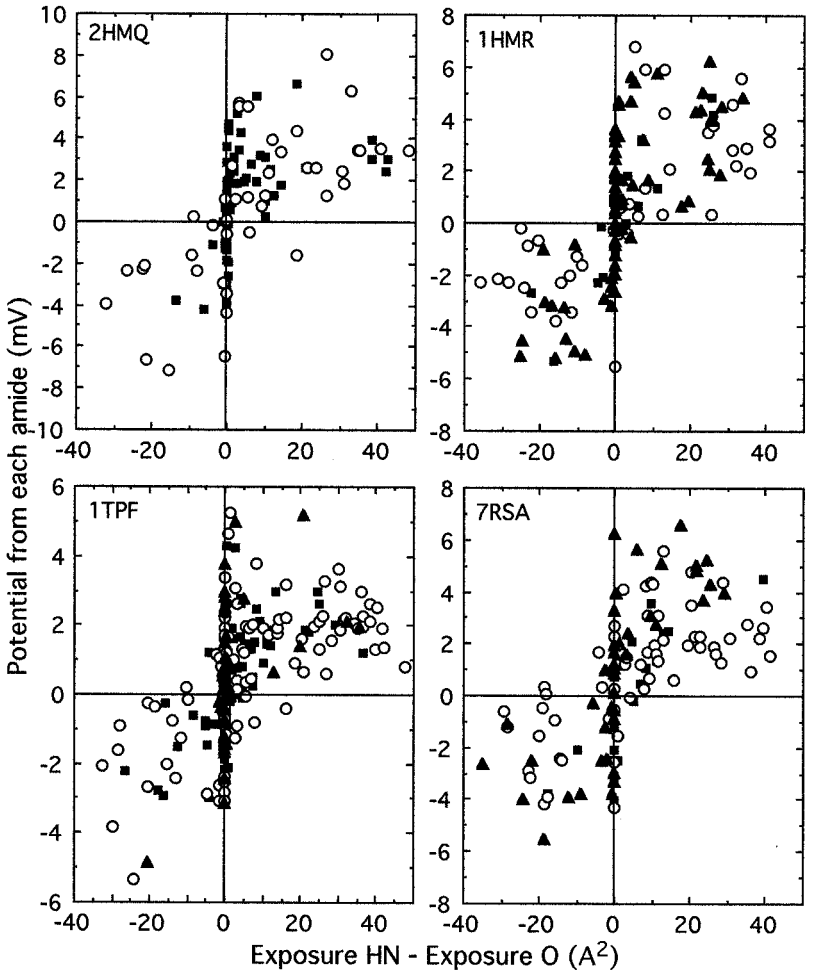

FIGURE 10 The difference in the exposure of the HN and O vs. the contribution of that amide to the average potential within the four-helix bundle 2HMQ, the β-barrel 1HMR, the α/β barrel 1TPF, and the $\alpha + \beta$ protein 7RSA. Residues in α-helices (■), in β-sheets (▲), and in loops (○). The structures were modified as described in the Methods section to equalize the length and size of the HN-N and C-O dipoles. The potential was calculated with $\epsilon_{protein} = 4$, $\epsilon_{solv} = 80$.

residues where ΔG_{rxn} is 4.1–6.8 kcal/mol, 56% are acids. Of the residues where ΔG_{rxn} is >6.8 kcal/mol 62% are acids, representing 17% of the acids and 12% of the bases.

Interaction of ionized residues with the backbone

A buried acid or base with a large ΔG_{rxn} will be neutral at physiological pH unless specific elements of the protein stabilize the charge (Eq. 6). Nearby charges or appropriately oriented dipoles can compensate for the loss of reaction field energy. The free energy of stabilization of each acidic and basic residue due to the electrostatic potential from the protein amide dipoles (ΔG_{bkn}) was calculated with Eq. 1 using CHARMM charges for the backbone (Table 1). Fig. 12 compares ΔG_{rxn} and ΔG_{bkn} for individual amino acids. No surface-exposed residue ($\Delta G_{rxn} \sim 0$) has a large ΔG_{bkn}. However, buried groups have a wide range of interactions with the backbone. The straight line of slope 1 in Fig. 12 shows where $-\Delta G_{bkn} = \Delta G_{rxn}$. If there were no other interactions (e.g., with the other protein side chains) the pK_a of groups along this line would be identical to that found in solution. There are a small number of residues where stabilization by the potential from the backbone dipoles is larger than the destabilization due to removal from the water dipoles (Fig. 12 and Table 10). In the absence of other interactions the protein would shift the pK_a of acids to lower and bases to higher pH values. Prior calculations have shown that hyper-stabilized residues can be functionally important. For example, in the photosynthetic reaction center a cluster of buried acids remain significantly ionized because they exist in a region where $-\Delta G_{bkn} > \Delta G_{rxn}$ (Lancaster et al., 1996).

There are fewer residues with large ΔG_{bkn} than large ΔG_{rxn} (Tables 9 and 10). Only 14% of the acidic or basic residues have ΔG_{bkn} larger than ±4.1 kcal/mol. The different types of side chains have the same order of propensities for large values of ΔG_{bkn} as for ΔG_{rxn} (Asp > Glu ≥ Arg > Lys). However, the difference between acids and bases is far more striking. For example, ΔG_{bkn} is −4.1 kcal/mol for 20% of the acids, while only 6.5% of the bases have interactions above this threshold. For most residues ΔG_{bkn} is favorable. However, 80% of the strong, favorable interactions with the backbone are to acids, only 20% to bases. Of the small number of residues with unfavorable ΔG_{bkn},

TABLE 8 The contribution of amides to the potential in the protein depends on the amide orientation relative to the protein surface

Protein Motif		PDB file 2HMQ α	1HMR β	1TPF α/β	7RSA $\alpha + \beta$
Sum of the potential (mV) from all residues with given amide orientation	$S_H > S_O$*	159	150	162	157
	$S_H = S_O$	−22	21	16	−12
	$S_H < S_O$	−46	−83	−51	−56
Number of residues with given amide orientation	$S_H > S_O$	61	55	97	59
	$S_H = S_O$	27	42	117	34
	$S_H < S_O$	24	32	33	30
Average potential (mV) from amides with given orientation	$S_H > S_O$	3.1	2.7	1.7	2.7
	$S_H = S_O$	−0.5	0.5	0.1	−0.3
	$S_H < S_O$	−3.1	−2.6	−1.5	−1.9

*$S_H > S_O$, the HN has at least 1 Å more surface area exposed than the O for this amide; $S_H = S_O$, the HN and O surface exposure differ by less than 1 Å; $S_H < S_O$, the O has at least 1 Å more surface area exposed than the NH.
The protein coordinate files were modified for the analysis of amide exposure. The amine HN—N bond was lengthened to 1.23 Å, equivalent to the C—O bond. The radius of both HN and O were taken as 1.6 Å. The potential was calculated in a standard, unmodified structure.

93% are bases (Figs. 11, 12). Thus, acids are more likely to be buried than bases and they are much more likely to be stabilized inside the protein by the potential from the amide dipoles. These distinctions are as expected if the potential from the protein backbone creates a bias to favor buried acids and raise the energy of buried bases.

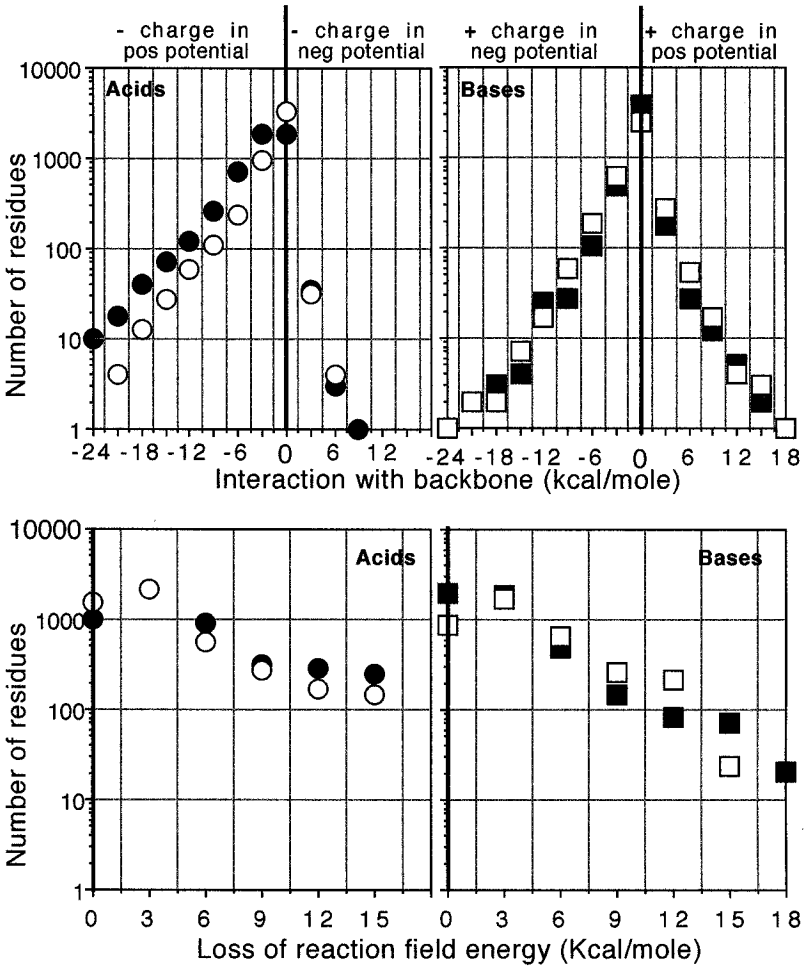

FIGURE 11 The distribution of acidic and basic side chains with different values of ΔG_{rxn} and ΔG_{bkn} in 305 proteins with different motifs. CHARMM charges were used for side chains and amides. The net charges in each run were +1 on the bases or −1 on the acids. $\epsilon_{protein} = 4$, $\epsilon_{solv} = 80$. Acids: ●, Asp; ○, Glu. Bases: ■, Arg; □, Lys.

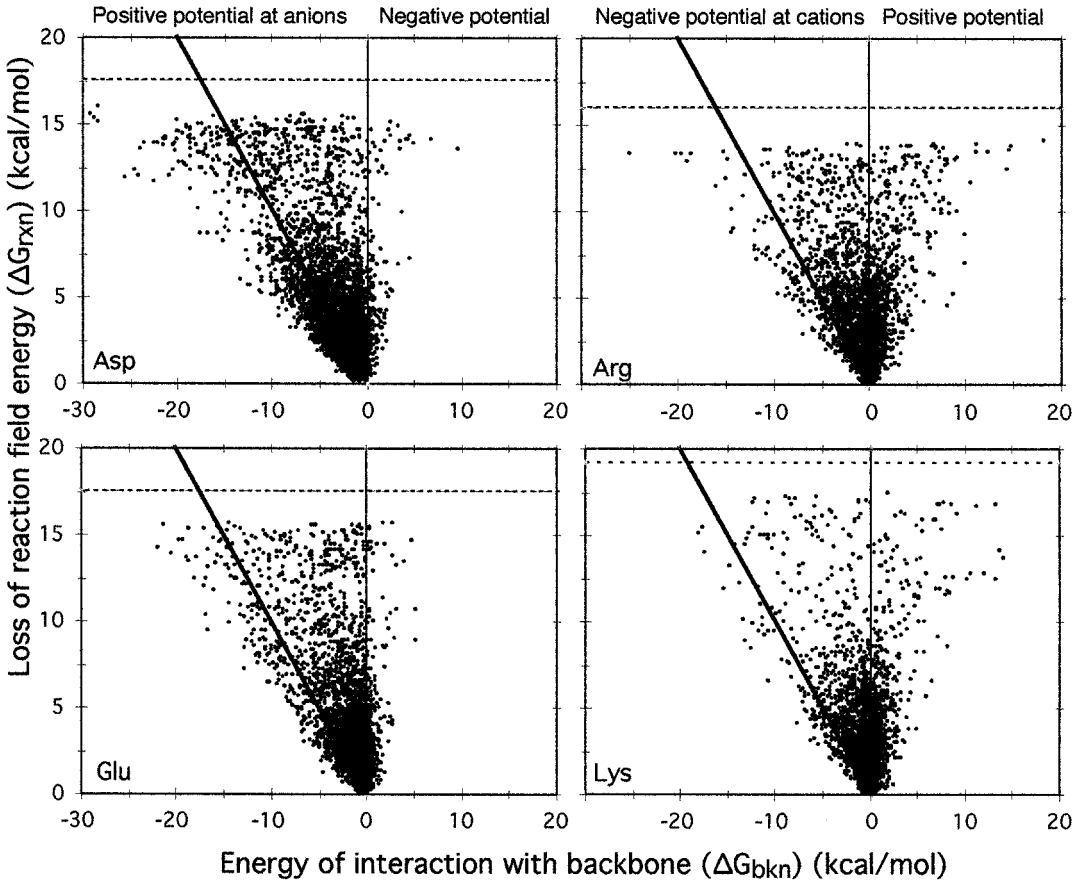

FIGURE 12 The relationship between ΔG_{bkn} and ΔG_{rxn} for the acidic and basic amino acids in 305 proteins. The bold line is for $-\Delta G_{bkn} = \Delta G_{rxn}$. The dashed line shows the maximum value for ΔG_{xn}, when $G_{rxn} = 0$ and $\Delta G_{rxn} = -G_{rxn\ in\ soln}$ (Table 2). The ± 1.5 kcal/mol has been removed.

The role of hydrogen bonds in creating favorable interactions between backbone and side chain

A hydrogen bond between the terminus of an acidic side chain and the amide HN or a basic side chain and the amide O generally indicates that the backbone will stabilize the charged residue. The necessity of hydrogen bonds for generating large values of ΔG_{bkn} was investigated (Table 11). Of the 1942 acids stabilized by >4.1 kcal/mol, 710 make no hydrogen bonds to the backbone. In contrast, of the 526 bases only 70 make no hydrogen bonds. This result highlights the bias toward the protein being positive inside. Thus, negative regions are almost always formed with local, hydrogen bonds while positive regions can be generated by longer-range interactions.

DISCUSSION

The average potential from the neutral amide dipoles (V_P) is found to be positive in every protein (Table 4, Fig. 4). Larger regions of each protein are at positive rather than negative potential (Fig. 3) and this potential is often large (Tables 5 and 6). The numerical value of the potential depends on the charge distribution used for the amide and the dielectric constant for the protein. However, the average remains positive even when these parameters are varied (Table 4). The potential from the backbone is positive within all proteins for two reasons. First, the side chains of all residues come off the backbone into the positive end of both their neighboring amides (Fig. 2 *A*). The regions of phi/psi space where side chains are close to the carbonyl oxygen are disallowed because of van der Waals overlap (Ramachandran and Sasisekharan, 1968). The HN proton is much smaller, so the side chain can come closer. In addition, the orientation of the amide at the protein surface influences the interior potential. The larger, more highly charged carbonyl O is more than twice as likely to be oriented into the solvent then the amine HN. The amides, with their O's more surface-exposed, raise the interior potential (Fig. 10). The restrictions in phi/psi space influence the interactions between amides and their neighboring side chains. The distribution of amide orientation at the protein surface raises the potential at distal side chains and bound ligands.

It is remarkable, given the complexity and uniqueness of individual proteins, that the neutral backbone yields a po-

TABLE 9 Loss of reaction field energy (ΔG_{rxn}) for ionized acids and bases within proteins

	Number of Residues within Given Range of ΔG_{rxn}				Percentage of Residues within Given Range of ΔG_{rxn}						
ΔpH	>5	3 to 5	0 to 3		>5	3 to 5	0 to 3				
meV	>300	180 to 300	0 to 180		>300	180 to 300	0 to 180		Average	Buried	Number of
Kcal/Mole	>6.8	4.1 to 6.8	0 to 4.1		>6.8	4.1 to 6.8	0 to 4.1		(meV)	(%)	Residues
					(%)	(%)	(%)				
Asp	986	976	2975	—	20.0	19.8	60.3	—	194	18	4937
Glu	670	622	3541	—	13.9	12.9	73.3	—	152	12	4833
Lys	389	556	3766	—	8.3	11.8	79.9	—	120	9	4711
Arg	600	702	2427	—	16.1	18.8	65.1	—	167	19	3729
Acids	1656	1598	6516	—	16.9	16.3	66.8	—	173	15	9770
Bases	989	1258	6193	—	12.2	15.3	72.5	—	141	13	8440
All	2645	2856	12709	—	14.5	15.8	69.6	—	158		18210

ΔG_{rxn} calculated with Eq. 2. The positive ΔpH units implies that the neutral form of the side chain will be stabilized, shifting the pK_a of acids to higher and bases to lower pH. Residues were considered buried if the terminal oxygens in Asp and Glu, or terminal nitrogens in Arg and Lys had less than 10% of their surface exposed to a solvent with a radius of 1.4 Å as determined by the program SURFV (Sridharan et al., 1992).

tential that is, on average, significantly positive in every protein. The question is how this bias affects protein structure and function. Empirical rules determined from the distribution of residues in protein structures have established the importance of other forces in proteins. Thus, the hydrophobic effect is recognized by many, though not all, non-polar residues being buried. Again, the solvation of charged residues stabilizes them on the surface where the majority are found (Table 9).

The analysis of the distribution of acidic and basic side chains reveals that despite the energetic penalty for removing charges from water, many are buried. However, there are significantly more buried acids than bases. This is as expected if the positive potential from the amides affects side chain location. There are 1.7 times as many acids that have lost 6.8 kcal/mol (5 ΔpH unit) reaction field energy in the proteins studied here (Table 9, Fig. 11). In addition, the proteins have more bound anions (phosphates, sulfates, heme propionic acids, etc) than cations (calcium, copper, zinc, etc.) (Table 6). The numerical value of the loss in reaction field energy is dependent on the parameters used for the protein dielectric constant and, to a lesser extent, the charge distribution on the side chains. However, similar results are found in studies that assess the surface exposure of side chains geometrically (Table 9). Prior surveys of residue surface accessibility in smaller numbers of proteins found three (Rashin and Honig, 1984) to almost nine (McDonald and Thorton, 1994) times as many buried acids as bases. The more modest imbalance of buried anions and cations reported here is probably more realistic. Although the positive potential from the backbone favors burial of anions, these will then repel each other. In addition, the buried anions will lower the electrostatic potential, as required to stabilize buried bases.

One challenge is to determine how the bias toward the backbone stabilizing anions is expressed within specific proteins. For example, the average potential from the backbone at a Val would stabilize an anion or destabilize a cation by 3.7 kcal/mol (160 meV) (Table 5). However, comparing the effects of mutating an arbitrary Val to an acidic or basic residue would not provide a simple test of this value. First and most important, the backbone is only one contributor to the electrostatic potential within a protein. Charged and polar side chains affect the energy of charges in the protein (Eq. 6) (Bashford and Karplus, 1990; Yang et al., 1993; Antosiewicz et al., 1996; Alexov and Gunner, 1997), but the analysis of the intra-side chain interactions is beyond the scope of this paper. In addition, the naturally occurring acids and bases have different structures, so they occupy different positions relative to the backbone. The neighboring dipoles (Fig. 2) affect Asp or Glu much more than the longer Arg or Lys (Table 7, Fig. 9) (Spassov et al., 1997). Also, the range of potentials from the distal amides is significant, so each position for mutation must be evaluated independently (Table 7). Lastly, while the amide dipoles are expected to have a favorable interaction with an acid, this is only rarely sufficient to be as large as the destabilization of the charge due to the loss of reaction field energy (Fig. 12). Therefore, without interactions with other side chains or ligands, buried acids would often be neutral. Thus, as found experimentally, random mutations that bury charges can destabilize a protein (Dao-pin et al., 1991) or yield neutral side chains (Stites et al., 1991). However, sometimes the residue will remain charged (Varadarajan et al., 1989; Perona et al., 1993). The results presented here suggest that in a protein with few other buried charges, acids will be less destabilizing and more likely to remain ionized than bases.

Thus, despite the penalty for moving charges into proteins, a significant number of acidic and basic side chains are buried (Table 9). As has been suggested previously, many of these buried residues have their charged state stabilized by the electrostatic potential from the amide back-

TABLE 10 The interaction of ionized acidic and basic side chains with the backbone (ΔG_{bkn})

	Number of Residues Falling in Range of Interaction Energies				Percentage of Residues Falling in Range of Interaction Energies					
ΔpH	<−5	−5 to −3	−3 to 0	0 to 3	<−5	−5 to −3	−3 to 0	0 to 3		
meV	<−300	−300 to −180	−180 to 0	0 to 180	<−300	−300 to −180	−180 to 0	0 to 180	Average	
Kcal/Mole	<−6.8	−6.8 to −4.1	−4.1 to 0	0 to 4.1	<−6.8	−6.8 to −4.1	−4.1 to 0	0 to 4.1	(meV)	*
					(%)	(%)	(%)	(%)		
Asp	617	800	3513	7	12.5	16.2	71.2	0.1	−145	12.2
Glu	247	278	4304	4	5.1	5.8	89.1	0.1	−68	3.1
Lys	74	122	4463	52	1.6	2.6	94.7	1.1	−15	1.0
Arg	134	198	3304	93	3.6	5.3	88.6	2.5	−27	2.1
Acids	864	1078	7817	11	8.8	11.0	80.0	0.1	−107	
Bases	208	320	7767	145	2.6	3.9	91.7	1.8	−20	
All	1072	1398	15584	156	5.7	7.5	85.9	1.0	−67	

ΔG_{bkn} calculated with equation 1.
*The percentage of residues where the backbone stabilizes the ionized side chain by 1 pH unit more than the loss of reaction field energy destablizes ionization.

bone (e.g., Hol et al., 1981; Gandini et al., 1996; Lancaster et al., 1996; Oberoi et al., 1996; Raychaudhuri et al., 1997; Spassov et al., 1997)). However, while the backbone destabilizes the ionization of few acids and bases in native proteins, it stabilizes many more acids than bases (Table 10, Fig. 12).

Studies in model systems

The positive potential from the backbones is not a result of proteins folding around buried anions. Because of the restrictions on angles at which side chains come off the backbone, the potential from the neighboring amides will tend to be positive in polypeptides as well as in proteins, although it will be diminished when the system is more solvent-exposed (Figs. 7 and 8).

The neighboring amides and shifts of amino acid pK_a values in peptides

The neighboring amides are predicted to stabilize the charge on the short Asp^- and Glu^-. The pK_a values of carboxylic acids are near 4.8, a value which is essentially independent of the length of the acid alkyl chain. In a tetrapeptide in solution Asp has a pK_a of 3.9, demonstrating the peptide backbone stabilizes the charge by 0.9 pH units (1.2 kcal/mol) (Richarz and Wüthrich, 1975). In the same study the pK_a of Glu is shifted by a smaller amount to 4.2. A survey of the measured pK_a values provided average values of 2.7 ± 0.9 and 4.0 ± 0.9 for Asp and Glu, respectively (Antosiewicz et al., 1996). This represents a −2.9 and −1.0 kcal/mol stabilization of Asp^- and Glu^- relative to a carboxylic acid in water. While the pK_a values of residues in proteins depend on a number of factors (Bashford and Karplus, 1991; Yang et al., 1993; Antosiewicz et al., 1994; Alexov and Gunner, 1997), the average backbone interaction with Asp^- and Glu^- of −3.4 and −1.6 kcal/mol are comparable to these shifts. The average interaction of His with the backbone would be expected to raise its pK_a by 0.6 pH units (Table 5). This is within experimental error of the finding that the average pK_a of His (6.9 ± 1.1) (Antosiewicz et al., 1996) is the same as that of imidazole.

TABLE 11 Acids or bases that are stabilized by the backbone by more than 4.1 kcal/mole (3 ΔpH unit) without making hydrogen bonds to the backbone

	*	†	%
Asp	1417	530	37
Glu	525	180	34
Arg	332	31	9
Lys	196	39	20

*Number of residues with ΔG_{bkn} <−4.1 kcal/mol.
†Number of residues with ΔG_{bkn} <−4.1 kcal/mol that make no hydrogen bonds with the backbone amides.
Two groups were defined as making a hydrogen bond if the H to O distance is 3 Å or less. No angular cutoffs were used.

The contribution of the neighboring amides to the helix propensity of the ionizable amino acids

The different helix propensities of amino acids have been recognized to result from a number of factors, including the loss of entropy and the burial of side chains when a helix is formed (Creamer and Rose, 1994; Pace and Scholtz, 1998). In addition, ionized residues interact with the charge of the helix macro-dipole in proteins and polypeptides (Hol, 1985; Shoemaker et al., 1987; Aqvist et al., 1991; Sitkoff et al., 1994; Nicholson et al., 1988; Sali et al., 1988). An anion is stabilized near the helix N-terminal and induces helix fraying near the C-terminal. A cation has the opposite effect.

However, the local interaction between the side chain and neighboring amides also depends on the phi and psi angles (Figs. 7 and 8). The interaction, especially with amide(n), is strongest when a residue is in an α-helix (Fig. 8). Thus, the neighboring amide interactions should modify helix propensity even at the helix midpoint, where the macro-dipole influence is negligible.

Baldwin and colleagues have compared the helix propensities of side chains incorporated at different positions along the helix to correct for the influence of the helix macro-dipole (Chakrabartty et al., 1994). Asp^- and Asp^0 have similar helix propensities (Huyghues-Despointes et al., 1993) and Glu^- is somewhat helix-destabilizing relative to Glu^0 (Scholtz et al., 1993), while His^+ is very helix destabilizing relative to His^0 (Armstrong and Baldwin, 1993). A series of amino acid analogs provides more evidence that a positive charge near the amide destabilizes a helix. In particular, the short side chain ($CH_2NH_3^+$) is much more helix-destabilizing than the analog with one more carbon (Padmanabhan et al., 1996). Surveys of the frequency of side chain positions do show that bases are prevalent at the middle of helices (Richardson and Richardson, 1988; Gandini et al., 1996). However, long tails of Arg and Lys distance their charge from their neighboring amides.

How the positive backbone potential can influence protein folding and stability

Protein stability is sensitive to pH and salt concentration. Thus, electrostatic forces influence the equilibrium between folded and unfolded states (Stigter et al., 1991). The constraints that lead to positive potential from the backbone may also be found in compact, non-native structures often found on folding pathways (Fink, 1995). The average potential from the backbone within a group of incorrectly folded proteins (Novotny et al., 1984) is similar to the average of the ensemble of proteins studied here (data not shown). Low-pK_a acids have been predicted in non-native, compact states of apomyoglobin (Yang and Honig, 1994). Measured carboxylate pK_a values in compact, unfolded proteins are on average 0.3 to 0.4 pH units lower than found in isolated residues (Oliveberg et al., 1995; Tan et al., 1995).

The addition of salts can also change the relative stability of the native and other states of proteins. Ion occupancy of specific cation or anion binding sites do stabilize native conformations (Pace and Grimsley, 1988). However, salts can also lead to protein unfolding. The compact, unfolded states appear to be stabilized by anion binding (Goto et al., 1990; Uversky et al., 1998). The preferential interaction of backbone dipoles with anions may provide one mechanism for the observed anion dependence of salt-induced denaturation.

Could the positive bias of the amide backbone have influenced the chemical nature of substrates or selection of amino acids found in proteins?

The acidic and basic amino acids in modern proteins have significantly different structures. The bases Lys and Arg are very long, so they have little interaction with their neighboring dipoles, and His has a pK_a near physiological pH, so it can be neutral without destabilizing the protein. In contrast, Asp and Glu are short, so their charge can be stabilized by their neighboring dipoles. It is tempting to speculate that the positive potential from the backbone amides may have had an impact on the selection of amino acids that are incorporated in proteins, as has been considered previously by Spassov (Spassov et al., 1997). Shorter analogs of Lys such as ornithine, diaminopropionic acid, and diamminobutryic acids are present in mixtures that may represent pre-biotic (bio)chemistry (Rohlfing and Saunders, 1978) and are also found in modern metabolism, but are not incorporated into proteins. Longer-chain acidic amino acids such as α-amino adipic acid are intermediates in metabolic pathways, but are not found in proteins.

Metals are the most common cations associated with proteins. Metal binding sites are generally at positive potential from the backbone, with amino acids such as Cys and His playing essential roles in binding. There are several important exceptions to this rule. Calcium is often bound by backbone carbonyls and is therefore at very negative potential from the backbone. Sodium is also found at negative potentials (Table 6). The potassium channel protein uses a ring of carbonyls pointing into the channel to bind the correct cation (Doyle et al., 1998). This motif actually uses the propensity of carbonyls to point out toward the protein surface to form the cation binding site.

While the backbone often destabilizes cation binding, it contributes significantly to the binding of many anionic substrates and cofactors (Table 6) (Quiocho et al., 1987; Jacobson and Quiocho, 1988; Luecke and Quiocho, 1990; Wilson et al., 1992; He and Quiocho, 1993; Yao et al., 1996). Anions such as carboxylic acids and phosphates play crucial roles in cellular metabolism. Phosphorylated substrates are used in polymerization of proteins, nucleic acids, and polysaccharides. Proteins interact with DNA or RNA in all stages of nucleic acid replication, transcription, and translation. Enzymes are phosphorylated to control their activity. Many cofactors have phosphates in their structure, which are not needed for catalysis, but are still removed from water and bound in the protein. The energy of interaction of the backbone with these phosphates can be sufficient to bind the ligand with little help from amino acid side chains (Yao et al., 1996). Thus, the amide group is found to be specifically well designed to bind the phosphate containing molecules that are the frequent partners of proteins in much of biochemistry.

We thank Robert Callender, Themis Lazaridis, Barry Honig, and Colin Wraight for patient, helpful discussions.

This work was supported by NSF-MCB Grant 9629047 and National Institutes of Health Grant GM08168 (to E.C.) and City College CRS (to M.A.S.).

REFERENCES

Alexov, E. G., and M. R. Gunner. 1997. Incorporating protein conformational flexibility into the calculation of pH-dependent protein properties. *Biophys. J.* 72:2075–2093.

Antosiewicz, J., J. A. McCammon, and M. K. Gilson. 1994. Prediction of pH-dependent properties in proteins. *J. Mol. Biol.* 238:415–436.

Antosiewicz, J., J. A. McCammon, and M. K. Gilson. 1996. The determinants of pK_a's in proteins. *Biochemistry.* 35:7819.

Aqvist, J., H. Luecke, F. A. Quiocho, and A. Warshel. 1991. Dipoles localized at helix termini of proteins stabilize charges. *Proc. Natl. Acad. Sci. USA.* 88:2026–2030.

Armstrong, K. M., and R. L. Baldwin. 1993. Charged histidine affects α-helix stability at all positions in the helix by interacting with the backbone charges. *Proc. Natl. Acad. Sci. USA.* 90:11337–11340.

Baker, E. N., and R. E. Hubbard. 1984. Hydrogen bonding in globular proteins. *Prog. Biophys. Mol. Biol.* 44:97–179.

Bashford, D., and M. Karplus. 1990. The pK_a's of ionizable groups in proteins: atomic detail from a continuum electrostatic model. *Biochemistry.* 29:10219–10225.

Bashford, D., and M. Karplus. 1991. Multiple-site titration curves of proteins: an analysis of exact and approximate methods for their calculation. *J. Phys. Chem.* 95:9556–9561.

Bernstein, F. C., T. F. Koetzle, G. J. B. Williams, E. F. Meyer, M. D. Brice, J. R. Rodgers, O. Kennard, T. F. Shimanouchi, and M. Tasumi. 1977. The protein data bank: a computer based archival file for macromolecular structures. *J. Mol. Biol.* 112:535–542.

Beroza, P., D. R. Fredkin, M. Y. Okamura, and G. Feher. 1995. Electrostatic calculations of amino acid titration electron transfer, $Q_A^- Q_B \rightarrow Q_A Q_B^-$, in the reaction center. *Biophys. J.* 68:2233–2250.

Brooks, B. R., R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus. 1983. CHARMM: a program for macromolecular energy, minimization and dynamics calculations. *J. Comp. Chem.* 4:187–217.

Chakrabartty, A., T. Kortemme, and R. L. Baldwin. 1994. Helix propensities of the amino acids measured in alanine-based peptides without helix-stabilizing side-chain interactions. *Protein Sci.* 3:843–852.

Churg, A. K., and A. Warshel. 1986. Control of the redox potential of cytochrome c and microscopic dielectric effects in proteins. *Biochemistry.* 25:1675–1681.

Creamer, T. P., and G. D. Rose. 1994. α-Helix forming propensities in peptides and proteins. *Proteins.* 19:85–97.

Dao-pin, S., D. E. Anderson, W. A. Baase, F. W. Dahluist, and B. W. Matthews. 1991. Structural and thermodynamic consequences for burying a charged residue within the hydrophobic core of T4 lysozyme. *Biochemistry.* 30:11521–11529.

Doyle, D. A., J. M. Cabral, R. A. Pfuetzner, A. Kuo, J. M. Gulbis, S. L. Cohen, B. T. Chait, and R. MacKinnon. 1998. The structure of the potassium channel: molecular basis of K^+ conduction and selectivity. *Science.* 280:69–77.

Fink, A. L. 1995. Compact intermediate states in protein folding. *Annu. Rev. Biophys. Biomol. Struct.* 24:495–522.

Gandini, D., L. Gogioso, M. Bolognesi, and D. Bordo. 1996. Patterns in ionizable side chain interactions in protein structures. *Proteins: Struct., Funct., Genet.* 24:439–449.

Gilson, M. K., K. A. Sharp, and B. H. Honig. 1987. Calculating the electrostatic potential of molecules in solution: method and error assessment. *J. Comp. Chem.* 9:327–335.

Goto, Y., N. Takahashi, and A. L. Fink. 1990. Mechanism of acid-induced folding of proteins. *Biochemistry.* 29:3480–3488.

Gunner, M. R., E. Alexov, E. Torres, and S. Lipovaca. 1997. The importance of the protein in controlling the electrochemistry of heme metalloproteins: methods of calculation and analysis. *J.B.I.C.* 2:126–134.

Gunner, M. R., and B. Honig. 1991. Electrostatic control of midpoint potentials in the cytochrome subunit of the *Rhodopseudomonas viridis* reaction center. *Proc. Natl. Acad. Sci. USA.* 88:9151–9155.

He, J. J., and F. A. Quiocho. 1993. Dominant role of local dipoles in stabilizing uncompensated charges on a sulfate sequestered in a periplasmic active transport protein. *Protein Sci.* 2:1643–1647.

Hol, W. G. J. 1985. The role of the α-helix dipole in protein function and structure. *Prog. Biophys. Mol. Biol.* 45:149–195.

Hol, W. G., L. M. Halie, and C. Sander. 1981. Dipoles of the α-helix and β-sheet: their role in protein folding. *Nature.* 294:532–536.

Hol, W. G. J., P. T. van Duijnen, and H. J. C. Berendsen. 1978. The α-helix dipole and the properties of proteins. *Nature.* 273:443–446.

Holmes, M. A., and R. E. Stenkamp. 1991. The structures of met and azidomet hemerythrin at 1.66 Å resolution. *J. Mol. Biol.* 220:723–737.

Huyghues-Despointes, B. M. P., J. M. Scholtz, and R. L. Baldwin. 1993. Effect of a single aspartate on helix stability at different positions in a neutral alanine-based peptide. *Protein Sci.* 2:1604–1611.

Jacobson, B. L., and F. A. Quiocho. 1988. Sulfate-binding protein dislikes protonated oxyacids. A molecular explanation. *J. Mol. Biol.* 24:783–787.

James, M., A. Sielecki, G. Brayer, L. Delbaere, and C. Bauer. 1980. Structures of product and inhibitor complexes of *Streptomyces griseus* protease A at 1.8 Å resolution. A model for serine protease catalysis. *J. Mol. Biol.* 144:43–88.

Kabsch, W., and C. Sander. 1983. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers.* 22:2577–2637.

Kishan, K. V. R., J. P. Zeelen, M. E. M. Noble, T. V. Borchert, and R. K. Wierenga. 1994. Comparison of the structures and the crystal contacts of trypanosomal triosephosphate isomerase in four different crystal forms. *Protein Sci.* 3:779–787.

Lancaster, C. R. D., H. Michel, B. Honig, and M. R. Gunner. 1996. Calculated coupling of electron and proton transfer in the photosynthetic reaction center of *Rhodopseudomonas viridis*. *Biophys. J.* 70: 2469–2492.

Langen, R., G. M. Jensen, U. Jacob, P. J. Stephens, and A. Warshel. 1992. Protein control of iron-sulfur cluster redox potentials. *J. Biol. Chem.* 267:25625–25627.

Luecke, H., and F. A. Quiocho. 1990. High specificity of a phosphate transport protein determined by hydrogen bonds. *Nature.* 347:402–406.

McDonald, I. K., and J. M. Thorton. 1994. Satisfying hydrogen bonding potential in proteins. *J. Mol. Biol.* 238:777–793.

McPhalen, C. A., N. C. J. Strynadka, and M. N. G. James. 1991. Calcium binding sites in proteins: a structural perspective. *Adv. Protein Chem.* 42:77–144.

Murzin, A. G., S. E. Brenner, T. Hubbard, and C. Chothia. 1995. SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.* 247:536–540.

Nicholls, A., and B. Honig. 1991. A rapid finite difference algorithm utilizing successive over-relaxation to solve the Poisson-Boltzmann equation. *J. Comp. Chem.* 12:435–445.

Nicholls, A., K. Sharp, and B. Honig. 1991. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins.* 11:281–296.

Nicholson, H., W. J. Becktel, and B. W. Matthews. 1988. Enhanced protein thermostability from designed mutations that interact with α-helix dipoles. *Nature.* 336:651–656.

Novotny, J., R. Bruccoleri, and M. K. Karplus. 1984. An analysis of incorrectly folded protein models: implications for structure predictions. *J. Mol. Biol.* 177:787–818.

Oberoi, H., J. Trikha, X. Yuan, and N. Allewell. 1996. Identification and analysis of long-range electrostatic effects in proteins by computer modeling: aspartate transcarbamylase. *Proteins.* 25:300–314.

Oliveberg, M., V. L. Arcus, and A. R. Fersht. 1995. pK_a values of carboxyl groups in the native and denatured states of barnase. The pK_a values of

the denatured state are on average 0.4 units lower than those of model compounds. *Biochemistry.* 34:9424–9433.

Pace, C. N., and G. R. Grimsley. 1988. Ribonuclease T1 is stabilized by cation and anion binding. *Biochemistry.* 27:3242–3246.

Pace, C. N., and J. M. Scholtz. 1998. A helix propensity scale based on experimental studies of peptides and proteins. *Biophys. J.* 75:422–427.

Padmanabhan, S., E. J. York, J. M. Stewart, and R. L. Baldwin. 1996. Helix propensities of basic amino acids increase with the length of the side-chain. *J. Mol. Biol.* 257:726–734.

Perona, J. J., C. A. Tsu, M. E. McGrath, C. S. Craik, and R. J. Fletterick. 1993. Relocating a negative charge in the binding pocket of trypsin. *J. Mol. Biol.* 230:934–949.

Quiocho, F. A., J. S. Sack, and N. K. Vyas. 1987. Stabilization of charges on isolated ionic groups sequestered in proteins by polarized peptide units. *Nature.* 329:561–564.

Ramachandran, G. N., and V. Sasisekharan. 1968. Conformation of polypeptides and proteins. *Adv. Protein Chem.* 23:283–437.

Ramachandran, G. N., A. S. Kolaskar, C. Ramakrishnan, and V. Sasisekharan. 1974. The mean geometry of the peptide unit from crystal structure data. *Biochim. Biophys. Acta.* 359:298–302.

Rashin, A. A., and B. Honig. 1984. On the environment of ionizable groups in globular proteins. *J. Mol. Biol.* 173:515–521.

Raychaudhuri, S., F. Younas, P. A. Karplus, C. H. Faerman, and D. R. Ripoll. 1997. Backbone makes a significant contribution to the electrostatics of α/â-barrel proteins. *Protein Sci.* 6:1849–1857.

Richardson, J. S., and D. C. Richardson. 1988. Amino acid preferences for specific locations at the ends of α-helices. *Science.* 240:1648–1652.

Richarz, R., and K. Wüthrich. 1975. Carbon-13 NMR chemical shifts of the common amino acid residues measured in aqueous solutions of the linear tetrapeptides H-Gly-Gly-X-L-Ala-OH. *Biopolymers.* 17: 2133–2141.

Rohlfing, D. L., and M. A. Saunders. 1978. Evolutionary processes possibly limiting the kinds of amino acids in proteins to twenty: a review. *J. Theor. Biol.* 71:487–503.

Sali, D., M. Bycroft, and A. R. Fersht. 1988. Stabilization of protein structure by interaction of alpha-helix dipole with a charged side chain. *Nature.* 335:740–743.

Sancho, J., L. Serrano, and A. R. Fersht. 1992. Histidine residues at the N- and C-terminal of α-helices: perturbed pK_a's and protein stability. *Biochemistry.* 31:2253–2258.

Scholtz, J. M., H. Qian, V. H. Robbins, and R. I. Baldwin. 1993. The energetics of ion-pair and hydrogen-bonding interactions in a helical peptide. *Biochemistry.* 32:9668–9676.

Shoemaker, K. R., P. S. Kim, E. J. York, J. M. Stewart, and R. L. Baldwin. 1987. Tests of the helix dipole model for stabilization of α-helices. *Nature.* 326:563–567.

Sitkoff, D., D. J. Lockhart, K. Sharp, and B. Honig. 1994. Calculation of electrostatic effects at the amino terminus of an α-helix. *Biophys. J.* 67:2251–2260.

Spassov, V. Z., R. Ladenstein, and A. D. Karshikoff. 1997. Optimization of the electrostatic interactions between ionized groups and peptide dipoles in proteins. *Protein. Sci.* 6:1190–1196.

Sridharan, S., A. Nicholls, and B. Honig. 1992. A new vertex algorithm to calculate solvent accessible surface areas. *Biophys. J.* 61:174*a*. (Abstr.).

Stickle, D. F., L. G. Presta, K. A. Dill, and G. D. Rose. 1992. Hydrogen bonding in globular proteins. *J. Mol. Biol.* 226:1143–1163.

Stigter, D., D. O. V. Alonso, and K. A. Dill. 1991. Protein stability: electrostatics and compact denatured states. *Proc. Natl. Acad. Sci. USA.* 88:4176–4180.

Stites, W. E., A. G. Gittis, E. E. Lattman, and D. Shortle. 1991. In a staphylococcal nuclease mutant the side-chain of a lysine replacing valine 66 is fully buried in the hydrophobic core. *J. Mol. Biol.* 221:7–14.

Strydnaka, N. C. J., and M. N. E. James. 1989. Crystal structures of the helix-loop-helix calcium-binding proteins. *Annu. Rev. Biochem.* 58: 951–998.

Swartz, P. D., B. W. Beck, and T. Ichiye. 1996. Structural origins of redox potentials in Fe-S proteins: electrostatic potentials of crystal structures. *Biophys. J.* 71:2958–2969.

Tan, Y., M. Oliveberg, B. Davis, and A. R. Fersht. 1995. Perturbed pK_a-values in the denatured states of proteins. *J. Mol. Biol.* 254: 980–992.

Uversky, V. N., A. S. Karnoup, D. J. Segel, S. Seshadri, S. Doniach, and A. L. Fink. 1998. Anion-induced folding of Staphylococcal nuclease: characterization of multiple equilibrium partially folded intermediates. *J. Mol. Biol.* 278:879–894.

Varadarajan, R., D. G. Lambright, and S. G. Boxer. 1989. Electrostatic interactions in wild-type and mutant recombinant human myoglobins. *Biochemistry.* 28:3771–3781.

Wada, A. 1976. The α-helix as an electric macro-dipole. *Adv. Biophys.* 9:1–63.

Wilson, D. K., K. M. Bohren, K. H. Gabbay, and F. A. Quiocho. 1992. An unlikely sugar substrate site in the 1.65 Å structure of the human aldose reductase holoenzyme implicated in diabetic complications. *Science.* 257:81–84.

Wlodawer, A., L. A. Svensson, L. Sjolin, and G. l. Gilliland. 1988. Structure of phosphate-free ribonuclease A refined at 1.26 Ångstroms. *Biochemistry.* 27:2705.

Yang, A.-S., M. R. Gunner, R. Sampogna, K. Sharp, and B. Honig. 1993. On the calculation of pK_a's in proteins. *Proteins.* 15:252–265.

Yang, A.-S., and B. Honig. 1994. Structural origins of pH and ionic strength effects on protein stability acid denaturation of sperm whale apomyoglobin. *J. Mol. Biol.* 273:602–614.

Yang, A., and B. Honig. 1995a. Free energy determinants of secondary structure formation. I. α-Helices. *J. Mol. Biol.* 252:351–365.

Yang, A., and B. Honig. 1995b. Free energy determinants of secondary structure formation. II. Antiparallel β-sheet. *J. Mol. Biol.* 252:366–376.

Yao, N., P. S. Ledvina, A. Choudhary, and F. A. Quiocho. 1996. Modulation of a salt link does not affect binding of phosphate to its specific active transport receptor. *Biochemistry.* 35:2079–2085.

Zanotti, G., G. Scapin, P. Spadon, J. H. Veerkamp, and J. C. Sacchettini. 1992. Three-dimensional structure of recombinant human muscle fatty acid-binding protein. *J. Biol. Chem.* 267:18541–18550.