doi:10.1016/j.jmb.2005.03.051 J. Mol. Biol. (2005) 348, 1283–1298

JMB

Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

Are Acidic and Basic Groups in Buried Proteins Predicted to be Ionized?

Jinrang Kim, Junjun Mao and M. R. Gunner*

Physics Department J-419 City College of New York 138th Street and Convent Avenue, New York, NY 10031 USA

*Corresponding author

Ionizable residues play essential roles in proteins, modulating protein stability, fold and function. Asp, Glu, Arg, and Lys make up about a quarter of the residues in an average protein. Multi-conformation continuum electrostatic (MCCE) calculations were used to predict the ionization states of all acidic and basic residues in 490 proteins. Of all 36,192 ionizable residues, 93.5% were predicted to be ionized. Thirtyfive percent have lost 4.08 kcal/mol solvation energy ($\Delta\Delta G_{rxn}$) sufficient to shift a pK_a by three pH units in the absence of other interactions and 17% have $\Delta\Delta G_{rxn}$ sufficient to shift pK_a by five pH units. Overall 85% of these buried residues ($\Delta\Delta G_{rxn} > 5\Delta pK$ units) are ionized, including 92% of the Arg, 86% of the Asp, 77% of the Glu, and 75% of the Lys. Ion-pair interactions stabilize the ionization of both acids and bases. The backbone dipoles stabilize anions more than cations. The interactions with polar side-chains are also different for acids and bases. Asn and Gln stabilize all charges, Ser and Thr stabilize only acids while Tyr rarely stabilize Lys. Thus, hydroxyls are better hydrogen bond donors than acceptors. Buried ionized residues are more likely to be conserved than those on the surface. There are 3.95 residues buried per 100 residues in an average protein.

Keywords: pK_a; continuum electrostatics; multi-conformation; buried charges; structural genomics

Introduction

Ionizable residues play essential roles in proteins. Asp, Glu, Arg, and Lys make up about a quarter of the residues in an average protein. They account for almost half of active site residues while about another fifth are His[1]. Charged groups in proteins are essential for protein–protein recognition,[2,3] for modulating electrostatic fields at protein active sites,[4] and for forming proton conduction pathways.[5,6] The strongest long-range interactions in proteins are the electrostatic interactions especially between buried charged groups. The pH and salt dependence of protein stability shows the importance of residue ionization states and screening of electrostatic interactions in protein folding.[7] Residue pK_as have been measured by NMR[8] or FTIR,[9–12] by the difference in the pH dependence of denaturation energies,[13,14] by potentiometric titration,[15] and by following the pH dependence of protein activity. Computational analysis has also developed to a point where reasonably accurate predictions of residue pK_as can be made.[7,16–20] Calculation allows for a systematic computational study of a large number of ionizable groups to determine their distribution, ionization state, and factors that control ionization. Large-scale comparisons of this kind cannot readily be carried out experimentally.

The interactions of residues with water and with each other determine a protein's folded structure. Hydrophobic residues prefer to be out of contact with water while polar and charged groups are stabilized by interactions with water, imposing an energy barrier for burial.[21,22] Thus, proteins have most charged groups and many polar residues on their surface with the majority of hydrophobic residues buried inside. The charged surface residues keep the protein soluble and are used for the recognition of binding partners, such as substrates, other proteins, or the lipid membrane. The earliest surveys suggested that buried ionizable residues were very rare.[23,24] However, only a few small proteins with little internal volume were analyzed.

Abbreviations used: MCCE, multi-conformation continuum electrostatic; ASA, accessable surface area; QM-MM, quantum mechanics–molecular mechanics; CE, continuum electrostatics.

E-mail address of the corresponding author: gunner@sci.ccny.cuny.edu

More recent surveys of the solvent exposure of ionizable groups,[25] ion-pairs[26] or the desolvation energy of these residues[27] show that more than 30% of ionizable residues are buried. Seventy percent of functional ionizable residues have less than 5% solvent exposure when substrate is bound.[1] However, it is not generally known if these buried residues are ionized.

Charges and dipoles have the most solvation energy and conformational freedom when surrounded by water. However, attractive pair-wise interactions are weak in water because they are screened by the high dielectric solvent keeping charged groups apart. Measurements of polypeptides show salt bridges are not stable in water; however, they are stable in solvents such as hydrated octanol.[28] A backbone amide group or a polar side-chain loses solvation energy when removed from water.[29–31] Partition coefficients for oil–water transfer show large favorable water–ion interactions are lost when a charged group is moved out of water.[32,33] The loss of solvation energy shifts the pK_as of acids and bases stabilizing the neutral form. Thus, groups can only remain ionized when buried if the interactions with water are replaced by stabilizing interactions with dipolar or charged groups.[34] When an ion pair is brought into an environment with a lower dielectric constant, pair-wise interactions are stronger. However, early analysis showed the loss of solvation energy could barely be repaid by the favorable interaction within an ion-pair.[35] Thus, salt bridges must be properly arranged to compensate for the loss of reaction field energy.

Protein interiors are by many measures a polar medium. Every residue has an amide group with a larger dipole moment than water. Polar residues make up 25% and ionizable residues 25% of an average protein. The concentration of polar moieties inside proteins has been estimated as being on the order of 25 M.[28] Hence, proteins are highly polar but not polarizable, differing from water in the relative rigidity of the polar groups, which cannot reorient to stabilize changes in charge state. Thus, protein interiors will not tolerate randomly introduced charges but they can stabilize particular charges in particular locations. Many experimental and computational studies have shown that specific charged residues can stabilize[26,36–38] or destabilize[39–43] proteins depending on their context.[40,44,45]

Methods have been developed to calculate pK_as of residues in proteins. Quantum mechanics (QM),[46] quantum mechanics–molecular mechanics (QM-MM),[47,48] and molecular dynamics (MD) free energy simulations[49] can calculate the free energies of residue ionization considering a small region of the protein in detail. Recent constant pH molecular dynamics simulations allow pH titrations to be followed in MD simulations.[50,51] Methods that do not involve quantum mechanics simulations calculate how specific interactions with polar or charged groups replace the lost favorable interactions with water when acidic and basic side-chains are buried in the protein. The most routine methods, which allow evaluation of all residues in a protein, are based on continuum electrostatics (CE) using the Poisson-Boltzmann equation[52–55] although other approaches are used successfully.[16,56–58] To obtain higher accuracy, more recent methods use non-uniform dielectric constants in the Poisson-Boltzmann equation[59–62] or average the results in multiple protein structures[63,64] to provide more flexibility in modeling the protein dielectric response. In CE methods, the interactions with solvent water are encapsulated in the reaction field (Born solvation) energy. As a charge is moved into the protein interior, this favorable energy is diminished thus stabilizing the neutral form. Pair-wise interactions amongst backbone, polar, charged side-chains, and ligands are screened by the high dielectric water and by the dielectric properties assigned to protein itself. In a protein with n ionizable groups there are 2^n charged states. If the decrease in solvation energies and the compensating pair-wise interactions between charged groups can be calculated, the Boltzmann distribution of ionization states can be obtained as a function of pH. Given the large number of microstates this is generally obtained by Monte Carlo sampling.[52] Hybrid methods that combine molecular mechanics and continuum electrostatics have also been developed.[1,19,65–69] These keep the background dielectric constant low but allow consideration of multiple positions of protein side-chains. The aim is to incorporate some of the reorganization of the protein to changes in charge into explicit conformational changes rather than averaging them in the dielectric constant.

Multi-conformation continuum electrostatics (MCCE) samples conformational and ionization states in one Monte Carlo simulation. Traditional pK_a calculations allow residues to be ionized or neutral. MCCE uses Monte Carlo sampling to equilibrate side-chain conformations and ionization changes in the same simulation.[19,66] This lets the protein change the orientation of surrounding dipoles when a residue undergoes a change in charge, providing more accurate pK_a.[19] Changes in polar and charged side-chain positions are followed explicitly as a function of pH rather than averaged into the dielectric constant, allowing the unique surroundings of each charged group to be taken into account. Here, MCCE was used to calculate the degree of burial of ionizable groups in 490 proteins. This moves beyond many previous surveys of buried charged groups in that the focus is not on the geometry of the environment but on the energy of ionized and neutral states. The ionization states of all residues were calculated to determine if buried residues are likely to be ionized at physiological pH. The environment of each residue was analyzed to find which interactions in the protein are most important to keep acids and bases ionized. The energetic role of salt bridges, backbone dipoles, and polar side-chains in favoring ionization was

Table 1. Protein and residue composition

SCOP class	α	β	α/β	α+β	Mix and others
Number of proteins	93	110	107	112	69
Number of residues	≤100		100< to ≤300		>300
Number of proteins	35		255		200
Residues/protein	70±20		190±56		519±231
Ires/100res	26.4±13.7		24.2±8.9		23.0±10.0
Ires/100res with $\Delta\Delta G_{rxn}$> 5Δp*K*units	1.87±2.28		2.89±2.20		4.49±1.95
Ires/100res with $\Delta\Delta G_{rxn}$ Sufficient of yield 99% neutrality	1.38±1.90		1.93±1.75		3.03±1.31

Proteins were chosen from the pre-compiled Culled-PDB list[101] with a resolution of 2.0 Å or better, less than 20% percent identity, and *R*-factor less than 0.25. Mix, proteins with domains in different SCOP classes; others, other SCOP classes; Ires, the ionizable residues Arg, Asp, Glu, and Lys. $\Delta\Delta G_{rxn}$(1%): $\Delta\Delta G_{rxn}$ sufficient to shift residue pK_as so they would be 99% neutral at pH 7 in the absence of other interactions. The $\Delta\Delta G_{rxn}$ values are: Arg, 7.5; Lys, 5.8; Asp, 5.1; and Glu, 4.9Δp*K* units. One Δp*K* unit is 1.36 kcal/mol.

determined. Lastly, the conservation of buried surface ionizable residues was compared.

Results

A total of 490 proteins were analyzed ranging in size from 36 to 1357 residues and sampling all SCOP motifs[70] (Table 1). The biological units for 209 are monomeric while 281 are oligomeric proteins. Of the 154,537 residues, 23% are ionizable, and their frequency is ordered Glu>Lys>Asp>Arg with 33% more Glu than Arg (Table 1). Several criteria can be used to define which ionizable residues are buried. The loss in reaction field energy, $\Delta\Delta G_{rxn}$, is used here. This term measures by how much removing the residue from water destabilizes the ionized form more than the neutral form of a residue. $\Delta\Delta G_{rxn}$ provides a numerical value for burial that directly influences the resultant, calculated pK_a. A $\Delta\Delta G_{rxn}$ value of 6.8 kcal/mol is sufficient to shift a residue pK_a by five pH units if there are no other interactions in the protein. Given a reference, solution p$K_{a,sol}$ of 12.5 for Arg, 10.8 for Lys, 3.9 for Asp, and 4.1 for Glu, a $\Delta\Delta G_{rxn}$ of 5 Δp*K* units shifts *in situ* pK_a to 7.5 (Arg), 5.8 (Lys), 8.9 (Asp), and 9.1 (Glu). With these pK_a values, the acids are 1% ionized, Lys 6% ionized and Arg 76%

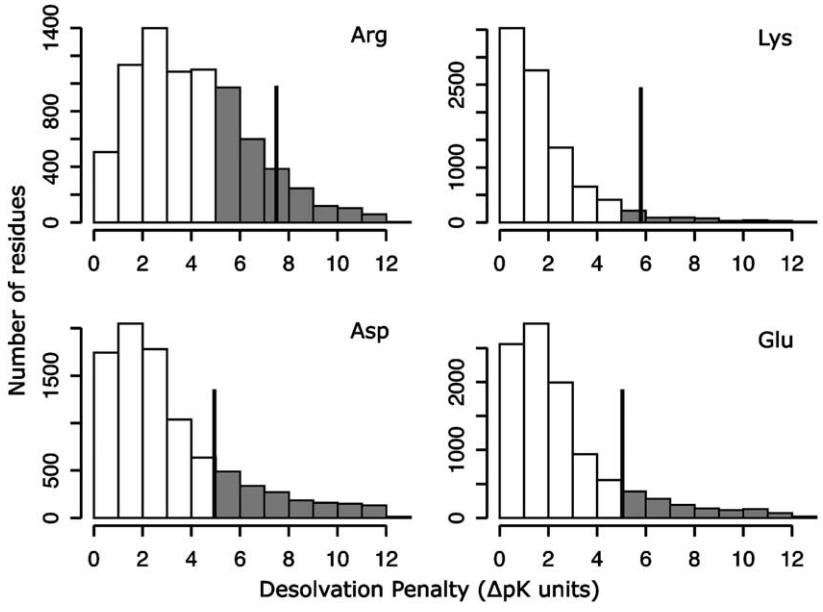

Figure 1. Distribution of the loss of reaction field (solvation) energies ($\Delta\Delta G_{rxn}$) for basic and acidic sidechains in 490 proteins. Shaded region, $\Delta\Delta G_{rxn}$>5 Δp*K* units (6.8 kcal/mol); vertical line, $\Delta\Delta G_{rxn}$(1%) (see Table 1).

Table 2. Number of buried ionizable residues

	Arg	Asp	Glu	Lys
Number of residues	7707	8976	10,232	9277
$\Delta\Delta G_{rxn}$>5 Δp*K* units	2483 (32.2%)	1731 (19.3%)	1326 (13.0%)	566 (6.1%)
ΔG_{rxn}>$\Delta\Delta G_{rxn}$ (1%)	681 (8.8%)	1677 (18.7%)	1355 (13.2%)	382 (4.1%)

See the legend to Table 1 for $\Delta\Delta G_{rxn}$limits for $\Delta\Delta G_{rxn}$(1%).

ionized at pH 7. A second criterion takes into account the differences of the solution $pK_{a,sol}$ value of the different groups, defining the burial thresholds so 99% of the residues would be neutral at pH 7 in the absence of any stabilizing interactions ($\Delta\Delta G_{rxn}(1\%)$). The $\Delta\Delta G_{rxn}(1\%)$ value is 7.5 ΔpK units for Arg, 5.8 for Lys, 5.1 for Asp, and 4.9 for Glu. This will bring the pK_a values of basic amino acids down to 5.0 and those of acidic amino acids up to 9.0 in the absence of other interactions.

The accessible surface area (ASA) or the distance to the surface[71] can also be used to determine the degree of residue burial. These describe the location of the residue but do not directly provide an energy penalty for burial. ASA, the most common measurement, treats an atom as solvent accessible if a water molecule with a specified radius (1.4 Å) can touch it.[72–74] The ASA of the functional side-chain oxygen and nitrogen atoms of acidic and basic residues are roughly correlated with their $\Delta\Delta G_{rxn}$ value (Figure 2). Differences between the two measures can arise because a residue in a concave cavity has less interaction with solvent than one exposed on a convex surface. In addition, a buried residue near the surface would have a greater reaction field energy than one that is deeply buried though they might both have little surface exposure. DelPhi[75] also fills cavities in the protein with a dielectric constant of 80, increasing the favorable reaction field energy of some buried residues far from the surface. Overall residues with $\Delta\Delta G_{rxn} > 5.0$ ΔpK units are >80% buried. Arg with its extended guanidinium group has the poorest correlation between ASA and $\Delta\Delta G_{rxn}$.

Using a $\Delta\Delta G_{rxn} > 5.0$ ΔpK units there are 3049 buried bases and 3057 buried acids (Figure 1 and Table 2). Arg residues are the most likely to be buried and Lys the least. There are 4.4 times as many buried Arg as Lys and 30% more buried Asp than Glu. Because of the very high $pK_{a,sol}$ value for Arg, using the criterion of $\Delta\Delta G_{rxn}(1\%)$ defines many fewer as buried, while changing the number of other residues little. With this criterion, there are 2.8 times as many buried acids as bases, with 1.8 times as many Arg as Lys. Overall, there are on average 3.95/100 residues with $\Delta\Delta G_{rxn} > 5.0$ ΔpK units and 2.65/100 with $\Delta\Delta G_{rxn}(1\%)$. There are more buried ionizable residues/100 found in larger proteins, which have more internal volume (Table 1). Proteins with >300 residues have 4.5 ± 2 ionizable residues with $\Delta\Delta G_{rxn} > 5.0$ ΔpK units. In addition, approximately one quarter of His and Tyr are buried with $\Delta\Delta G_{rxn} > 5.0$ ΔpK units.

Charge states of the ionizable residues

MCCE was used to determine the charge states of the ionizable residues. Residues that are ≥90% ionized at pH 7 are defined as ionized. Overall, 93% of all Asp, Glu, Arg and Lys residues are ionized here. The percentage ionized decreases very slowly as the desolvation penalty increases (Figure 3). Residues with $\Delta\Delta G_{rxn} > 7.5$ ΔpK units are still 78% ionized. Thus, most buried ionizable residues have sufficient interactions with other charges and dipoles in the protein to compensate for the loss of solvation energy incurred when they are buried.

When analyzing the ionizable residues, they are grouped by their desolvation penalties. Residues with $\Delta\Delta G_{rxn} \leq 2.5$ ΔpK units (the surface residues) are defined as being in class I, with $2.5 < \Delta\Delta G_{rxn} \leq 5.0$ ΔpK units in class II, with $5.0 < \Delta\Delta G_{rxn} \leq 7.5$ ΔpK units in class III, and with $\Delta\Delta G_{rxn} > 7.5$ ΔpK units (the deeply buried residues) in class IV.

Arginine

Arg is the most likely to remain charged when buried in large part because of its very high $pK_{a,sol}$

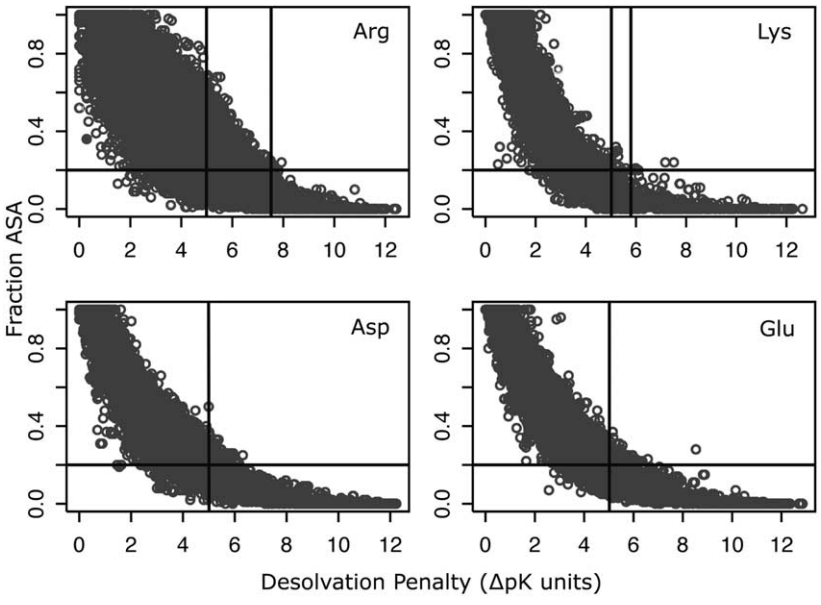

Figure 2. The relationship between fraction ASA and desolvation energy ($\Delta\Delta G_{rxn}$) for each ionizable residue. ASA values of side-chain polar oxygen and nitrogen atoms calculated with SURFV.[103–105] The fully exposed surface determined in Gly-X-Gly tripeptides. Horizontal line at 20% ASA, and vertical lines at $\Delta\Delta G_{rxn} > 5$ ΔpK units and $\Delta\Delta G_{rxn}(1\%)$.

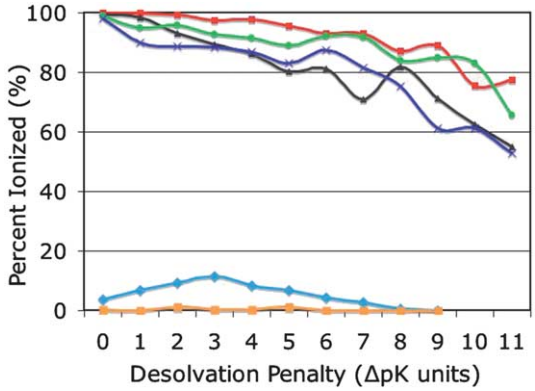

Figure 3. Fraction of residues which are ≥90% ionized at pH 7. Arg (■), Lys (▲), Asp (●), Glu (×), His (◆), and Tyr (■).

value. The class I Arg (the surface Arg) have small and on average slightly unfavorable interactions with the backbone (ΔG_{pol}) and slightly favorable interactions with other charged or polar groups (ΔG_{res}) (Table 3). Only three of the class I Arg residues are calculated to be less than 90% ionized. Two of them, ArgA32 of 1JAY and ArgA18 of 1OAA, are in NADP binding sites. In this study, many substrates including NADP do not have atomic charges. Groups like the anionic phosphates screen the residues from interacting with water, but they do not stabilize ionization of the Arg as they would if they had been properly charged. At the interface of the homo-dimeric protein, 1OR3, ArgA136 and ArgB147 are close and their unfavorable pair-wise interactions do not allow both to be ionized at pH 7. The class I Arg residues do not interact strongly with other groups due to the screening of water, and only 5.2% have even one favorable interaction of ≥2.5 Δp*K* units with other residues or backbone dipoles (Table 4 and Figure 5).

Burying an Arg residue reduces the favorable interaction of the charge with water. However, 86% of the deeply buried Arg residues are ionized (Table 3). The average pair-wise interaction with side-chains (ΔG_{res}) for Arg in this class is −9.6 (±6.1)Δp*K* units stabilizing the charged states, of the same order as the destabilizing $\Delta\Delta G_{rxn}$ (9.0 (±1.2)Δp*K*). In contrast, the average backbone dipole interactions are only slightly favorable (−0.6 (±3.5)Δp*K*) (Table 3). The amide dipoles tend to yield a positive potential inside proteins with relatively small negative regions that can be arranged to stabilize cations.[27] Thus, 81% of

Table 3. Electrostatic interactions at each level of burial

Res		$\Delta\Delta G_{rxn}$	A (%)	B (%)	$\Delta\Delta G_{rxn}$	$\Delta\Delta G_{pol}$	$\Delta\Delta G_{res}$
A. *Charged residues (≥90% ionized at pH 7)*							
Arg	I	≤2.5	30.9	99.9	1.6±0.6	0.1±0.5	−0.6±1.1
7451 res	II	2.5–5.0	36.1	98.0	3.7±0.7	0.0±1.1	−2.4±2.6
96.7%	III	5.0–7.5	22.1	94.4	6.0±0.7	−0.4±2.1	−5.6±4.2
	IV	>7.5	7.6	86.2	9.0±1.2	−0.2±3.5	−9.6±6.1
Lys	I	≤2.5	75.3	98.7	1.1±0.6	0.0±0.5	−0.9±1.4
8870 res	II	2.5–5.0	15.7	89.2	3.4±0.7	−0.4±1.7	−3.1±3.2
95.6%	III	5.0–7.5	3.0	79.3	5.9±0.7	−1.3±2.9	−5.0±4.4
	IV	>7.5	1.6	69.7	9.3±1.4	−2.5±5.3	−8.9±7.1
Asp	I	≤2.5	51.4	96.7	1.3±0.7	−1.0±0.9	−0.8±1.3
8427 res	II	2.5–5.0	25.7	93.5	3.5±0.7	−2.2±1.7	−3.0±3.0
93.9%	III	5.0–7.5	9.8	90.6	6.1±0.7	−4.1±2.9	−6.3±6.4
	IV	>7.5	6.9	82.0	9.4±1.3	−6.5±4.1	−8.4±6.4
Glu	I	≤2.5	59.5	93.0	1.2±0.7	−0.4±0.6	−0.8±1.3
9190 res	II	2.5–5.0	20.3	88.1	3.4±0.7	−1.2±1.4	−3.2±2.9
89.8%	III	5.0–7.5	6.4	85.0	6.1±0.7	−3.1±2.8	−6.2±4.6
	IV	>7.5	3.5	65.8	9.4±1.3	−4.7±3.4	−9.6±6.0
B. *Neutral residues (<90% ionized)*							
Arg	I	≤2.5	0.0	0.1	2.3±0.3	1.2±0.8	1.4±1.2
256 res	II	2.5–5.0	0.8	2.0	3.9±0.7	0.9±1.0	1.4±2.1
3.3%	III	5.0–7.5	1.3	5.6	6.2±0.7	1.4±2.3	−0.4±2.8
	IV	>7.5	1.2	13.8	9.5±1.3	1.1±3.7	−1.8±3.5
Lys	I	≤2.5	1.0	1.3	1.9±0.5	0.3±0.8	1.8±1.4
407 res	II	2.5–5.0	1.9	10.8	3.6±0.7	0.1±1.2	0.3±1.6
4.4%	III	5.0–7.5	0.8	20.7	6.0±0.8	0.1±1.5	−0.7±1.8
	IV	>7.5	0.7	30.3	9.8±1.6	1.8±2.8	−2.8±3.8
Asp	I	≤2.5	1.8	3.3	1.6±0.5	−0.4±0.8	1.1±1.1
549 res	II	2.5–5.0	1.8	6.5	3.7±0.7	−1.5±1.7	0.8±2.2
6.1%	III	5.0–7.5	1.0	9.4	6.1±0.7	−3.1±2.8	3.1±3.5
	IV	>7.5	1.5	18.0	10.1±1.4	−4.5±3.5	3.2±4.8
Glu	I	≤2.5	4.5	7.0	1.7±0.5	−0.2±0.6	0.7±0.9
1042 res	II	2.5–5.0	2.7	11.9	3.5±0.7	−0.5±1.6	0.1±2.1
10.2%	III	5.0–7.5	1.1	15.0	6.0±0.7	−1.4±2.2	0.6±3.3
	IV	>7.5	1.8	34.2	9.9±1.3	−3.4±3.4	2.6±4.9

Units are in Δp*K* (1 Δp*K* unit=1.36 kcal/mol). A, percentage of all amino acids of this type at this reaction field penalty with this charge state. B, percentage of amino acid with this degree of burial which have this charge state. For both ionized and neutral residues Δ*G* values are Δ*G* (ionized form)-Δ*G* (neutral form). All free energy terms are for standard state at pH 7 ($\Delta G^{o\prime}$).

Table 4. Charged ionizable residues and break down of pair-wise interactions

					% With at least one strong interaction ($\Delta G_{elec} <= -2.5$ ΔpK units, -3.4 kcal/mol)					
Res		$\Delta\Delta G_{rxn}$	Total	% None	% Strong	Salt	Polar	bkbn	Others	alt conf.
Arg	I	≤ 2.5	2377	94.8	5.2	4.9	0.0	0.3	0.1	0.6
	II	2.5–5.0	2781	64.8	35.2	31.5	0.8	3.2	0.7	6.8
	III	5.0–7.5	1696	21.0	79.0	64.4	3.2	18.1	1.1	6.6
	IV	>7.5	586	5.6	94.4	80.9	8.2	26.1	5.1	2.6
Lys	I	≤ 2.5	6970	88.4	11.6	11.3	0.1	0.3	0.1	1.4
	II	2.5–5.0	1451	47.6	52.4	44.7	0.8	8.2	0.8	16.3
	III	5.0–7.5	276	27.5	72.5	59.8	2.2	27.2	0.7	26.4
	IV	>7.5	152	2.0	98.0	74.3	9.9	46.7	5.3	8.6
Asp	I	≤ 2.5	4605	84.9	15.1	8.8	0.3	6.6	0.2	4.8
	II	2.5–5.0	2299	31.6	68.4	40.5	2.7	39.5	2.8	12.7
	III	5.0–7.5	879	5.2	94.8	63.9	12.3	67.9	3.4	4.9
	IV	>7.5	619	0.5	99.5	73.3	31.2	82.4	7.1	1.6
Glu	I	≤ 2.5	6046	89.7	10.3	9.6	0.0	0.9	0.1	5.2
	II	2.5–5.0	2071	44.4	55.6	47.8	1.4	14.7	1.4	23.0
	III	5.0–7.5	657	8.8	91.2	70.2	12.8	49.3	3.5	11.1
	IV	>7.5	360	0.6	99.4	84.2	26.7	71.9	6.4	3.0

Count of pair-wise interactions $\leq -2.5\Delta pK$ units (3.4 kcal/mol) for residues at least 90% ionized at pH 7. Total, number of residues; %none, have no individual interactions; %strong, residues with at least one interaction; Ion pair, having at least one favorable interaction with residue of opposite charge. This can be Asp or Glu to Lys, Arg, Ntr or ionized His; Lys or Arg to Asp, Glu, or Ctr; polar, making a favorable interaction to Ser, Thr, Asn, Gln, or neutral His or Tyr; bkbn, the total interaction to the whole backbone is $\leq -2.5\Delta pK$ units; Others, interactions to ligands bound to protein, alt conf., residues take alternate positions at pH 7. Here the Boltzmann weighted $\Delta\Delta G_{rxn}$is at least $1\Delta pK$ unit smaller than found for the original PDB file conformer.

the class IV Arg residues are stabilized by ion pairs with acidic amino acids, while only 26% are stabilized by backbone dipoles. In addition, 8% of these Arg residues are stabilized by polar groups, usually Asn, Gln and Tyr. Very few stabilizing interactions with Ser or Thr are found (Table 5). Five percent also have favorable interactions with the non-amino acid ligands provided with appropriate charges. These include PO_4^{-2}, SO_4^{-2} or the propionic acids on buried heme groups. However, the percentage favorable interactions with the non-amino acids would be higher had all groups treated properly. Few of the less deeply buried Arg (classes II and III) have more than one strong interaction partner (Figure 5). However, only 5% of the class IV Arg residues are ionized without having any strong

Table 5. Interactions between ionizable residues and polar residues

		Number of charge–polar residue interactions			Side-chain O↔N distance	
Res–type	Int–with	$<=-2.5$	$<=-1.5$	$<=-1.0$	$<=3.9$ Å	$<=3.2$ Å
Arg	Asn	45	97	203	582	311
	Gln	32	90	173	502	239
	His	7	27	69	321	47
	Ser	2	32	53	474	241
	Thr	5	18	44	370	187
	Tyr	47	86	162	517	199
Lys	Asn	19	56	103	297	157
	Gln	3	34	72	239	117
	His	13	21	38	83	29
	Ser	0	5	22	179	82
	Thr	0	1	22	181	96
	Tyr	9	30	68	259	84
Asp	Asn	55	165	248	665	380
	Gln	30	95	184	466	297
	His	258	403	557	491	332
	Ser	158	404	567	985	746
	Thr	105	361	517	812	572
	Tyr	45	200	300	445	369
Glu	Asn	16	66	133	472	232
	Gln	25	66	122	509	341
	His	201	349	476	558	378
	Ser	72	207	286	660	396
	Thr	71	188	252	562	294
	Tyr	40	180	285	596	457

Numbers of strong pair-wise interactions. These include all residues independent of their level of burial. The count of residues at a given distance does not take into account the angle between the residues.

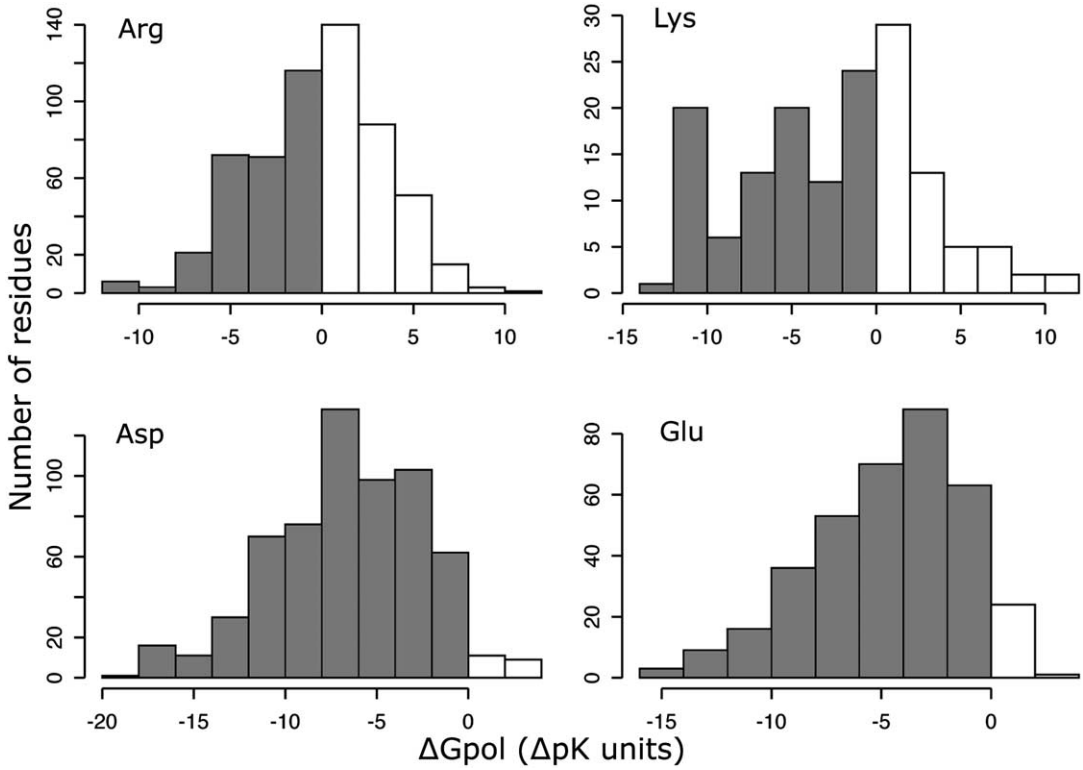

Figure 4. Distribution of ΔG_{pol} (pair-wise interaction with backbone dipoles) for class IV residues ($\Delta\Delta G_{rxn} > 7.5$ ΔpK units). Shaded region, favorable interactions.

stabilizing interactions. One- third of these occupy a conformer with significantly less $\Delta\Delta G_{rxn}$ value than the native Arg rotamer used to classify the burial of residues (Table 4). The rest are stabilized by several interactions below the -2.5 ΔpK unit cutoff.

Lysine

Lys is significantly less likely to be buried than Arg with only 6% having >5.0 ΔpK units $\Delta\Delta G_{rxn}$ (Table 2). While 95% of all Lys residues are ionized, 30% of these in class IV are not. Class IV Lys residues are twice as likely as the buried Arg to be at least partially neutral. As with Arg, pair-wise interactions, on average, pay back the $\Delta\Delta G_{rxn}$. Overall, 98% of the class IV Lys residues have at least one strong interaction partner; 74.3% of them have at least one strong interaction with an acidic residue, 47% with backbone dipoles, and 9.9% with polar side-chains (Table 4). As with Arg, Ser and Thr do not stabilize Lys ionization (Table 5). The interactions with backbone dipoles are more important for stabilizing charged Lys than Arg (Tables 3 and 4). All but the most surface exposed Lys residues are more likely to have multiple strong pair-wise interactions than do Arg (Figure 5).

Conformational changes also stabilize Lys ionization. A significant fraction of class II and III Lys residues find more exposed positions in MCCE Monte Carlo sampling (Table 4). The class IV Lys residues do not generally have available conformers significantly closer to the surface.

Aspartic and glutamic acids

Nineteen percent of the Asp and 12% of the Glu residues have >5 ΔpK units $\Delta\Delta G_{rxn}$ and of these 87% of the Asp and 77% of the Glu residues were ionized. Three percent of the surface (class I) Asp and 7% of these Glu residues were calculated to be $<90\%$ ionized largely because small destabilizing pair-wise interactions add to the small $\Delta\Delta G_{rxn}$ (Table 4). The backbone dipoles help stabilize ionization of the acids significantly more than the bases at all levels of burial. The backbone interactions stabilize class IV Asp and Glu residues by -6.5 (± 4.1) and -4.7 (± 3.4)ΔpK units, while they only stabilize the class IV Arg and Lys residues by -0.2 (± 3.5) and -2.5 (± 5.3)ΔpK units (Table 3 and Figure 4).

The average residue pair-wise interactions for the ionized acids and bases are similar at each level of burial. Seventy-three percent of the class IV Asp and 84% of the class IV Glu residues have at least one strong ion pair, a slightly smaller percentage than found for the bases. About 10% of these are made to ionized His residues. However, the polar side-chains (including neutral His) play a much larger role in stabilizing the acids with 31% of the deeply buried Asp and 27% of the Glu having at least one strong, favorable polar interaction. The hydroxyl containing side-chains Ser and Thr which do not stabilize the bases do interact favorably with the acids, while Asn and Gln and Tyr residues keep both acids and bases ionized. Three percent of the

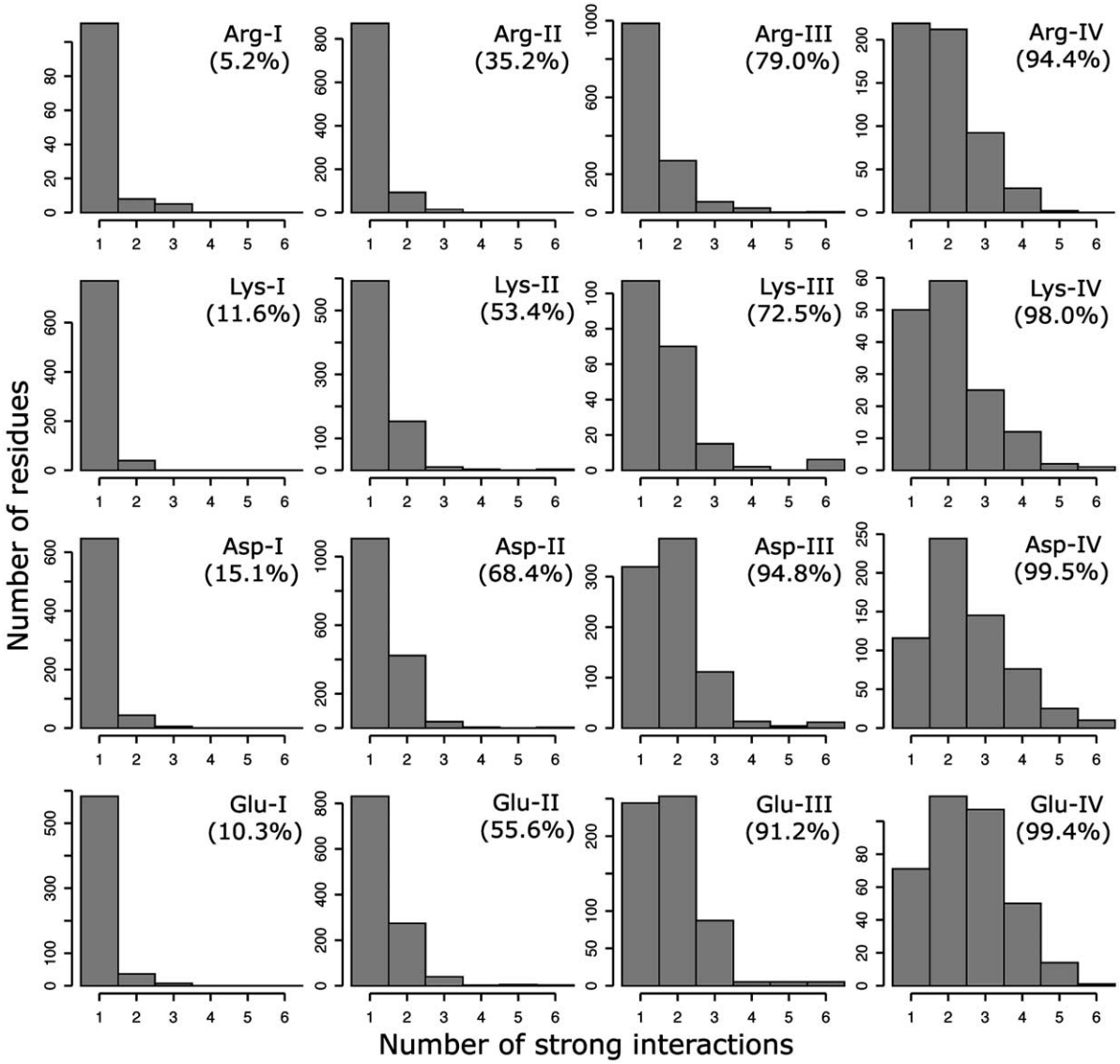

Figure 5. Distribution of the number of individual strong interactions (≤ -2.5 ΔpK units) to other residues, or ligands. A net $\Delta G_{pol} \leq 2.5$ ΔpK units counts as one interaction. Residues with no strong interaction are not shown. The percentage value in each panel shows the percentage of residues with one or more interaction.

Asp and 4% of the Glu residues interact with the non-amino acid ligands considered here especially ions such as Zn^{2+}, Ca^{2+}, and Mg^{2+}.

Overall, 95% of the class III Asp and 91% of the class III Glu residues have at least one strong interaction partner, and the percentages increase for those in class IV. Class IV acids have more multiple interactions than do the bases (Figure 5).

Environment of neutral acids and bases

A small fraction of the Asp, Glu, Arg, and Lys residues are <90% ionized at pH 7 (Figure 3 and Table 3). The neutral residues are often at positive potential from the backbone, stabilizing ionization of acids while destabilizing the bases. Overall the ΔG_{res} value changes more between the ionized and neutral acids and bases than does the ΔG_{pol}. It is notable that the average ΔG_{res} tends to destabilize ionization of the neutral acids, while they still stabilize ionized form of the bases. This may arise because there are more class IV acids than bases.

Subunit interface and ionizable residues

There are 250 oligomeric proteins analyzed here with 2030 Arg, 2448 Lys, 2354 Asp, and 2715 Glu. The desolvation penalties calculated in isolated monomeric units of oligomeric proteins were compared (Figure 6). Residues were identified as being buried at an interface if they have a monomer $\Delta\Delta G_{rxn} \leq 2.5$ and oligomer$\Delta\Delta G_{rxn} > 5.0$ ΔpK units. There are 77 Arg, 29 Lys, 56 Asp, and 70 Glu residues in this class. This represents ≈3% (Arg and Asp) and ≈6% (Lys and Glu) of the class III and IV residues. There are small changes between the interfacial and other buried acids (Table 6). The Arg and Lys residues at the interface are as likely to

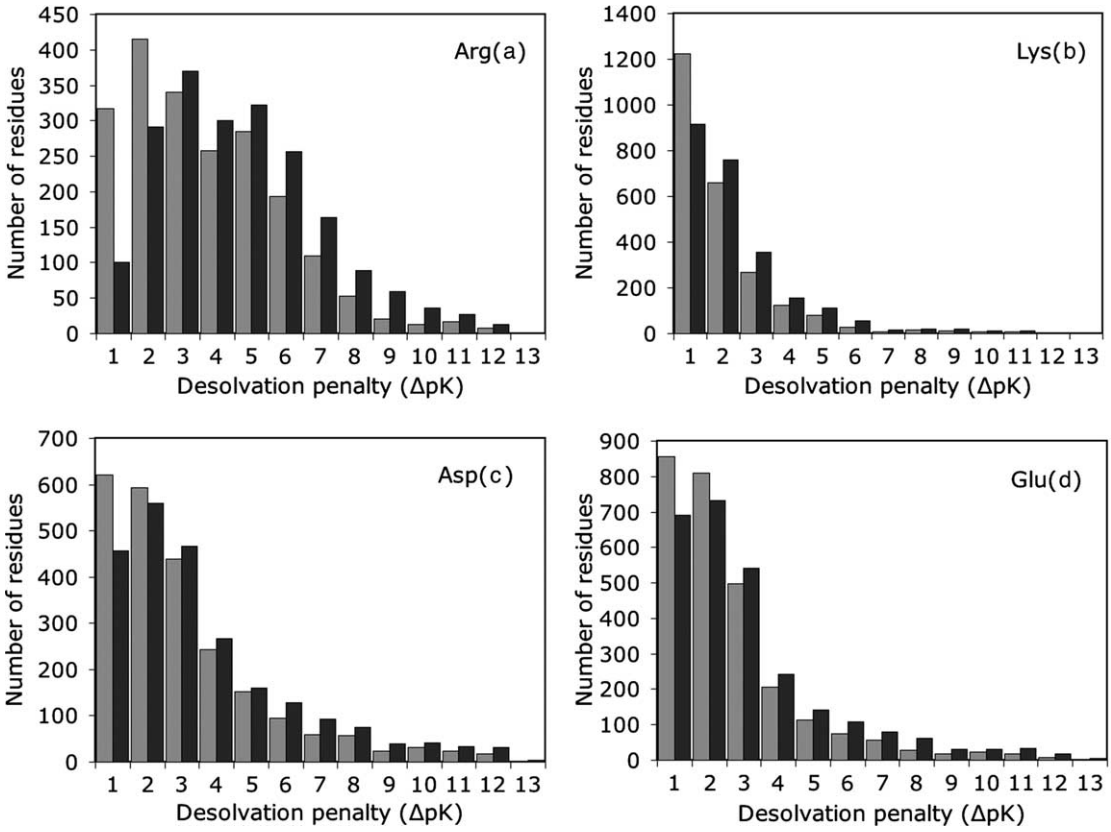

Figure 6. Distribution of the solvation energies of the ionizable residues in oligomer decomposed into monomers and in the holoprotein. Dark (■) bar, oligomers; gray (■) bar, monomeric units.

be ionized as are other class III and IV residues. Interfacial Glu are ≈5% more likely and Asp ≈5% less likely to be charged than the other buried residues. The probability of having strong interactions with the backbone is smaller for the interfacial bases while these are more likely to have larger pair-wise interactions with other residues.

Conservation of acidic and basic amino acids

Residue conservation, defined in the HSSP database,[76] was determined for residues at each level of burial (Figure 7). Surface residues are less likely to be conserved than those buried in the protein core. For example, 25% of the surface residues ($\Delta\Delta G_{rxn} \leq 1.0$ ΔpK unit) and 85% of the most deeply buried residues ($\Delta\Delta G_{rxn} > 10$ ΔpK units) are at least 50% conserved.

Burial and ionization of histidine and tyrosine

His has a solution $pK_{a,sol}$ value of 6.5 so it is only 24% ionized at pH 7 when free in solution. Tyr has a $pK_{a,sol}$ value of 10.8 and is only 0.016% ionized at pH 7 in solution. Twenty-three percent of the His and 26% of the Tyr residues have $\Delta\Delta G_{rxn} > 5.0$ ΔpK. Overall 0.037% of Tyr and 6% of His residues are calculated to be ≥90% ionized at pH 7 (Figure 3). For the buried residues ($\Delta\Delta G_{rxn} > 5.0$ ΔpK), 4% of the His (33/841) and ≪1% of the Tyr (4/1395) are ionized. For the His it may be more appropriate to ask how many are as ionized in the protein as in the solution. In total 30% (1078/3645) are at least 25% ionized in the protein and 9% (71/841) of those with $\Delta\Delta G_{rxn} > 5.0$ are this ionized. Thus, in these calculations the protein destabilizes His ionization relative to solution. His residues are often involved in binding metals or cofactors and these His need to be neutral.

Discussion

The ionization states of more than 35,000 ionizable residues in 490 proteins with a range of sizes and folds were calculated. A significant minority of these residues are found to be buried in the protein as defined by solvent accessibility or by loss of reaction field (solvation) energy. MCCE calculations show that the large majority of these residues are ≥90% ionized at pH 7. The fraction ionized falls off very slowly with residue burial (Figure 3), showing that in general electrostatic interactions with other charges and dipoles in a protein are predicted to replace the solvation energy lost when a group is removed from water. When residues with $\Delta\Delta G_{rxn} >$ 5 ΔpK units (6.8 kcal/mol) are considered, there are

Table 6. Characteristics of ionized residues buried at subunit interfaces

					%ΔG_{res} <= −2.5 ΔpH units (−3.4 kcal/mol) (%)				
Res	Total	Avg ΔG_{rxn}	Avg ΔG_{pol}	Avg ΔG_{res}	% All	Salt	Polar	bkbn	Others
Arg	69	6.8±1.5	0.3±2.6	−7.4±5.8	76.8	72.5	4.3	13.0	0.0
Lys	22	6.7±1.7	−0.6±3.8	−7.8±5.8	86.4	81.8	0.0	18.2	0.0
Asp	43	6.9±1.7	−4.1±3.5	−7.3±6.7	95.6	73.3	17.8	68.9	4.3
Glu	60	6.6±1.3	−3.7±3.3	−7.2±6.0	93.3	63.3	16.7	55.0	3.3

Residues were defined as buried at the subunit interface if monomer $\Delta\Delta G_{rxn} \leq 2.5$ and oligomer $\Delta\Delta G_{rxn} > 5.0$. Only residues ≥90% ionized at pH 7 are considered here. Table 3 provides values for average interaction energies and Table 4 the percentage of strong interactions for all residues.

essentially the same number of buried acids (3057) and bases (3049). Considering only the most deeply buried group ($\Delta\Delta G_{rxn} > 7.5$ ΔpK units), there are more acids (1308) than bases (899). The ability of proteins to support deeply buried acids seems to be greater than bases. The backbone dipoles tend to raise the potential inside any protein.[27] In addition, while the amide group of Asn and Gln can stabilize ionization of both acids and bases, hydroxyl groups are much better hydrogen bond donors than acceptors so they tend to stabilize anions but not cations (Table 5).

The data presented here represent the prediction of the ionization states of more than 35,000 residues. pK_as are not presented because of the level of uncertainty found in this implementation of MCCE (MCCE1.0). Continuum electrostatics calculations tend to overestimate the interactions between surface charged groups.[7,77] Benchmark studies on small proteins using no conformational sampling find better match to experimental data with dielectric constants of 10 to 20 reducing the interactions.[7] In contrast, MCCE calculations find good matches to benchmarks with dielectric constants of 4 to 8. The version used, here (MCCE1.0) has been benchmarked in calculations of known amino acids[19] and of hemes in small cytochromes.[78] Calculations of 166 ionizable residues in 12 small soluble proteins such as lysozyme and barnase showed root-mean-square deviation between the calculated and measured of 0.83 pH unit with >90% having errors of <1 pH unit using $\varepsilon = 4$.[19] However, this was carried out using a SOFT function that decreased very strong interactions by about 50%. This is necessary because MCCE1.0 does provide enough well solvated conformers for surface conformers. However, as all conformers are present in all calculations of pair-wise interactions adding more surface conformers increases the region with a low dielectric constant, artificially increasing the interactions. This is not a problem for calculations of $\Delta\Delta G_{rxn}$, since these are determined in a calculation with only one conformer per residue (see Georgescu *et al.*[19] for a more complete description).

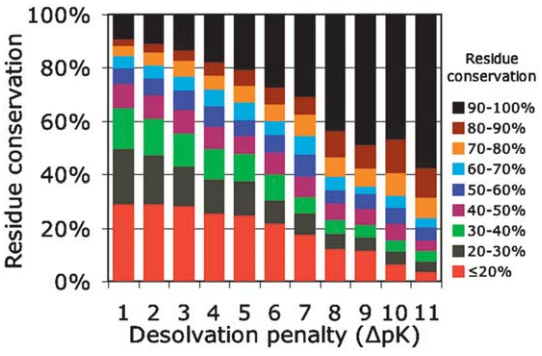

Figure 7. Degree of conservation as a function of the $\Delta\Delta G_{rxn}$. HSSP[76] conservation weight was used.

While ionization states for both surface and buried residues are presented here, the focus is on the ionization state of the buried charged groups. Studies of buried active sites in photosynthetic reaction centers,[79–81] bacteriorhodopsin,[82,83] halorhodopsin,[84] and fumarate reductase[85] using MCCE and other continuum electrostatics methods provide the best benchmarks for the analysis of buried charges. These sites are far from the surface so boundary errors are negligible. Calculations with large dielectric constants or with SOFT do not find good agreement with the data. The value of $\varepsilon = 4$ used here provides the best results. Thus, earlier calculations with MCCE1.0 suggest that the ionization states and even pK_as of buried residues are likely to be well predicted.[81,83] The predicted pK_as without SOFT for surface residues, especially those in ion pairs are likely to overestimate the stability of the ionized form, often pushing acid pK_a to be less than zero and bases to move to be over 14.[19] However, the ionization states at pH 7 are likely to be well estimated. A new version of the program (MCCE2.0) has recently become available which has much better rotamer sampling and significantly corrects the problems that arise from errors in the boundary. This provides a consistently good match to buried and surface residues using the same assumptions and parameters. Calculation of pK_as for the set of proteins examined here is in progress.

The work presented here complements earlier surveys of ionizable groups. Some of these looked at geometrically defined features such as surface accessibility or found active sites in H-bonded clusters, and others looked at how specific interactions contribute to stability without calculating site ionization. Contributors such as protein backbone dipoles[86–88] and ion-pairs[26,89] have been

explored in detail. The asymmetry of anion and cation interactions have been seen in surveys of main-chain to side-chain hydrogen bonds which are far more common for acidic than basic side-chains.[90] Surveys of side-chain to side-chain hydrogen bonds also show somewhat more interactions between acids and hydroxyl side-chains.[91] However, when the pattern is defined by the geometrically determined hydrogen bond, it is not nearly as obvious as when interaction energies are considered (Table 5). Previous calculations suggest, as found here, that on average proteins are arranged to enhance interaction with their charged groups.[37] Charge–charge interactions[83,92] replace the loss of solvation energy when residues are buried. The trend that the protein will stabilize buried charges has also been found in surveys of measured pK_as. Surface carboxylic acids[8] and His[93] tend to have relatively unperturbed pK_as. The range of pK_as increases when these groups are buried,[8] not unexpected given the complexities of the interactions within proteins. For both buried acids and His, the average pK_a value shifts show the protein environment stabilizes the ionized form of the residue.

Active site residues are found to be in complex electrostatic environments. Some surveys suggest these buried residues destabilize the protein implying the protein will favor the neural state.[94] Other surveys suggest that active sites are characterized by regions of large electrostatic potential.[95] Many active sites have clusters of residues of like charge forming an internal buffer where a proton is shared between several residues.[79,83,96] These motifs contain residues where ionization states change by proton transfer amongst sites, leading to complex, non-Henderson–Hasselbach titrations.[97,98] One problem is that with clusters of ionizable residues it is often difficult to identify correctly which site has the proton[19] although the net charge of the cluster is easier to ascertain.[83] It seems likely that at least some of the buried, neutral residues are in protein active sites.

The most contentious issue in continuum electrostatics analysis of proteins is the value for the protein dielectric constant. This is a parameter that accounts implicitly for protein rearrangement on changes in charge.[34,56] Values from $\varepsilon=4$[99] to $\varepsilon=20$[100] are used routinely. The higher the value, the larger the average protein response to changes in charge. If carried out correctly, where desolvation energy and pair-wise interactions are both scaled appropriately, most calculated pK_as depend only modestly on the dielectric constant used.[19] This is because pair-wise interactions in proteins tend to stabilize the charged states. Thus, the reduction of the unfavorable desolvation penalty and of the favorable pair-wise interactions found when the dielectric constant is increased tend to cancel. Therefore, with a higher dielectric constant, it is likely that the percentage of residues that are charged would remain high. However, if a higher dielectric constant were used, all of the individual terms used in the analysis of the environment of the ionizable groups would be smaller. Consequently fewer residues would have large $\Delta\Delta G_{rxn}$ values and the compensating ΔG_{res} and ΔG_{pol} would be smaller.

Materials and Methods

Sample protein selection

A total of 490 crystal structures with a resolution of 2.0 Å or better, with less than 20% identity, and an *R*-factor of less than 0.25 were selected from the 1059 proteins in the Dunbrack Culled-PDB list.[101] The subset analyzed here was selected to represent the different SCOP[70] classes and protein sizes. The proteins range in size from 36 to 1357 residues. Four membrane proteins were included while the rest are water-soluble and classified mainly in the four SCOP classes (Table 1). Where possible the biological unit in the EMBL-EBI Likely Quaternary Structure or Macro-Molecular Assembly definitions were used.[102] There are 209 monomers, 25 proteins with non-identical oligomers, and 256 oligomers formed of identical subunits. Additional calculations were run where the subunits of the oligomeric proteins were analyzed separately allowing the comparison of residues buried in a monomer interior and those buried at subunit interfaces. HSSP[76] provided the degree of conservation of each ionizable residue.

Calculation of accessible surface area

The reference surface area for fully exposed side-chain oxygen or nitrogen atoms from ten different Gly–X–Gly polypeptides was determined with SURFV.[103–105]

The list of 490 proteins

12as, 13pk, 16pk, 1a05, 1a17, 1a1x, 1a26, 1a3a, 1a48, 1a4i, 1a6f, 1a6m, 1a73, 1a8d, 1a8e, 1a8 l, 1af7, 1agj, 1agq, 1ah7, 1aie, 1aj2, 1ajj, 1akr, 1aoc, 1aoh, 1aol, 1aop, 1aqz, 1aw7, 1axn, 1ayf, 1ayl, 1ayo, 1ayx, 1azo, 1b12, 1b2p, 1b66, 1b6a, 1b93, 1baz, 1bbz, 1bd3, 1beh, 1bfg, 1bft, 1bg2, 1bgc, 1bgf, 1bkf, 1bkr, 1bm8, 1bpi, 1bqu, 1btk, 1bxa, 1bxe, 1bxm, 1by2, 1byi, 1byq, 1byr, 1c1k, 1c1 l, 1c3d, 1c3w, 1c52, 1c7k, 1ccw, 1ccz, 1cdy, 1ceo, 1cex, 1chd, 1 cmc, 1cnv, 1cpo, 1cqm, 1csh, 1ctj, 1ctq, 1cv8, 1cxq, 1cyo, 1d06, 1d0d, 1d2o, 1d2v, 1d4o, 1d9c, 1daq, 1dbf, 1dcs, 1dek, 1dfm, 1dfn, 1dg6, 1dlw, 1dmg, 1dnl, 1dow, 1dqe, 1dvo, 1dxg, 1dzk, 1dzo, 1e29, 1e30, 1e5k, 1e7 l, 1e85, 1eaz, 1eb6, 1edh, 1eej, 1eer, 1ej2, 1ej8, 1ejg, 1ekr, 1elk, 1elp, 1elw, 1ep0, 1eqo, 1es9, 1et1, 1euj, 1euv, 1euw, 1ex2, 1ext, 1eyh, 1ezi, 1f2t, 1f3u, 1f46, 1f5m, 1f7 l, 1f86, 1f9m, 1faz, 1fcy, 1fe6, 1fi2, 1fip, 1fj2, 1flm, 1fpo, 1fqi, 1fqt, 1fs1, 1fsg, 1fso, 1ft5, 1fvg, 1fvk, 1fw1, 1fw9, 1fye, 1g0 s, 1g13, 1g1j, 1g2b, 1g2i, 1g2r, 1g2y, 1g3p, 1g4i, 1g5t, 1g61, 1g66, 1g6u, 1g6x, 1g7a, 1g8e, 1g8q, 1gak, 1gbg, 1gbs, 1gef, 1geq, 1gmi, 1gmm, 1gmu, 1gmx, 1gnu, 1gou, 1gp0, 1gpr, 1gr3, 1gsa, 1gsm, 1gu9, 1gux, 1gx1, 1gxj, 1gy7, 1gyo, 1h4x, 1h6 h, 1h7c, 1h8p, 1h8u, 1h97, 1h99, 1han, 1hd2, 1hdo, 1he1, 1hfo, 1hoe, 1hpc, 1hqk, 1hsl, 1htw, 1huf, 1huw, 1hxi, 1hxr, 1hyp, 1hzt, 1i0 h, 1i0r, 1i0v, 1i12, 1i1j, 1i2t, 1i40, 1i4m, 1i4u, 1i52, 1i5 g, 1i6p, 1i6w, 1i8a, 1i8d, 1i8o, 1i9 s, 1iab, 1iap, 1iaz, 1icx, 1ido, 1ifr, 1iib, 1ij2, 1ijb, 1ijv, 1ijy, 1ikt, 1im5, 1ioo, 1iow, 1iu8, 1ixh, 1j83, 1j8r, 1j98, 1j9b, 1j9 l, 1jat, 1jay, 1jb3, 1jbe, 1jbk, 1jc4, 1jd5, 1jdw, 1je5, 1jek, 1jeo, 1jf0, 1jf4, 1jf8, 1jfx, 1jhf, 1jhg, 1jhj, 1jid, 1jjv, 1jk3, 1jke, 1jkx, 1jl1, 1jm0, 1jm1,

1jmv, 1jni, 1jr8, 1jya, 1jyh, 1jyo, 1jzg, 1jzt, 1k04, 1k0m, 1k12, 1k1a, 1k3 s, 1k4i, 1k55, 1k6d, 1k94, 1kaf, 1kcq, 1kdo, 1kfn, 1kgb, 1khc, 1km4, 1kng, 1knm, 1knq, 1koe, 1koi, 1kpf, 1kpt, 1kqr, 1kr7, 1ksk, 1kso, 1ktg, 1kug, 1kyf, 1l3k, 1l6p, 1l7m, 1l8b, 1l8r, 1lf7, 1lg7, 1lgp, 1lh0, 1li1, 1lki, 1lkk, 1lm5, 1lmh, 1ln4, 1lo7, 1lr5, 1ly2, 1m0d, 1m48, 1m6p, 1mai, 1mfm, 1mgt, 1mka, 1mla, 1mml, 1mof, 1mrj, 1mrp, 1msc, 1msk, 1mug, 1nbc, 1nfp, 1nkd, 1nkr, 1nox, 1npk, 1nsj, 1nxb, 1oaa, 1ois, 1opc, 1or3, 1oyc, 1pbv, 1pgs, 1pgt, 1phm, 1pin, 1pmi, 1pne, 1ppn, 1ppt, 1psr, 1ptq, 1puc, 1qa7, 1qau, 1qb0, 1qb7, 1qcs, 1qdd, 1qde, 1qfo, 1qgv, 1qhv, 1qip, 1qjb, 1qjc, 1qjp, 1qkr, 1ql0, 1qmy, 1qna, 1qre, 1qst, 1qtn, 1qto, 1r51, 1rh4, 1rhs, 1rss, 1sac, 1sbp, 1sei, 1sfp, 1shk, 1sig, 1slu, 1sra, 1sur, 1svb, 1swu, 1t1d, 1tca, 1ten, 1tfe, 1thv, 1tif, 1tig, 1tl2, 1tml, 1tx4, 1ubi, 1utg, 1vcc, 1vhh, 1vid, 1vie, 1vii, 1vls, 1vns, 1vsr, 1wfb, 1whi, 1wwc, 1yge, 256b, 2a0b, 2acy, 2ahj, 2arc, 2cpg, 2ctc, 2cua, 2eng, 2erl, 2fcb, 2hft, 2hrv, 2igd, 2ilk, 2lbd, 2lis, 2mcm, 2mhr, 2mlt, 2nlr, 2pgd, 2pii, 2psp, 2ptd, 2pth, 2pvb, 2scp, 2sns, 2spc, 2tgi, 2tps, 2wrp, 3bam, 3cao, 3cyr, 3grs, 3pnp, 3prn, 3pvi, 3std, 3vub, 4bcl, 4eug, and 8abp.

MCCE: multiconformation continuum electrostatics

Multi-conformation continuum electrostatic (MCCE) is a hybrid of continuum electrostatics and molecular mechanics.[19] MCCE starts with atomic coordinates from the Protein Data Bank (PDB)[106] as inputs and solves the Poisson-Boltzmann equations by the finite difference method for multiple conformers.[75] MCCE keeps the protein backbone rigid while letting side-chains of polar and ionizable groups sample additional rotamer positions.[19,66] Thus, residues have pre-assigned conformers differing in atomic positions or ionization state. Each ionizable side-chain can be charged or a neutral dipole. The neutral acids have several positions for the terminal proton. Hydroxyl side-chains have protons in their torsion minima and making hydrogen bonds. Asn and Gln have conformers that interchange their terminal oxygen and nitrogen atoms. Both neutral His tautomer conformers are available.

Monte Carlo sampling yields the Boltzmann distribution of conformers for each residue given the electrostatic and non-electrostatic self and pair-wise energy terms. The electrostatic energies for each conformer are calculated with Delphi[75] using dielectric constant of 4 as described by Georgescu *et al.*[19] The self-energy term is the loss in reaction field (desolvation penalty) that occurs when the conformer is moved from water into its position in the protein. The $\Delta\Delta G_{rxn}$ term represents the loss of the reaction field energy of ionized conformers minus the loss for the neutral conformers when the residue is buried. ΔG_{pol} is the difference in electrostatic pair-wise interactions of the ionized and neutral conformers with the amide backbone. Pair-wise electrostatic interactions are also obtained for the interactions of each conformer with all other conformers. The non-electrostatic energy terms are the torsion energy and the pair-wise Lennard-Jones energies of a conformer with the backbone and with all other conformers. Torsion and Lennard-Jones parameters as described by Alexov *et al.*[79] and Parse partial charges[107] are used for the DelPhi calculations. The solution reaction field energies were 12.8ΔpK units (17.4 kcal/mol) for Asp, 12.9 (17.5 kcal/mol) for Glu, 11.7 (15.9 kcal/mol) for Arg, 13.6 (18.5 kcal/mol) for Lys, 8.1 (11.0 kcal/mol) for His, and 10.3 (14.0 kcal/mol) for Tyr. In each case, these values are the difference between the DelPhi reaction field energy for the ionized and neutral form of the residue isolated from the rest of the protein.

The Boltzmann distribution of conformers as a function of pH is determined by Monte Carlo sampling. Protein microstates are created given one conformer for each group including residues, cofactors, and ligands. The energy of microstate x (ΔG_x) is:

$$\Delta G_x = \sum_{i=1}^{M} \delta_x(i)\{\gamma(i)[2.3\, k_b T(\mathrm{pH} - \mathrm{p}K_{sol,i})] + (\Delta\Delta G_{rxn,i} + \Delta G_{torsion,i} + \Delta G_{pol,i} + \Delta G_{fixed,i}^{nonel})\} + \sum_{i=1}^{M} \delta_x(i) \sum_{j=i+1}^{M} \delta_x(j)[\Delta G_{ij}^{el} + \Delta G_{ij}^{nonel}] \quad (1)$$

The first line accounts for the reference chemistry of the group in solution (p$K_{a,sol}$) and the solvent conditions (pH), the second for conformer self energies and interactions with fixed portions of the protein, while the third treats interactions between conformers in the considered microstate. k_bT is 25 meV (0.43 ΔpK units or 0.58 kcal/mol at 20 °C), M the total number of conformers, $\delta_x(i)$ is 1 for selected conformers and 0 for all others. $\gamma(i)$ is 1 for bases, −1 for acids, and 0 for neutral conformers (polar groups, water molecules, neutral acids and bases). p$K_{sol,i}$ is the p$K_{a,sol}$ of group i, $\Delta\Delta G_{rxn,i}$ the difference between the conformer reaction field energy in solution and protein; $\Delta G_{torsion,i}$ the conformer torsion energy; $\Delta G_{pol,i}$ the conformer electrostatic interaction with the backbone dipoles and $\Delta G_{fixed,i}^{nonel}$ the Lennard-Jones interaction with the backbone and side-chains with no conformers. ΔG_{ij}^{el} and ΔG_{ij}^{nonel} are pair-wise interactions between side-chain conformer i and j. Monte Carlo sampling yields conformer occupancies in a Boltzmann distribution of states as a function of pH. The SOFT function, an empirical function used in previous studies[19,78] to reduce the strong pair-wise interactions, is not used here. While small, flexible proteins appear to be best calculated with a higher value for the continuum protein dielectric constant in MCCE, analysis of larger membrane proteins such as bacteriorhodopsin[82,83] and photosynthetic reaction centers[79,108] provides the best match to experimental results with $\varepsilon = 4$.

Crystal water molecules were deleted and the cavities filled by DelPhi with a dielectric of 80, greatly reducing the effort needed to model water properly. The results with explicit and continuum water have been found to be comparable in many[19,78,83] but not all cases.[66,109,110] These proteins contain numerous non-amino acid groups co-crystallized with the protein. Hemes and their propionic acids are treated as described in our earlier studies.[78,111] Metal ions such as Cl^-, Na^+, Ni^+, Cu^+, Mg^{2+}, Zn^{2+}, and Ca^{2+} and PO_4^{2-} and SO_4^{2-} are included in this calculation with appropriate charges. They each have a conformer that does not interact with the protein. This models the ligand leaving the protein if favorable interactions cannot pay the reaction field penalty incurred when it is buried. Other ligands included in the protein data files are simply treated as uncharged, low dielectric material. Of 490 proteins 312 have at least a non-amino acid group, and 225 of them have at least one ligand, which is treated as uncharged here. The most prevalent ligands where were not given appropriate charges are glycerol (GOL), *N*-acetyl-D-glucosamine (NAG), acetate ion (ACT), acetyl group (ACE), FMN, Cd^{2+}, EPE, and ADP.

Determination of interactions of ionizable groups with the rest of the protein in the Monte Carlo sampled ensemble of protein microstates

Each of the multiple conformers used in MCCE has a different reaction field energy. The loss of reaction field energy ($\Delta\Delta G_{rxn}$) for the original conformer in the PDB file is used to decide if a particular acid or base is buried.

The free energy of interaction of a conformer with the backbone dipoles is:

$$\Delta G_{pol,i} = \sum_{a=1}^{n} \Psi_a^{bkbn} q_{a,i} \quad (2)$$

where $q_{a,i}$ is the charge on atom a. The interaction of a residue with the backbone is the Boltzmann weighted $\Delta G_{pol,i}$ for all ionized conformers minus the $\Delta G_{pol,i}$ for neutral conformers at pH 7.0.

Interaction between residues

The free energy of interaction between individual residues uses a Mean Field analysis. The interaction between residue A and B is:

$$\Delta G_{res,A\leftrightarrow B} = \sum_{i=1}^{Aconf} c_i \sum_{j=1}^{Bconf} c_j \Delta G_{Ai,Bj} \quad (3)$$

where Aconf and Bconf are the number of conformers of the two residues and c_{Ai} and c_{Bj} are the Boltzmann weighted occupancy of conformer i and j of residue A and B at pH 7 and $\Delta G_{Ai\leftrightarrow Bj}$ is the conformer to conformer interaction energy. This includes both Lennard-Jones and electrostatic interactions. ΔG_{res} for a given residue sums $\Delta G_{res,A\leftrightarrow B}$ over all conformers of all surrounding residues.

Acknowledgements

This work was supported by a grant from the National Science Foundation (MCB 0212696). We thank Yifan Song for helpful discussion.

Supplementary Data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmb.2005.03.051.

All MCCE1.0 calculated pK_as and associated energy terms for the 490 proteins described here can be found in a searchable database at http://www.sci.ccny.cuny.edu/~mcce. MCCE can be downloaded from the same site.

References

1. Bartlett, G. J., Porter, C. T., Borkakoti, N. & Thornton, J. M. (2002). Analysis of catalytic residues in enzyme active sites. *J. Mol. Biol.* **324**, 105–211.
2. Tiede, D. M., Vashishta, A.-C. & Gunner, M. R. (1993). Electron transfer kinetics and electrostatic properties of the *Rhodobacter sphaeroides* reaction center and soluble *c*-cytochromes. *Biochemistry*, **32**, 4515–4531.
3. Sims, P. A., Wong, C. F. & McCammon, J. A. (2004). Charge optimization of the interface between protein kinases and their ligands. *J. Comp. Chem.* **25**, 1416–1429.
4. Gunner, M. R., Nicholls, A. & Honig, B. (1996). Electrostatic potentials in *Rhodopseudomonas viridis* reaction center: implications for the driving force and directionality of electron transfer. *J. Phys. Chem.* **100**, 4277–4291.
5. Ferguson-Miller, S. & Babcock, G. T. (1996). Heme/copper terminal oxidases. *Chem. Rev.* **96**, 2889–2907.
6. Sham, Y., Muegge, I. & Warshel, A. (1999). Simulating proton translocations in proteins: probing proton transfer pathways in the *Rhodobacter sphaeroides* reaction center. *Proteins: Struct. Funct. Genet.* **36**, 484–500.
7. Garcia-Moreno, E. B. & Fitch, C. A. (2004). Structural interpretation of pH and salt-dependent processes in proteins with computational methods. *Methods Enzymol.* **380**, 20–51.
8. Forsyth, W. R., Antosiewiez, J. M. & Robertson, A. D. (2002). Empirical relationships between protein structure and carboxyl pK_a values in proteins. *Proteins: Struct. Funct. Genet.* **48**, 388–403.
9. Fahmy, K., Weidlich, O., Engelhard, M., Sigrist, H. & Siebert, F. (1993). Aspartic acid-212 of bacteriorhodopsin is ionized in the M and N photocycle intermediates: an FTIR study on specifically ^{13}C-labeled reconstituted purple membranes. *Biochemistry*, **32**, 5862–5869.
10. Hienerwadel, R., Grzybek, S., Fogel, C., Kreutz, W., Okamura, M. Y., Paddock, M. L. & Breton, J. (1995). Protonation of Glu L212 following Q_B-formation in the photosynthetic reaction center of Rhodobacter sphaeroides: evidence from time-resolved infrared spectroscopy. *Biochemistry*, **34**, 2832–2843.
11. Mitchell, D. M., Shapleigh, J. P., Archer, A. M., Alben, J. O. & G.R., B. (1996). A pH-dependent polarity at the binuclear center of reduced cytochrome c oxidase detected by FTIR difference spectroscopy of the CO adduct. *Biochemistry*, **35**, 9446–9450.
12. Gennis, R. B. (2003). Some recent contributions of FTIR difference spectroscopy to the study of cytochrome oxidas. *FEBS Letters*, **555**, 2–7.
13. Shortle, D., Stites, W. E. & Meeker, A. K. (1990). Contributions of the large hydrophobic amino acids to the stability of staphylococcal nuclease. *Biochemistry*, **29**, 8033–8041.
14. Dwyer, J. J., Gittis, A. G., Karp, D. A., Lattman, E. E., Spencer, D. S., Stites, W. E. & Garcia-Moreno, B. (2000). High apparent dielectric constants in the interior of a protein reflect water penetration. *Biophys. J.* **79**, 1610–1620.
15. Bolen, D. W. & Santoro, M. M. (1988). Unfolding free energy changes determined by the linear extrapolation method. 2. Incorporation of $\Delta G^{\circ}_{N\text{-}U}$ values in a thermodynamic cycle. *Biochemistry*, **27**, 8069–8074.
16. Sham, Y. Y., Chu, Z. T. & Warshel, A. (1997). Consistent calculations of pK_a's of ionizable residues in proteins: Semi-microscopic and microscopic approaches. *J. Phys. Chem.* **101**, 4458–4472.
17. Hassan, S. A. & Mehler, E. L. (2001). A general screened coulomb potential based implicit solvent model: calculation of secondary structure of small peptides. *Int. J. Quant. Chem.* **83**, 193–202.

18. Simonson, T. (2001). Macromolecular electrostatics: continuum models and their growing pains. *Curr. Opin. Struct. Biol.* **11**, 243–252.
19. Georgescu, R. E., Alexov, E. G. & Gunner, M. R. (2002). Combining conformational flexibility and continuum electrostatics for calculating pK_a's in proteins. *Biophys. J.* **83**, 1731–1748.
20. Bashford, D. (2004). Macroscopic electrostatic models for protonation states in proteins. *Front Biosci.* **9**, 1082–1099.
21. Warshel, A. & Russell, S. T. (1984). Calculations of electrostatic interactions in biological systems and in solutions. *Quant. Rev. Biophys.* **17**, 283–422.
22. Honig, B. & Yang, A.-S. (1995). Free energy balance in protein folding. *Advan. Protein Chem.* **46**, 27–58.
23. Rashin, A. A. & Honig, B. (1984). On the environment of ionizable groups in globular proteins. *J. Mol. Biol.* **173**, 515–521.
24. Barlow, D. J. & Thornton, J. M. (1986). The distribution of charged groups in proteins. *Biopolymers*, **25**, 1717–1733.
25. Kajander, T., Kahn, P. C., Passila, S. H., Cohen, D. C., Lehtio, L., Adolfsen, W. *et al*. (2000). Buried charged surface in proteins. *Struct. Fold Des.* **8**, 1203–1214.
26. Kumar, S. & Nussinov, R. (1999). Salt bridge stability in monomeric proteins. *J. Mol. Biol.* **293**, 1241–1255.
27. Gunner, M. R., Saleh, M. A., Cross, E., ud-Doula, A. & Wise, M. (2000). Backbone dipoles generate positive potentials in all proteins: origins and implications of the effect. *Biophys. J.* **78**, 1126–1144.
28. Wimley, W. C., Gawrisch, K., Creamer, T. P. & White, S. H. (1996). Direct measurement of salt-bridge solvation energies using a peptide model system: implications for protein stability. *Proc. Natl Acad. Sci. USA*, **93**, 2985–2990.
29. Fersht, A. R., Shi, J., Knill-Jones, J., Lowe, D. M., Wilkinson, A. J., Blow, D. M. *et al*. (1985). Hydrogen bonding and biological specificity analysed by protein engineering. *Nature*, **314**, 235–238.
30. Yang, A. & Honig, B. (1995). Free energy determinants of secondary structure formation: I. α-helices. *J. Mol. Biol.* **252**, 351–365.
31. Yang, A. & Honig, B. (1995). Free energy determinants of secondary structure formation: II. Antiparallel β-sheet. *J. Mol. Biol.* **252**, 366–376.
32. Wolfenden, R. (1983). Waterlogged molecules. *Science*, **222**, 1087–1093.
33. Wolfenden, R. & Radzicka, A. (1994). On the probability of finding a water molecule in a nonpolar cavity. *Science*, **265**, 936–937.
34. Gunner, M. R. & Alexov, E. (2000). A pragmatic approach to structure based calculation of coupled proton and electron transfer in proteins. *Biochim. Biophys. Acta*, **1458**, 63–87.
35. Honig, B. H. & Hubble, W. L. (1984). Stablilty of "salt bridges' in membrane proteins. *Proc. Natl Acad. Sci. USA*, **81**, 5412–5416.
36. Anderson, D. E., Becktel, W. J. & Dahlquist, F. W. (1990). pH-induced denaturation of proteins: A single salt bridge contrubutes 3–5 kcal/mol to the free energy of folding of T-4 lysozyme. *Biochemistry*, **29**, 2403–2408.
37. Spassov, V. Z. & Atanasov, B. P. (1994). Spatial optimization of electrostatic interactions between the ionized groups in globular proteins. *Proteins: Struct. Funct. Genet.* **19**, 222–229.
38. Waldburger, C. D., Schildbach, J. F. & Sauer, R. T. (1995). Are buried salt-bridges imporant for protein stability and conformational specificity? *Nature Struct. Biol.* **2**, 122–128.
39. Novotny, J. & Sharp, K. (1992). Electrostatic fields in antibodies and antibody/antigen complexes. *Prog. Biophys. Molec. Biol.* **58**, 203–224.
40. Hendsch, Z. & Tidor, B. (1994). Do salt bridges stabilize proteins? A continuum electrostatic analysis. *Protein Sci.* **3**, 211–226.
41. Elcock, A. H. & McCammon, J. A. (1998). Electrostatic contributions to the stability of halophilic proteins. *J. Mol. Biol.* **280**, 731–748.
42. Sheinerman, F. B., Norel, R. & Honig, B. (2000). Electrostatic aspects of protein–protein interactions. *Curr. Opin. Struct. Biol.* **10**, 153–159.
43. Lee, L. P. & Tidor, B. (2001). Optimization of binding electrostatics: charge complementarity in the barnase–barstar protein complex. *Protein Sci.* **10**, 362–377.
44. Giletto, A. & Pace, C. N. (1999). Buried, charged, non-ion-paired aspartic acid 76 contributes favorably to the conformational stability of ribonuclease T1. *Biochemistry*, **38**, 13379–13384.
45. Karshikoff, A. & Ladenstein, R. (2001). Ion pairs and the thermotolerance of proteins from hyperthermophiles: a "traffic rule" for hot roads. *TIBS*, **26**, 550–556.
46. Zhong, W., Gallivan, J. P., Zhang, Y., Li, L., Lester, H. A. & Dougherty, D. A. (1998). From ab initio quantum mechanics to molecular neurobiology: A cation–π binding site in the nicotinic receptor. *PNAS*, **95**, 12088–12093.
47. Gao, J. & Xia, X. (1992). A priori evaluation of aqueous polarization effects through Monte Carlo QM-MM simulations. *Science*, **258**, 631–635.
48. Murphy, R. B., Philipp, D. M. & Friesner, R. A. (2000). A mixed quantum mechanics/molecular mechanics (QM/MM) method for large-scale modeling of chemistry in protein environments. *J. Comput. Chem.* **21**, 1442–1457.
49. Simonson, T., Carlsson, J. & Case, D. A. (2004). Proton binding to proteins: pK_a calculations with explicit and implicit solvent models. *J. Am. Chem. Soc.* **126**, 4167–4180.
50. Mongan, J., Case, D. A. & McCammon, J. A. (2004). Constant pH molecular dynamics in generalized Born implicit solvent. *J. Comput. Chem.* **25**, 2038–2048.
51. Kuhn, B., Kollman, P. A. & Stahl, M. (2004). Prediction of pK_a shifts in proteins using a combination of molecular mechanical and continuum solvent calculations. *J. Comput. Chem.* **25**, 1865–1872.
52. Beroza, P., Fredkin, D. R., Okamura, M. Y. & Feher, G. (1991). Protonation of interacting residues in a protein by a Monte Carlo method: application to Lysozyme and the photosynthetic reaction center of *Rhodobacter sphaeroides*. *Proc. Natl Acad. Sci. USA*, **88**, 5804–5808.
53. Yang, A.-S., Gunner, M. R., Sampogna, R., Sharp, K. & Honig, B. (1993). On the calculation of pK_a's in proteins. *Proteins: Struct. Funct. Genet.* **15**, 252–265.
54. Bashford, D. & Karplus, M. (1990). The pK_a's of ionizable groups in proteins: atomic detail from a continuum electrostatic model. *Biochemistry*, **29**, 10219–10225.
55. Antosiewicz, J., McCammon, J. A. & Gilson, M. K. (1994). Prediction of pH-dependent properties in proteins. *J. Mol. Biol.* **238**, 415–436.
56. Schutz, C. N. & Warshel, A. (2001). What are the

"dielectric constants" of proteins and how to validate electrostatic models? *Proteins: Struct. Funct. Genet.* **44**, 400–417.
57. Mehler, E. L. (1996). Self-consistent, free energy based approximation to calulate pH dependent electrostatic effects in proteins. *J. Phys. Chem.* **100**, 16006–16018.
58. Mehler, E. L. & Guarnieri, F. (1999). A self-consistent, microenvironment modulated screened Coulomb potential approximation to calculate pH-dependent electrostatic effects in proteins. *Biophys. J.* **77**, 3–22.
59. Demchuk, E. & Wade, R. C. (1996). Improving the continuum dielectric approach to calculating pK_a's of ionizable groups in protein. *J. Phys. Chem.* **100**, 17373–17387.
60. Krishtalik, L. I., Kuznetsov, A. M. & Mertz, E. L. (1997). Electrostatics for proteins: description in terms of two dielectric constants simultaneously. *Proteins: Struct. Funct. Genet.* **28**, 174–182.
61. Voges, D. & Karshikoff, A. (1998). A model of a local dielectric constant in proteins. *J.Chem.Phys.* **108**, 2219–2227.
62. Rocchia, W., Alexov, E. & Honig, B. (2001). Extending the applicability of the nonlinear Poisson-Boltzmann equation: multiple dielectric constants and multivalent ions. *J. Phys. Chem. B*, **105**, 6507–6514.
63. Ripoll, D. R., Vorobjev, Y. N., Liwo, A., Vila, J. A. & Scheraga, H. A. (1996). Coupling between folding and ionization equilibria: Effects of pH on the conformational preferences of polypeptides. *J. Mol. Biol.* **264**, 770–783.
64. Van Vlijmen, H. W. T., Schaefer, M. & Karplus, M. (1998). Improving the accuracy of protein pK_a calculations: conformational averaging versus the average structure. *Proteins: Struct. Funct. Genet.* **33**, 145–158.
65. Beroza, P. & Case, D. (1996). Including side chain flexibility in continuum electrostatic calculations of protein titration. *J. Phys. Chem.* **100**, 20156–20163.
66. Alexov, E. G. & Gunner, M. R. (1997). Incorporating protein conformational flexibility into the calculation of pH-dependent protein properties. *Biophys. J.* **72**, 2075–2093.
67. You, T. J. & Bashford, D. (1995). Conformation and hydrogen ion titration of proteins: a continuum electrostatic model with conformational flexibility. *Biophys. J.* **69**, 1721–1733.
68. Nielsen, J. E. & Vriend, G. (2001). Optimizing the hydrogen-bond network in Poisson-Boltzmann equation-based pK_a calculations. *Proteins: Struct. Funct. Genet.* **43**, 403–412.
69. Koumanov, A., Ruterjans, H. & Karshikoff, A. (2002). Continuum electrostatic analysis of irregular ionization and proton allocation in proteins. *Proteins: Struct. Funct. Genet.* **46**, 85–96.
70. Murzin, A. G., Brenner, S. E., Hubbard, T. & Chothia, C. (1995). SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.* **247**, 536–540.
71. Chakravatry, S. & Varadarajan, R. (1999). Residue depth: a novel parameter for the analysis of protein structure and stability. *Structure*, **7**, 723–732.
72. Richards, F. M. (1977). Areas, volumes, packing, and protein structure. *Annu. Rev. Biophys. Bioengng.* **6**, 151–176.
73. Richmond, T. J. (1984). Solvent accessible surface area and excluded volume in proteins. Analytical equations for overlapping spheres and implications for the hydrophobic effect. *J. Mol. Biol.* **178**, 63–89.
74. Richmond, T. J. & Richards, F. M. (1978). Packing of a-helices: geometrical constraints and contact areas. *J. Mol. Biol.* **119**, 537–555.
75. Nicholls, A. & Honig, B. (1991). A rapid finite difference algorithm utilizing successive over-relaxation to solve the Poisson-Boltzmann equation. *J. Comput. Chem.* **12**, 435–445.
76. Sander, C. & Schneider, R. (1991). Database of homology-derived protein structures and the structural meaning of sequence alignment. *Proteins: Struct. Funct. Genet.* **9**, 56–68.
77. Robertson, A. D. (2002). Intramolecular interactions at protein surfaces and thier impact on protein function. *TIBS*, **27**, 521–526.
78. Mao, J., Hauser, K. & Gunner, M. R. (2003). How cytochromes with different folds control heme redox potentials. *Biochemistry*, **42**, 9829–9840.
79. Alexov, E. G. & Gunner, M. R. (1999). Calculated protein and proton motions coupled to electron transfer: electron transfer from Q_A^- to Q_B in bacterial photosynthetic reaction centers. *Biochemistry*, **38**, 8253–8270.
80. Rabenstein, B., Ullmann, G. M. & Knapp, E. W. (2000). Electron transfer between the quinones in the photosynthetic reaction center and its coupling to conformational changes. *Biochemistry*, **39**, 10487–10496.
81. Zhu, Z. & Gunner, M. R. (2005). Energetics of quinone-dependent electron and proton transfers in *Rhodobacter sphaeroides* photosynthetic reaction centers. *Biochemistry*, **44**, 82–96.
82. Spassov, V. Z., Luecke, H., Gerwert, K. & Bashford, D. (2001). pK_a alculations suggest storage of an excess proton in a hydrogen-bonded water network in bacteriorhodopsin. *J. Mol. Biol.* **312**, 203–219.
83. Song, Y., Mao, J. & Gunner, M. R. (2003). Calculation of proton transfers in Bacteriorhodopsin bR and M intermediates. *Biochemistry*, **42**, 9875–9888.
84. Song, Y. & Gunner, M. R. (2005). Halorhodopsin and bacteriorhodopsin: comparing the mechanism of a Cl and a proton pump. In the press.
85. Haas, A. H. & Lancaster, C. R. (2004). Calculated coupling of transmembrane electron and proton transfer in dihemic quinol:fumarate reductase. *Biophys. J.* **87**, 4298–4315.
86. Spassov, V., Karshokoff, A. & Ladenstein, R. (1994). Optimization of the electrostatic interactions between ionized groups and peptide dipoles in proteins. *Protein Sci.* **3**, 1556–1569.
87. Gandini, D., Gogioso, L., Bolognesi, M. & Bordo, D. (1996). Patterns in ionizable side chain interactions in protein structures. *Proteins: Struct. Funct. Genet.* **24**, 439–449.
88. Spassov, V. Z., Ladenstein, R. & Karshikoff, A. D. (1997). Optimization of the electrostatic interactions between ionized groups and peptide dipoles in proteins. *Protein Sci.* **6**, 1190–1196.
89. Spector, S., Wang, M., Carp, S. A., Robblee, J., Hendsch, Z., Fairman, R. *et al*. (2000). Rational modification of protein stability by the mutation of charged surface residues. *Biochemistry*, **39**, 872–879.
90. Eswar, N. & Ramakrishnan, C. (2000). Deterministic features of side-chain main-chain hydrogen bonds in globular protein structures. *Protein Engng.* **13**, 227–238.
91. McDonald, I. K. & Thorton, J. M. (1994). Satisfying hydrogen bonding potential in proteins. *J. Mol. Biol.* **238**, 777–793.

92. Pace, C. N. (2000). Polar group burial contributes more to protein stability than nonpolar group burial. *J. Am. Chem. Soc.* **39**, A–D.
93. Edgcomb, S. P. & Murphy, K. P. (2002). Variability in the pK_a of histidine side-chains correlates with burial within proteins. *Proteins: Struct. Funct. Genet.* **49**, 1–6.
94. Elcock, A. H. (2001). Prediction of functionally important residues based solely on the computed energetics of protein structure. *J. Mol. Biol.* **312**, 885–896.
95. Bate, P. & Warwicker, J. (2004). Enzyme/non-enzyme discrimination and prediction of enzyme active site location using charge-based methods. *J. Mol. Biol.* **340**, 263–276.
96. Rabenstein, B., Ullmann, G. M. & Knapp, E.-W. (1998). Energetics of electron-transfer and protonation reactions of the quinones in the photosynthetic reaction center of *Rhodopseudonas viridis*. *Biochemistry*, **37**, 2488–2495.
97. Onufriev, A., Case, D. A. & Ullmann, G. M. (2001). A novel view of pH titration in biomolecules. *Biochemistry*, **40**, 3413–3419.
98. Ondrechen, M., Clifton, J. & Ringe, D. (2001). THEMATICS: a simple computational predictor of enzyme function from structure. *Proc. Natl Acad. Sci. USA*, **98**, 12473–12478.
99. Havranek, J. J. & Harbury, P. B. (1999). Tanford-Kirkwood electrostatics for protein modeling. *Proc. Natl Acad. Sci. USA*, **96**, 11145–11150.
100. Antosiewicz, J., McCammon, J. A. & Gilson, M. K. (1996). The determinants of pK_a's in proteins. *Biochemistry*, **35**, 7819–7833.
101. Wang, G. & Dunbrack, R. L. (2003). PISCES: a protein sequence culling server. *Bioinformatics*, **19**, 1589–1591.
102. Henrick, K. & Thornton, J. M. (1998). PQS: a protein quaternary structure file server. *Trends Biochem. Sci.* **23**, 358–361.
103. Nicholls, A., Sharp, K. A. & Honig, B. (1991). Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbonds. *Proteins: Struct. Funct. Genet.* **11**, 281–296.
104. Bharadwaj, R., Windemuth, A., Sridharan, S., Honig, B. & Nicholls, A. (1995). The fast multipole boundary element method for molecular electrostatics: an optimal approach for large systems. *J. Comp. Chem.* **16**, 898–913.
105. Sridharan, S., Nicholls, A. & Honig, B. (1992). A new vertex algorithm to calculate solvent accessible surface areas. *Biophys. J.* **61**, A174.
106. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H. *et al*. (2000). The Protein Data Bank. *Nucl. Acids Res.* **28**, 235–242.
107. Sitkoff, D., Sharp, K. A. & Honig, B. (1994). Accurate calculation of hydration free energies using macroscopic solvent models. *J. Phys. Chem.* **98**, 1978–1988.
108. Alexov, E., Miksovska, J., Baciou, L., Schiffer, M., Hanson, D., Sebban, P. & Gunner, M. R. (2000). Modeling the effects of mutations on the free energy of the first electron transfer from Q_A^- to Q_B in photosynthetic reaction centers. *Biochemistry*, **39**, 5940–5952.
109. Gibas, C. & Subramaniam, S. (1996). Explicit solvent models in protein pK_a calculations. *Biophys. J.* **71**, 138–147.
110. Garcia-Moreno, B., Dwyer, J. J., Gittis, A. G., Lattman, E. E., Spencer, D. S. & Stites, W. E. (1997). Experimental measurement of the effective dielectric in the hydrophobic core of a protein. *Biophys. Chem.* **64**, 211–224.
111. Hauser, K., Mao, J. & Gunner, M. R. (2004). pH dependence of heme electrochemistry in cytochromes investigated by multiconformation continuum electrostatic calculations. *Biopolymers*, **74**, 51–54.

Edited by J. Thornton

(Received 12 January 2005; received in revised form 11 March 2005; accepted 17 March 2005)