The structure, stability, solubility, and function of proteins depend on their net charge and on the ionization state of the individual residues. Consequently, biochemists are interested in the pK values of the ionizable groups in proteins and how these pKs depend on their environment. We review what has been learned about pK values of ionizable groups in proteins from experimental studies, and discuss the important contributions they make to protein stability and solubility.

**Historical perspective**

Sorensen defined pH in 1909 and in 1917 published the first experimental study of the titration of a protein, egg albumin (1). In succeeding years, hydrogen ion titration curves were determined for several proteins, and it was possible to make rough estimates of the pKs of the ionizable groups of proteins (2). In special cases, it was possible to determine the pKs of individual groups, but it was only when NMR became available that the pKs of individual groups could be readily determined, at least for small proteins (3,4). This led to rapid progress and over 500 pKs have been determined for individual ionizable groups in folded proteins (5), and a more limited number in unfolded proteins (6).

The landmark paper by Debye and Hückel on the theory of electrolyte solutions was published in 1923 (7), and the ideas were extended to proteins by Linderstrom-Lang in 1924 (8). He recognized that net charge on a protein would influence the ionization of individual groups, and incorporated this into the first model developed to understand acid/base properties of proteins. This model was extended by Tanford and Kirkwood (9) in an important paper that triggered an interest in factors that determine pKs of the ionizable groups in proteins that continues to the present day (10).

**Protein ionizable groups and their intrinsic pKs**

Seven amino acid side chains contain groups that ionize between pH 1 and 14. For Asp, Glu, Tyr, and Cys, the ionizable groups are uncharged below their pK and negatively charged above their pK. For His, Lys, and Arg, the ionizable groups are positively charged below their pK, and uncharged above their pK. It is useful to know what the pKs of these groups would be in a protein if they are completely exposed to solvent, not hydrogen bonded, and not affected by the presence of any formal charges. These are generally referred to as the intrinsic pKs (pK\textsubscript{int}). The pK\textsubscript{int} values given in Table 1 are the pKs observed for the ionizable side chains when they are present in blocked pentapeptides with the structure: Ala – Ala – Xxx – Ala – Ala, where Xxx is the amino acid whose side-chain pK was measured (11). The \(\alpha\)-carboxyl and the \(\alpha\)-amino groups of proteins can also ionize and their pKs were determined in similar pentapeptides and are also given in Table 1. These pK\textsubscript{int} values reflect the inductive effects of neighboring peptide bonds, but will not be influenced by charge-charge interactions, and only minimally by hydrogen bonding or burial of the ionizable group. They should serve as good models for the unperturbed pK values of the ionizable groups in proteins.
Content and environment of the ionizable groups

Amino acids with ionizable side chains make up, on average, 29% of the amino acids in proteins (12). The average content for each is given in Table 1. As discussed below, the extent of burial of the ionizable groups in proteins is important in determining their pK values. The average % burial for the ionizable group in each side chain is given in Table 1 (13). The most buried ionizable groups are the –SH of Cys, the imidazole of His, and the –OH of Tyr. These groups are often buried because they are generally uncharged at pH 7. The least buried are the guanidinium of Arg, the carboxylates of Asp and Glu, and the amino groups of Lys. These groups will generally be charged at pH 7. It is surprising that Arg is buried to such an extent because of the high pK, and the fact that the Arg side chain can donate five hydrogen bonds. However, in water the guanidinium group is one of the most weakly hydrated cations, probably because of charge delocalization, and this makes the Arg side chain easier to bury (14). Buried Arg side chains are charged, extensively hydrogen bonded, and frequently interact by stacking with other planar side chain groups in proteins (15). They make many important contributions to the stability and function of proteins.

Measured pK values in folded proteins

Most of the pK values for ionizable groups in folded proteins were determined by measuring the pH dependence of chemical shifts using NMR (3,4). A smaller number were measured using indirect techniques (16). Recently, 541 pK values from 78 proteins were compiled and the results are summarized in Table 1 (5). Many of the pKs are perturbed far above and below the pK_int values. For example, the pK of one sulfhydryl group is lowered more than 6 pK units, and the pK of one carboxyl side chain is raised over 5 pK units.

Perturbation of the pK values

In proteins, the pKs of the ionizable groups may be substantially raised or lowered from the intrinsic pK values by environment effects (See Table 1). The three most important effects are summarized in Figure 1. Each of these will be discussed in general terms, and illustrated with experimental results. Another review discusses the perturbation of the pK values of catalytic groups in enzyme active sites (17).

Dehydration (Born effect)--It is energetically unfavorable to transfer a charged group from water to the interior of a protein where the dielectric constant, ε_protein, is lower. Consequently, the neutral state of the ionizable group will be favored and the pKs of Asp, Glu, Cys, and Tyr will be raised and the pKs of His, Lys and Arg will be lowered when the groups are buried, partially or completely, in a folded protein. To illustrate this, the pK of acetic acid is increased from 4.8 in water (ε = 78) to 10.1 in ethanol (ε = 24).

Studies of staph nuclease provide an example of this effect in proteins (18). Valine 66 is buried in the hydrophobic core of the enzyme. When it is replaced by Asp, the carboxyl group has a pK = 8.9, five pK units higher than the pK_int. When it is replaced by Lys, the amino group has a pK = 5.5, 4.9 units lower than the pK_int. If these changes resulted only from the Born effect, it would require ε_protein = 7.2. The authors concluded (18): “Regardless of how the pKc calculations were performed, they all showed that the shift in the pKc value of Asp-66 is governed by the loss of hydration of the carboxylic group in the buried state that is not offset by interactions with charges or with polar atoms of the protein.”

Charge-Charge Interactions (Coulombic interactions)--The net charge on a protein is zero at the isoelectric pH = pl. Below the pl, the net charge on a protein is positive, and, above the pl, net charge is negative. At the pK of a given ionizable group, net charge will be −½ for Asp, Glu, Tyr, and Cys, and + ½ for His, Lys, and Arg. The energy of interaction of the ionizable group, i, and the other charges, j, on the protein can be calculated with Coulomb’s law:
\[ \Delta G_{ij} = \sum q_i q_j / \epsilon r_{ij} \]

where \( q_i \) is the charge on the ionizable group of interest, \( q_j \) is the charge on the other groups at the pH = pK of the ionizing group, \( \epsilon \) is the dielectric constant, and \( r_{ij} \) is the distance between the two charges. (When opposite charges are 4.2 Å apart in water, \( \Delta G = -1 \) kcal/mol and this is reduced to – 0.5 kcal/mol at the ionic strength inside a cell.) The distance between charges can be calculated from the structure of the protein, keeping in mind that the distance between the groups may differ in solution and in a crystal and may vary as the protein is titrated. (For RNase A, structures were determined as a function of pH so that the effect of titration on the distances could be observed (19).) Protein interiors are heterogeneous so the effective value of \( \epsilon \) will depend on which two charges are considered. Values of \( \epsilon \) ranging from 2 to 80 have been used (20).

A good example of the effect of Coulombic interactions on the pKs of ionizable groups in a protein is provided by a study of ribonuclease Sa (RNase Sa) (21). RNase Sa is an acidic protein with a pI = 3.5 that contains no Lys residues (0K). By replacing Asp and Glu residues on the surface with 5 Lys residues, a basic protein was created with a pI = 10.2 (5K). At pH 7, the net charge on 0K is –7 and the net charge on 5K is +3, a difference of ~10 units. Crystal structures and NMR studies show that the structures of 0K and 5K are similar (21). Consequently, except for net charge, the ionizable groups will have similar environments in the two proteins, and Coulombic interactions will be the main determinant of pK differences. For the 11 common groups, the pKs were always higher in 5K than in 0K, as expected because of the greater positive charge. The differences ranged from 0.03 to 2.19 with an average difference of 0.75. The pK differences (pK\textsubscript{0K} – pK\textsubscript{5K}) calculated as described above with Coulomb’s law were in good agreement with the measured values. A value of \( \epsilon = 45 \) gave the best agreement between calculated and experimental values. It is not surprising that the \( \epsilon \) value that gave the best agreement here is considerably higher than the \( \epsilon \) value that gave the best results for the Born effect (22). Based on this it was concluded (21): “Taken together, the results are evidence that charge-charge interactions are the chief perturbant of the pK values of ionizable groups on the protein surface, which is where the majority of the ionizable groups are positioned in proteins.”

These conclusions are supported by other experimental and theoretical studies. A global analysis of available data for short to long range Coulombic interactions in staph nuclease led to an effective \( \epsilon \approx 42 \) (23). A theoretical analysis was successful in predicting the contribution of electrostatic effects to the stability of four proteins (22). When both the Born effect and Coulombic interactions were included in their analysis, the best agreement was found with \( \epsilon = 20 \) to 80 for the Born effect and \( \epsilon = 40 \) for Coulombic interactions. The significance of these \( \epsilon \) values, and why a large \( \epsilon \) is needed for the Born effect were discussed (22).

**Charge-Dipole interactions (Hydrogen Bonds)**--Ionizable groups can also interact with the partial charges or dipoles on neighboring polar groups. These interactions will be referred to as hydrogen bonds, but keep in mind that some of these interactions can be important and not meet the definitions generally used for hydrogen bonds. The effect of hydrogen bonding on a pK will depend on whether the interactions are more favorable with the protonated state of the group in which case the pK will be raised, or with the deprotonated state of the group in which case the pK will be lowered. Hydrogen bonds generally contribute 1 to 2 kcal/mol to the stability of a protein (24), and, when the hydrogen bonds are to ionizable groups, they can raise or lower the pKs by several pK units (25). An example is provided by the buried, charged, non-ion-paired, carboxyl group of Asp 76 in RNase T1(16).

The side chain carboxyl of Asp 76 in RNase T1 has a very low pK = 0.6 and forms three intramolecular hydrogen bonds to the side chains of Asn 9, Tyr 11, and Thr 91(16). To see if these hydrogen bonds were responsible for the low pK, the hydrogen bonds were removed one at a time and the pK of Asp 76 measured. When
single hydrogen bonds are removed, the average pK increases to 3.3. When two hydrogen bonds are removed, the average pK increases to 5.1. When all three hydrogen bonds are removed, the pK increases to 6.4. Thus, in the absence of the hydrogen bonds, the pK is elevated 2.5 units above the pK_{int}. This increase results from the Born effect because the carboxyl group of Asp 76 is buried and from the negative charge on the protein at pH 6.4. Hydrogen bonding and the positive charge on the protein at low pH combine to lower the pK of this carboxyl group by 5.8 pK units. A comprehensive description of the pK shifts due to hydrogen bonding is available (25). The pK shift per hydrogen bond can be as high as 1.6 for carboxyl groups and higher for sulfhydryl groups.

**Contribution of ionizable residues to protein stability**

Major forces favoring folding of globular proteins are the hydrophobic effect and hydrogen bonding, and they are just able to overcome the major force favoring unfolding, conformational entropy, so that most globular proteins have a surprisingly low conformational stability of just 2 to 10 kcal/mol (26). It is now clear that many proteins are unfolded under physiological conditions (27), but fold when it is required for their function. Because of this, minor forces in the 1 to 3 kcal/mol range become important.

**Charge – Charge Interactions**—Charged groups in proteins are generally arranged so that Coulombic interactions among charges are favorable. However, the arrangement is more favorable on some proteins than on others, and is sometimes unfavorable (28). Studies of the pH dependence of protein stability show that Coulombic interactions do not make a large contribution to protein stability, probably at most 10 kcal/mol (29). Despite this, Coulombic interactions are important to proteins in a number of ways.

There is considerable interest in developing methods to increase protein stability by amino acid substitutions. This is difficult to do by burying hydrophobic surface or adding hydrogen bonds, but can be done by improving the charge – charge interactions on the surface of a protein (30). Several groups showed that reversing the charge on a single side chain on the protein surface can improve Coulombic interactions and increase stability by over 1 kcal/mol (30-34). This approach was used to increase the stability of several proteins and guidelines are available for doing so with other proteins (30). (Another good approach is to improve the β-turns on the surface of a protein (35).)

Just as attractive charge-charge interactions can stabilize a protein, repulsive charge-charge interactions can destabilize a protein. Recently, it has become clear that many proteins are unfolded or have regions of the polypeptide chain that are disordered under physiological conditions (27). These proteins are referred to as intrinsically disordered proteins (IDPs) and the number identified is now over 500 (36). This revelation was surprising, but, on reflection, IDPs offer new functions and can improve some known functions of proteins (37). Two factors that are important in determining whether a protein will be folded or unfolded are hydrophobicity and net charge (38). IDPs generally have a low hydrophobicity, a high net charge, or both. The border between folded proteins and unfolded proteins is defined well by the equation: 

\[ \langle R \rangle = 2.785 \langle H \rangle - 1.151 \]

where \( \langle R \rangle \) and \( \langle H \rangle \) are the mean net charge and the mean hydrophobicity of the protein, respectively (38). Recent experimental studies show that favorable and unfavorable Coulombic interactions can influence the denatured state ensemble of a protein and influence both protein stability and the mechanism of folding (39).

**Buried Ionizable Residues**—As discussed above, many ionizable groups are buried and these often make important contributions to the function and stability of proteins. An improved version of a finite difference Poisson-Boltzmann method was used to estimate the number of buried ionizable residues (40). For each ionizable group, the dehydration penalty was calculated. If it was great enough to shift the pK value by > 5 pK units, the residue was classed as
buried. These residues would generally be > 80% buried when measured by accessible surface area as in Table 1. Using this criterion, 32 % of the Arg residues are buried, 19 % of the Asp, 13 % of the Glu, and 6% of the Lys. This amounts to 4 buried residues per 100 residues, and the composition is 41% Arg, 28% Asp, 22% Glu, and 9% Lys.

These buried ionizable residues can be used to stabilize or destabilize proteins. Despite being buried, the pK of Asp 76 in RNase T1 is lowered to 0.6 by hydrogen bonding and the positive charge on the protein (16). This buried carboxyl group makes a large contribution to the stability and to the pH dependence of the stability. The D76A mutation of RNase T1 is 3.8 kcal/mol less stable than wild type, and an analysis suggests that the hydrogen bonding and other interactions of the carboxyl group with the protein contribute ~ 8 kcal/mol to the stability. As discussed below, Asp 70 in T4 lysozyme is buried, has a pK = 0.5, and also makes a large contribution to the stability (41).

The two most buried carboxyl groups in RNase Sa are Asp 33 which is 99% buried and has a pK = 2.4 and Asp 79 which is 85% buried and has a pK = 7.4 (16,42). The D33A mutant is 6 kcal/mol less stable than the wild-type protein. The environment of the two carboxyls is different: Asp 33 forms three intramolecular hydrogen bonds and Asp 79 forms none. The net charge is +7 when Asp 33 ionizes which would lower the pK and −6 when Asp 79 ionizes which would raise the pK. These environmental factors combine to give a pK difference of 5 units and markedly different contributions to the stability for these two carboxyl groups. The D79F mutant of RNase Sa has a stability 3.7 kcal/mol greater than the wild-type enzyme and this raises the Tm 10°C (42). This is one of the largest increases in stability observed for a single mutation.

Based on what we have learned about the major forces that contribute to the stability of globular proteins, stability should increase as protein size increases. This is not observed: the conformational stability of globular proteins is independent of size. It has been shown that the number of buried charged groups increases substantially with protein size (40,43). For example, the number of buried charged groups is 1.9 per 100 residues in proteins with < 100 residues and 4.5 per 100 residues in proteins with > 300 residues (40). It seems likely that burial of charged side chains that are not hydrogen bonded or ion paired is one mechanism that evolution uses to lower protein stability (43).

Ion Pairs—Attractive interactions between nearby (< 5 Å) oppositely charged groups are called ion pairs or salt bridges. Coulombic interactions in proteins are made favorable by avoiding repulsive charge-charge interactions: on average there are 4 attractive ion pairs but only 1 repulsive ion pair per 100 residues (28). Whether the attractive ion pairs contribute favorably to protein stability is a controversial topic, experimentally (41,44) and theoretically (45).

An experimental study of a salt bridge between Asp 70 and His 31 in T4 lysozyme showed that a buried salt bridge can make a favorable contribution to protein stability (41). Asp 70 forms a salt bridge with His 31 and this substantially perturbs the pK values: Asp 70 has a pK = 0.5 in the folded protein and 3.5 to 4 in the unfolded protein, while His 31 has a pK = 9.1 in the folded protein and 6.8 in the unfolded protein. This salt bridge contributes ~ 4 kcal/mol to the stability of T4 lysozyme.

Proteins from thermophiles are stabilized in different ways (46), but improving charge-charge interactions and adding ion pairs is the strategy used most often (46,47). Thermophilic proteins generally contain more charged groups and salt bridges than proteins from mesophiles, and several lines of evidence suggest that the contribution of salt bridges to protein stability will increase at higher temperatures (47).

The most stable protein known is the CutA1 protein from the hyperthermophile Pyrococcus horikoshii (Ph-CutA1), with a Tm of ~ 150°C at pH 7 (48). This protein is found in bacteria, plants, and animals, including humans. All of the CutA1 proteins studied are remarkably
stable and this is thought to be due in part to a common trimeric structure. The most striking difference between Ph-CutA1 and the same protein from *E. coli* is the number of ion pairs in the monomer: 30 are found in the protein from the hyperthermophile, but only 1 in the protein from the mesophile. This suggests that ion pairs are of crucial importance to the stability of the most stable proteins.

**Contribution of ionizable residues to protein solubility**

Protein solubility is a concern to biochemists in experimental studies, and the recognition of its role in protein folding diseases makes it even more important. Because many proteins are now used as drugs, solubility is a concern in the biopharmaceutical industry. Key principles of protein solubility were summarized by Cohn in 1943 (49):

“A given protein is least soluble in the neighborhood of its isoelectric point in the presence, as well as the absence, of neutral salts. ∙∙∙ The solubility of proteins in the uncombined, salt-free state varies widely. This is true among those that separate in the crystalline state and in the amorphous state. ∙∙∙ The forces between molecules in the solid state as well as those between solvent and solute molecules determine solubility.”

The approach generally used to increase solubility of a protein is to replace the most hydrophobic residue on the surface with a charged or polar residue. Recent studies provide insights into protein solubility (50). Thr 76 has the most exposed side chain in RNase Sa. This Thr was replaced with the other 19 amino acids and the solubility measured. The most soluble variant was T76D (43 mg/ml) and least soluble was T76W (3.6 mg/ml), a 12-fold difference! One surprising finding was that His, Asn, Thr, and Glu make unfavorable contributions to the solubility relative to Ala near the pI of RNase Sa. Similarly, Arg and Lys make unfavorable contributions to the solubility relative to Ala when solubility is measured at pH 7 where the protein has a charge of \( \approx -5 \). In contrast, Asp and Ser always make favorable contributions to the solubility. These results suggest that the best approach for increasing solubility is to replace the most hydrophobic residue on the protein’s surface with an Asp or Ser (50).
REFERENCES

8. Linderstrom-Lang, K. U. (1924) *CR Trav Lab Carlsb* 15, 1-29
### TABLE 1
**Characteristics of ionizable side chains in proteins**

Summary of 541 pK values tabulated from the literature (5). The values were reported under various conditions for 78 folded proteins.

<table>
<thead>
<tr>
<th>Group</th>
<th>Content (%)</th>
<th>Buried (%)</th>
<th>pK value in alanine pentapeptides (pK_{int})</th>
<th>Average pK value</th>
<th>Low pK value</th>
<th>High pK value</th>
<th>Number of measurements</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asp</td>
<td>5.2</td>
<td>56</td>
<td>3.9</td>
<td>3.5 ± 1.2</td>
<td>0.5</td>
<td>9.2</td>
<td>139</td>
</tr>
<tr>
<td>Glu</td>
<td>6.5</td>
<td>48</td>
<td>4.3</td>
<td>4.2 ± 0.9</td>
<td>2.1</td>
<td>8.8</td>
<td>153</td>
</tr>
<tr>
<td>His</td>
<td>2.2</td>
<td>72</td>
<td>6.5</td>
<td>6.6 ± 1.0</td>
<td>2.4</td>
<td>9.2</td>
<td>131</td>
</tr>
<tr>
<td>Cys</td>
<td>1.2</td>
<td>90</td>
<td>8.6</td>
<td>6.8 ± 2.7</td>
<td>2.5</td>
<td>11.1</td>
<td>25</td>
</tr>
<tr>
<td>Tyr</td>
<td>3.2</td>
<td>67</td>
<td>9.8</td>
<td>10.3 ± 1.2</td>
<td>6.1</td>
<td>12.1</td>
<td>20</td>
</tr>
<tr>
<td>Lys</td>
<td>5.9</td>
<td>34</td>
<td>10.4</td>
<td>10.5 ± 1.1</td>
<td>5.7</td>
<td>12.1</td>
<td>35</td>
</tr>
<tr>
<td>Arg</td>
<td>5.1</td>
<td>56</td>
<td>12.3(^d)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>C-term</td>
<td>-</td>
<td>-</td>
<td>3.7</td>
<td>3.3 ± 0.8</td>
<td>2.4</td>
<td>5.9</td>
<td>22</td>
</tr>
<tr>
<td>N-term</td>
<td>-</td>
<td>-</td>
<td>8.0</td>
<td>7.7 ± 0.5</td>
<td>6.8</td>
<td>9.1</td>
<td>16</td>
</tr>
</tbody>
</table>

\(^a\) This is the current average % amino acid content of proteins from all three domains of life: Bacteria, Archaea, and Eukaryota (12). For comparison, the most abundant amino acid is Leu, 10.3%, and the least abundant is Trp, 1.1%.

\(^b\) The % buried is based on an analysis of 61 proteins by Lesser and Rose (13). The % buried is for just the ionizable group: the carboxyl of Asp and Glu, the imidazole of His, the-SH of Cys, the –OH of Tyr, the amino of Lys, and the guanidinium of Arg.

\(^c\) pK values are from (11) and the value for Asp was corrected in (5).

\(^d\) A value of 12.48 is given for Arg in (49). A correction for the negative charge on the carboxyl group present in Arg gives a pK \(\approx 12.3\).
FIGURE LEGEND

Fig. 1. Factors influencing the pK values of ionizable groups in proteins. A. A pK change due to
the Born effect results when an ionizable group is buried in the interior of the protein where the
dielectric constant is lower than water. The lower dielectric constant favors the neutral form of the ionizable group.
B. The pKs of all of the ionizable groups in a protein will be decreased by a positively charged
environment and increased by a negatively charged environment. C. The pKs of the ionizable groups will
be increased when hydrogen bonding is tighter to the protonated form and decreased when hydrogen
bonding is tighter to the deprotonated form.
A. Dehydration (Born Effect):

\[-\text{COOH} \leftrightarrow -\text{COO}^- + \text{H}^+ \quad \uparrow \text{nonpolar environment: } \text{pK}_a \uparrow\]

\[-\text{NH}_3^+ \leftrightarrow -\text{NH}_2 + \text{H}^+ \quad \uparrow \text{nonpolar environment: } \text{pK}_a \downarrow\]

B. Charge-Charge Interactions (Coulombic):

\[-\text{COOH} \leftrightarrow -\text{COO}^- + \text{H}^+ \quad \uparrow \text{positive charge: } \text{pK}_a \downarrow\]

\[-\text{NH}_3^+ \leftrightarrow -\text{NH}_2 + \text{H}^+ \quad \uparrow \text{positive charge } \text{pK}_a \downarrow\]

C. Charge-Dipole Interactions (Hydrogen Bonding):

\[-\text{COOH} \leftrightarrow -\text{COO}^- + \text{H}^+ \quad \uparrow \text{hydrogen bonding to protonated form: } \text{pK}_a \uparrow\]

\[-\text{NH}_3^+ \leftrightarrow -\text{NH}_2 + \text{H}^+ \quad \uparrow \text{hydrogen bonding to protonated form: } \text{pK}_a \uparrow\]
Protein ionizable groups: pK values and their contribution to protein stability and solubility
C. Nick Pace, Gerald R. Grimsley and J. Martin Scholtz

J. Biol. Chem. published online January 21, 2009

Access the most updated version of this article at doi: 10.1074/jbc.R800080200

Alerts:
• When this article is cited
• When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 0 references, 0 of which can be accessed free at
http://www.jbc.org/content/early/2009/01/21/jbc.R800080200.citation.full.html#ref-list-1