Arginine residues at internal positions in a protein are always charged

Michael J. Harms, Jamie L. Schlessman, Gloria R. Sue, and Bertrand Garcia-Moreno

Department of Biophysics, Johns Hopkins University, 3400 North Charles Street, Baltimore, MD 21218; and Chemistry Department, United States Naval Academy, 572 Holloway Road, Annapolis, MD 21402

Edited by Roderick MacKinnon, The Rockefeller University, New York, NY, and approved September 13, 2011 (received for review March 25, 2011)

Many functionally essential ionizable groups are buried in the hydrophobic interior of proteins. A systematic study of Lys, Asp, and Glu residues at 25 internal positions in staphylococcal nuclease showed that their pK\textsubscript{a} values can be highly anomalous, some shifted by as many as 5.7 pH units relative to normal pK\textsubscript{a} values in water. Here we show that, in contrast, Arg residues at the same internal positions exhibit no detectable shifts in pK\textsubscript{a}; they are all charged at pH ≤ 10. Twenty-three of these 25 variants with Arg are folded at both pH 7 and 10. The mean decrease in thermodynamic stability from substitution with Arg was \(6 \pm 2\) kcal/mol at this pH, comparable to that for substitution with Lys, Asp, or Glu at pH 7. The physical basis behind the remarkable ability of Arg residues to remain protonated in environments otherwise incompatible with charges is suggested by crystal structures of three variants showing how the guanidinium moiety of the Arg side chain is effectively neutralized through multiple hydrogen bonds to protein polar atoms and to site-bound water molecules. The length of the Arg side chain, and slight deformations of the protein, facilitate placement of the guanidinium moieties near polar groups or bulk water. This unique capacity of Arg side chains to retain their charge in dehydrated environments likely contributes toward the important functional roles of internal Arg residues in situations where a charge is needed in the interior of a protein, in a lipid bilayer, or in similarly hydrophobic environments.

Ionizable groups are usually found at the protein–water interface, where their charged moieties can interact with water. However, some ionizable groups are also found buried in the relatively hydrophobic interior of proteins. These ionizable groups are usually essential for function. Arginine (Arg) is the ionizable amino acid that is found most frequently buried in the protein interior (1–3). Arg residues in relatively hydrophobic environments are essential in a variety of biological processes, such as regulation of conformation or redox potentials (4, 5), viral capsid assembly (6), electrostatic steering (7), voltage sensing across the interiors of channels (8, 9), and peptide translocation across bilayers (10, 11), and peptide translocation across bilayers (12, 13), and peptide translocation across bilayers (14, 15). Internal Arg residues also play critical roles at protein–protein interfaces (3, 16), in enzyme active sites (1, 3, 17), and in a variety of transport channels (18, 19).

To understand the functional roles of internal Arg residues, it is necessary to know their charge state throughout their functional cycle. The low polarizability and polarity of the protein interior can lead to large shifts in the pK\textsubscript{a} values of internal groups relative to their normal values in water (20–26), usually in the direction that promotes the neutral state. Arg in water has the highest pK\textsubscript{a} among naturally occurring ionizable amino acid (27), making Arg the ionizable residue least likely to exist in the neutral state near physiological pH. In the proposed functional roles for Arg residues buried in hydrophobic environments, the side chain is usually assumed to be charged, but this is seldom demonstrated experimentally (2, 8–10, 12–15, 18, 19, 28–32). Computational studies indicating that the Arg side chain is rarely deprotonated support this conclusion (2, 15, 28, 29, 33). To our knowledge, only two experimental studies have examined explicitly the charge state of an internal Arg residue at physiological pH (34, 35). Both studies investigated internal Arg in channel proteins; both studies found that the Arg was neutral.

To study the properties of Arg residues in the protein interior systematically, 25 internal positions in a highly stable form of staphylococcal nuclease (SNase) known as Δ+PHS SNase were substituted with Arg (Fig. 1). To measure the pK\textsubscript{a} values and to characterize the charge state of the internal Arg residues, the pH dependence of thermodynamic stability of the 25 variants with internal Arg was measured (20, 21, 36). Using this approach, we previously demonstrated that Lys, Asp, and Glu at these 25 internal positions can have highly anomalous pK\textsubscript{a} values (20, 22, 26, 37–39). Of the 75 Lys, Asp, and Glu studied, 64 titrates with pK\textsubscript{a} values shifted relative to their normal values in water, some by as much as 5.7 pH units. The shifts are all in the direction that favors the neutral state (elevated pK\textsubscript{a} for acidic residues and depressed pK\textsubscript{a} for basic residues). This suggests that the ionizable moieties of these residues are at least partially dehydrated and that the polarity or polarizability in the protein interior is not sufficient to compensate fully for dehydration effects. A large number of crystal structures of other variants of SNase (already available in the PDB), and Trp fluorescence and CD and NMR spectroscopy studies show that most variants of SNase with internal ionizable groups are fully folded and native-like at pH values where the internal groups are neutral (21, 22, 26). Even upon ionization of the internal groups, most of the 75 proteins with internal Lys, Asp, and Glu remain folded; only about a third of them exhibit any evidence of conformational reorganization, minor in most cases, coupled to the ionization of the internal group (21, 22). In the current work we demonstrate that the electrostatic properties of Arg at these 25 internal positions in SNase are strikingly different from those of Lys, Glu, or Asp.

Results

Structural Consequences of Substitution with Arg. To examine how substitution of internal residues with Arg affected the conformational properties of the protein globally, far-UV CD spectra were measured for each variant at pH 7 (Fig. 24) and pH 10 (Fig. 2B). Of the 25 variants with Arg, 19 had spectra comparable to those of the parent protein (Δ+PHS nuclease), suggesting that these proteins were folded without alterations in secondary structure. The spectra of four variants (A58R, V39R, Y91R, and N100R) were slightly different from those of the parent protein. The A58R variant showed a decrease in signal intensity relative to the signal of...
Δ+PHS at both pH values, consistent with partial unfolding and loss of secondary structure or with the presence of some unfolded protein in equilibrium with fully folded protein. The V39R, Y91R, and N100R variants exhibited an increase in signal at 208 nm at pH 7 and a shift in spectral minimum at pH 10, consistent with a minor change in secondary structure relative to the reference protein.

Two other variants (I92R and V104R) showed much larger differences in both signal intensity and shape at both pH 7 and 10, relative to the reference protein. For comparison, the spectrum of an intrinsically unfolded T62P variant of SNase is included in Fig. 2. The I92R and V104R variants have deep 208-nm and shallow 222-nm minima at pH 7 and low overall signal intensity at pH 10, similar to those of the unfolded T62P variant. Although the I92R and V104R variants retain some secondary structure, they are clearly in a somewhat different conformation, more different from the native one than in any of the other variants.

**Enzymatic Activity.** To determine whether the variant proteins with Arg at internal positions could achieve a folded and enzymatically active conformation, the enzymatic activity of all variants was assayed as described under *Methods*. Eighteen variant proteins out of 25 that were studied had detectable activity. The remaining seven variants (G20R, L36R, T41R, A58R, I92R, V104R, and A109R) were inactive. Positions 20, 36, 41, 58, and 109 are clustered around the active site (Fig. 1); it is likely that Arg at these positions interferes directly with catalysis or with substrate binding without disrupting the native fold. The lack of enzymatic activity in the I92R and V104R variants is fully consistent with CD spectra showing that these proteins are unfolded. The fact that substrate binding did not induce folding suggests that these two variants are highly destabilized relative to the reference protein.

**Thermodynamic Stability.** The thermodynamic stability (Δ$G^\circ_{H_2O}$) of all variants was determined at pH 7 and 10 by denaturation with guanidinium chloride monitored by intrinsic Trp fluorescence at 295 nm. Points represent the normalized fluorescence measured at 295 nm; lines represent two-state linear extrapolation model fit to each curve. Color scheme is identical to A and B. The two-state model could not be fit to the titration curves of the I92R or V104R variants.

**Fig. 1.** Structure of Δ+PHS nuclease showing internal positions that were substituted with Arg. Spheres represent the C$_\alpha$ of residues substituted with Arg. Colors identify variants that were unfolded by the substitution with Arg (red), variants that were folded but enzymatically inactive (yellow), and variants that were folded and enzymatically active (blue). Bold numbers identify the positions (72, 90, and 109) that were substituted with Arg in the three crystal structures that were solved. The circle identifies the approximate location of the active site. Structure shown is Δ+PHS (PDB ID code 3BDC) (42).

**Fig. 2.** Far-UV CD spectra of Arg-containing variants at pH 7 (A) and pH 10 (B). Δ+PHS nuclease (solid black line), Δ+PHS nuclease in 6 M GdnHCl (dashed black line), and the unfolded SNase variant T62P in water (dashed gray line) are shown for reference. Solid gray lines correspond to the 19 Arg variants with no apparent structural rearrangement. Variants of interest are highlighted: A58R (green); V39R, Y91R, and N100R (blue); I92R (orange); and V104R (red). GdnHCl-induced unfolding of Arg-containing variants at pH 7 (C) and pH 10 (D). Points represent the normalized fluorescence measured at 295 nm; lines represent two-state linear extrapolation model fit to each curve. Color scheme is identical to A and B. The two-state model could not be fit to the titration curves of the I92R or V104R variants.
296 nm (20). The unfolding curves were analyzed using a two-state linear extrapolation model (40) (Fig. 2 C and D). The values of ΔΔG_H2O, midpoints of unfolding, and m values are listed for all proteins in Table S1. The unfolding transitions of 23 out of 25 variants exhibited apparent two-state behavior at both pH 7 and 10. Consistent with the CD spectra and with activity measurements, no unfolding transition was detected with Trp fluorescence for the I92R or the V104R variants, indicating that they are unfolded at both pH 7 and 10.

The effect of substitution with Arg on thermodynamic stability (ΔΔG_H2O) was measured for each variant by subtracting ΔG_H2O of the Arg-containing variant from ΔG_H2O of the reference protein (Table 1). At pH 7, the energetic cost of substitution of internal groups with Arg ranged from −1.2 kcal/mol in the L38R variant to −11.8 kcal/mol or more for the I92R and V104R variants. The mean ΔΔG_H2O was −6.2 ± 2.7 kcal/mol at pH 7 and −5.7 ± 2.7 kcal/mol at pH 10. The mean ΔΔG_H2O at pH 7 was comparable to the mean ΔΔG_H2O at pH 7 for substitution of the 25 internal positions with Lys (−5.7 ± 1.7 kcal/mol), Glu (−6.2 ± 1.8 kcal/mol), and Asp (−7.1 ± 2.3 kcal/mol) (21).

**Ionization State of Arg Residues.** The pKₐ values of internal ionizable residues can be measured by analysis of the pH dependence of ΔΔG_H2O (20–22, 26, 36, 41). If the pKₐ values of the Arg residues were depressed relative to the normal pKₐ of Arg in all proteins in Table S1, the unfolding transitions of 23 out of 25 variants exhibited apparent two-state behavior at both pH 7 and 10. However, 25 variants exhibited apparent two-state behavior at both pH 7 and 10. Consistent with the CD spectra and with activity measurements, no unfolding transition was detected with Trp fluorescence for the I92R or the V104R variants, indicating that they are unfolded at both pH 7 and 10.

The replacement of Ala-90 with Arg was tolerated with essentially no change in backbone position (Cα rmsd relative to Δ+PHS of 0.3 Å) (Fig. 3B). The Cα and Cβ atoms of Ile-72 and Arg-92 can be superimposed, the Arg side chain does not point into the β-barrel. Instead it reaches toward bulk water, thereby allowing Arg-72 to satisfy two of its hydrogen bonds with backbone carbonyl oxygen atoms and one with a bound water molecule (Table S3 and Fig. 3D). The Arg Nε atom is solvent inaccessible, but both Nη atoms are approximately 50% solvent-exposed, suggesting that the guanidinium moiety satisfies its remaining two hydrogen bonds with bulk water.

The replacement of Ala-90 with Arg was tolerated with essentially no change in backbone position (Cα rmsd relative to Δ+PHS of 0.3 Å) (Fig. 3B). The Cα and Cβ atoms of Ile-72 and Arg-92 can be superimposed, the Arg side chain does not point into the β-barrel. Instead it reaches toward bulk water, thereby allowing Arg-72 to satisfy two of its hydrogen bonds with backbone carbonyl oxygen atoms and one with a bound water molecule (Table S3 and Fig. 3D). The Arg Nε atom is solvent inaccessible, but both Nη atoms are approximately 50% solvent-exposed, suggesting that the guanidinium moiety satisfies its remaining two hydrogen bonds with bulk water.

**Microenvironments of the Arg Side Chains.** The crystal structures of the variants with I72R, A90R, and A109R were refined to final resolutions of 2.00 Å, 2.15 Å, and 1.90 Å, respectively. Crystallographic data collection parameters and refinement statistics are summarized in Table S2. These three variants had native-like CD spectra. I72R and A90R had detectable enzyme activity, but A109R was one of the variants that was enzymatically inactive. The structures suggested structural and physical origins of the ability of the Arg side chains at internal positions to retain their charge without significant structural reorganization (Fig. 3).

Substitution of Ile-72 with Arg led to a small concerted movement of the backbone relative to the structure of the Δ+PHS nuclease (42); β-strands 1, 2, and 3 moved 3 Å in one direction and the C terminus of helix-1 moved 1 Å in the opposite direction (Fig. 3A). The overall Cα rmsd relative to Δ+PHS nuclease is 0.7 Å. In the structure of Δ+PHS nuclease, Ile-72 points directly into the β-barrel. Although the Cα and Cβ atoms of Ile-72 and Arg-72 can be superimposed, the Arg side chain does not point into the β-barrel. Instead it reaches toward bulk water, thereby allowing Arg-72 to satisfy two of its hydrogen bonds with backbone carbonyl oxygen atoms and one with a bound water molecule (Table S3 and Fig. 3D). The Arg Nε atom is solvent inaccessible, but both Nη atoms are approximately 50% solvent-exposed, suggesting that the guanidinium moiety satisfies its remaining two hydrogen bonds with bulk water.

The replacement of Ala-90 with Arg was tolerated with essentially no change in backbone position (Cα rmsd relative to Δ+PHS of 0.3 Å) (Fig. 3B). The Cα and Cβ atoms of Ile-72 and Arg-92 can be superimposed, the Arg side chain does not point into the β-barrel. Instead it reaches toward bulk water, thereby allowing Arg-72 to satisfy two of its hydrogen bonds with backbone carbonyl oxygen atoms and one with a bound water molecule (Table S3 and Fig. 3D). The Arg Nε atom is solvent inaccessible, but both Nη atoms are approximately 50% solvent-exposed, suggesting that the guanidinium moiety satisfies its remaining two hydrogen bonds with bulk water.

**Table 1. Thermodynamic consequences of substitutions of internal hydrophobic groups in Δ+PHS SNase with Arg**

<table>
<thead>
<tr>
<th>Position</th>
<th>ΔΔG_H2O, kcal mol⁻¹</th>
<th>ΔΔG_H2O slope, kcal mol⁻¹ pH⁻¹</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>pH 7</td>
<td>pH 10</td>
</tr>
<tr>
<td>G20</td>
<td>−3.2 (0.1)</td>
<td>0.3 (0.2)</td>
</tr>
<tr>
<td>V23</td>
<td>−7.5 (0.1)</td>
<td>0.3 (0.2)</td>
</tr>
<tr>
<td>L25</td>
<td>−8.8 (0.1)</td>
<td>0.3 (0.2)</td>
</tr>
<tr>
<td>F34</td>
<td>−9.0 (0.1)</td>
<td>0.4 (0.2)</td>
</tr>
<tr>
<td>L36</td>
<td>−7.4 (0.1)</td>
<td>0.2 (0.2)</td>
</tr>
<tr>
<td>L37</td>
<td>−4.9 (0.1)</td>
<td>0.2 (0.2)</td>
</tr>
<tr>
<td>L38</td>
<td>−1.2 (0.1)</td>
<td>0.0 (0.2)</td>
</tr>
<tr>
<td>V39</td>
<td>−7.2 (0.1)</td>
<td>0.1 (0.2)</td>
</tr>
<tr>
<td>T41</td>
<td>−2.6 (0.1)</td>
<td>0.3 (0.2)</td>
</tr>
<tr>
<td>A58</td>
<td>−3.9 (0.1)</td>
<td>0.1 (0.2)</td>
</tr>
<tr>
<td>T62</td>
<td>−5.2 (0.1)</td>
<td>−0.1 (0.2)</td>
</tr>
<tr>
<td>V66</td>
<td>−9.0 (0.1)</td>
<td>0.2 (0.2)</td>
</tr>
<tr>
<td>I72</td>
<td>−6.2 (0.1)</td>
<td>0.2 (0.2)</td>
</tr>
<tr>
<td>V74</td>
<td>−6.5 (0.1)</td>
<td>0.2 (0.2)</td>
</tr>
<tr>
<td>A90</td>
<td>−6.3 (0.1)</td>
<td>0.3 (0.2)</td>
</tr>
<tr>
<td>Y91</td>
<td>−6.4 (0.1)</td>
<td>0.2 (0.2)</td>
</tr>
<tr>
<td>I92</td>
<td>−10.3 (0.1)</td>
<td>0.2 (0.2)</td>
</tr>
<tr>
<td>V99</td>
<td>−9.5 (0.1)</td>
<td>0.1 (0.2)</td>
</tr>
<tr>
<td>N100</td>
<td>−6.6 (0.1)</td>
<td>0.1 (0.2)</td>
</tr>
<tr>
<td>L103</td>
<td>−6.0 (0.1)</td>
<td>0.1 (0.2)</td>
</tr>
<tr>
<td>V104</td>
<td>−6.3 (0.1)</td>
<td>0.2 (0.2)</td>
</tr>
<tr>
<td>A109</td>
<td>−3.7 (0.1)</td>
<td>0.2 (0.2)</td>
</tr>
<tr>
<td>N118</td>
<td>−1.8 (0.1)</td>
<td>0.2 (0.2)</td>
</tr>
<tr>
<td>L125</td>
<td>−9.5 (0.1)</td>
<td>−0.1 (0.2)</td>
</tr>
<tr>
<td>A132</td>
<td>−7.0 (0.1)</td>
<td>0.4 (0.2)</td>
</tr>
</tbody>
</table>

Values in parentheses are errors propagated from model fit.

¹ΔΔG_H2O = ΔG_H2O,variant − ΔG_H2O,Δ+PHS
²ΔΔG_H2O slope, kcal mol⁻¹ pH⁻¹
³Slope = (∆ΔG_H2O, pH 10/ΔΔG_H2O, pH 7)/pH 10 − pH 7.

was 0.4 ± 0.2 kcal/mol/pH for Arg-36 and Arg-132. This slope could be consistent with an Arg pKₐ no lower than 9.2 (Fig. S1). The ΔΔG_H2O values demonstrated unequivocally that all Arg residues are ionized at pH 7 and that most if not all are also charged at pH 10.
in a conformation different than in the other two variants or in the reference protein. In all three structures the Arg side chain occupied locations where the guanidinium moiety could fulfill its hydrogen bonding potential through interactions with protein oxygen atoms and site-bound water molecules. Although most of the atoms in their side chains are solvent inaccessible, the guanidinium group is sufficiently long to be able to approach the protein–water interface, allowing at least one Nη atom in each side chain to achieve limited solvent-accessibility. This is accomplished without population of unusual rotamers of the Arg side chain (43). Although the Arg side chains have a formal charge of +1, the network of hydrogen bonds is extensive and seems likely to act by solvating and effectively neutralizing the guanidine moiety.

Discussion

The equilibrium thermodynamic data show unequivocally that the Arg side chain retains its charge even when removed from bulk water and buried, at least partially, in the relatively hydrophobic interior of a globular protein. In this respect, Arg is unique. Two-thirds of Lys, Glu, and Asp at the same 25 internal positions in SNase are neutral at pH 7, whereas all Arg retained their charge even up to pH 10. In 23 of the 25 Arg-containing variants studied, the internal Arg ionized without triggering large-scale reorganization of the protein detectable with Trp fluorescence or CD spectroscopy.

These unique properties of Arg could imply that it is less energetically costly to ionize Arg than other types of ionizable residues at internal positions. This cannot be demonstrated without knowing the pKα of Arg in water and in the protein. Unfortunately, the pKα of Arg in water is not known with certainty. It is assumed to be 12 (26), but actual experimental estimates range from 11.5 to 15.0 (44–46). If the pKα of Arg in water were 12, then the equilibrium thermodynamic experiments (Table 1) would have been able to detect shifts in pKα ≥ 2.5. In the previous study of pKα values of internal lysines in SNase, 12 Lys exhibited shifts in this range (26); therefore, if Arg and Lys behaved similarly, approximately 12 Arg residues would have exhibited shifted pKα values. If, on the other hand, the pKα of Arg in water were close to 15, shifts in pKα ≥ 5.5 would have been required to be detectable. A few internal Lys residues in SNase exhibited shifted shifts in this range; therefore, if the cost to ionize Arg and Lys was identical, a few Arg should have exhibited shifted pKα values. Although a pKα of 15 is higher than normally assumed (27), this value was measured in a study designed to eliminate sources of error in previous measurements (44). This measurement has not been validated independently or evaluated critically. Considering that the interpretation of recent studies of biological function of Arg depends on knowledge of its pKα in water, the accurate determination of this pKα should be a high priority. Owing to the uncertainty in the pKα of Arg in water, the absence of observable shifts in the pKα values of Arg residues in internal positions in SNase cannot be interpreted without ambiguity, but it is likely that at least some shifted pKα values would have been detected if they were present.

It is probably energetically less costly to ionize Arg at internal positions in a protein than it is to ionize Lys, Glu, or Asp. Because of its large size and distributed charge, the guanidinium moiety is hydrated more weakly than an amino or carboxylic acid moiety; thus dehydration is energetically less costly (47). For the same reason, this moiety is likely to be less sensitive to the nature of its microenvironment than the amino group of Lys, in which the excess charge is localized on a single heavy atom. The guanidinium moiety can also participate in more types of noncovalent interactions (17, 50), five hydrogen bonds (51, 52), and even favorable interactions with other guanidinium groups (53, 54). All these interactions promote the charged state of Arg. Owing to the length of its side chain, Arg can also sample conformational space more extensively than Lys, Glu, or Asp, thus...
enhancing the probability of placing the guanidinium moiety near polar protein atoms or internal or bulk water.

Many of the factors that promote the charged state of Arg were illustrated by the crystal structures (Fig. 3). The structures show that the Arg side chain reaches closer to the protein–water interface than other internal ionizable groups in SNase. In all three structures, at least one atom of the guanidinium moiety achieves some degree of exposure to bulk water, small but sufficient to partially hydrate the group (Table S3). These internal Arg side chains also fulfill all five possible hydrogen bonds with polar protein atoms or water, reminiscent of what is observed in crystal structures of proteins containing naturally occurring internal Arg residues (8–10, 12–16, 30–32). In contrast, in crystal structures of SNase variants with internal Lys, the amino moiety of the Lys side chain is often surrounded primarily by hydrophobic atoms (20, 37, 55).

The data suggest that the energy required to deform SNase slightly to keep the Arg minimally hydrated is lower than the energy required to dehydrate the group in the protein interior. If this is the case, the energy to substitute internal hydrophobic groups with ionized Arg does not only measure the energy to partially dehydrate the guanidinium group. It also measures the energy to deform the protein to accept the charged Arg side chain. This idea is supported by the high correlation between the thermodynamic cost of substitution of internal positions with charged Arg and charged Lys (Fig. 4). This indicates that, at a given site, a common process must determine the cost of substitution with both types of ionizable groups. Because the innate physicochemical properties of the amino and guanidinium moieties are radically different, the similarity in the cost of inserting Lys and Arg at pH 7 is likely dictated by a property of the protein, rather than the specific physicochemical details of the side chain. At least for the Arg-containing variants for which crystal structures are available (Fig. 3), the common process appears to involve a subtle deformation of tertiary structure to allow the hydration of the internal charged guanidinium moiety. This idea is consistent with detailed studies of the L38K, V66K, and I92K variants of SNase, all of which exhibit structural reorganization to hydrate the amino group of internal Lys residues upon ionization (55–58). An NMR study of other SNase variants with internal Lys residues (56) and computational studies on the same proteins (59) all suggest the involvement of minor conformational reorganization and water penetration in response to the ionization of internal Lys residues.

The deformation of SNase to allow hydration of internal charged groups suggests it would be difficult for any small globular protein to dehydrate and neutralize Arg at physiological pH. It might be possible to bury Arg in the neutral state in a larger protein with a larger hydrophobic core. Consistent with this, the only reports of Arg residues with perturbed $pK_a$ values involve large membrane proteins (34, 35). The hydrophobic interior of these proteins is much larger than for a globular protein; therefore, the Arg side chain can in principle be much farther away from bulk water than in SNase. Furthermore, minor structural deformation to allow hydration of the internal Arg side chain is less likely in a membrane protein. On the other hand, the low cost for transfer of Arg through a lipid bilayer in the interesting experiments of Freites et al. (60) has been rationalized in terms of favorable interactions between charged Arg and protein polar atoms (61), a situation comparable to what was observed with the Arg-containing variants of SNase. It also appears that specialized microenvironments do exist even in relatively small proteins that can promote the neutral form of Arg. This is the case of an internal Arg in the photoactive yellow protein, which in a neutron diffraction structure at pH 9 appears to be deprotonated (62).

Although the hydrophobic character of the interior of a protein is clearly different from that of a lipid bilayer, the ability of the Arg side chain to distort the protein to enhance its hydration and to remain charged is reminiscent of the manner in which the charged Arg side chain is thought to be able to deform a bilayer to maintain access to water and to contact polar head groups to stabilize the charged form (15, 29, 60). Our studies do not contribute definite insight into mechanisms for stabilization of charged Arg in a nonaqueous environment, but they do show that the Arg side chain has a unique capacity to retain its charge in environments that should be expected to promote the neutral form. This ability of Arg to stay charged appears to be fundamental for its functional role as a voltage sensor (8–10, as the essential filter in aquaporins (18, 30, 31), and in other proteins that perform fundamental energy transduction processes (4, 11–13).

Our data show that it is much more difficult to neutralize Arg than Lys, Asp, or Glu near physiological pH in hydrophobic protein environments. It is not clear if this occurs because the $pK_a$ of Arg in water is higher than 12, because of the unique side chain length, hydrogen bonding, and hydration properties of Arg, or some combination of both. The singular capacity of Arg to remain charged in relatively hydrophobic environments, without significantly reorganizing the protein significantly, may explain why naturally occurring internal Arg are abundant and why they seem to be the side chain of choice when biochemical function requires the presence of charge in hydrophobic environments.

### Methods

#### Protein Engineering and Optical Spectroscopy

All experiments were performed with variants of the highly stable $\Delta$-PHS SNase and were prepared and purified as described previously (57). pH titrations monitored by steady-state Trp fluorescence and CD scans from 250 to 200 nm were performed with an Aviv Automated Titration Fluorimeter and an Aviv Circular Dichroism Spectrometer Model 215, as described previously (57).

#### Enzymatic Activity and Thermodynamic Stability

The blue-plate assay was used to measure enzymatic activity at pH 7 as described previously (21). Thermodynamic stability was measured with GdnHCl titrations monitored with Trp fluorescence using an Aviv Automatic Titration Fluorimeter and an Aviv Circular Dichroism Spectrometer Model 215, as described previously.

#### Crystallography and Solvent-Accessibility Calculations

Crystal structures of three variants were solved under cryogenic conditions. Proteins were crystallized at 4 °C by vapor diffusion in hanging drops over reservoir solutions that...
contained 25 mM potassium phosphate buffer (pH 8–9) and 24–42% MPD (wt/vol) as a precipitant. The ΔpH = 1.72R and ΔpH/A90R structures were cocystallized with Ca2+ and with the inhibitor thymidine-3′,5′-diphosphate. Details of data collection and refinement statistics are provided in Table S2.

Solvent-accessibility calculations were performed with NACCESS 2.1.1 using a probe radius of 1.4 Å used with the default van der Waals's radii for each atom.

ACKNOWLEDGMENTS. We gratefully acknowledge Aaron C. Robinson for assistance with the activity measurements. This work was supported by National Institutes of Health Grant GM-061397 (to B.G.-M.E.), by a graduate research fellowship from the National Science Foundation (M.I.H.), and by a Research Experience for Undergraduates Fellowship (G.R.S.) and a Research Opportunity Award (J.L.S.) through the National Science Foundation Grant MCB-0743422 (to B.G.M.E.).

33. Chimenti MS, et al. (2009) Structural consequences of the ionization of internal Lys residues in 4-oxalocrotonate tautomerase on the pKa values of active site residues and on the conformational flexibility at the protein-water interface as a determinant of pKa values. Protein Sci 18:1222–1228.