Predicting Reactivities of Protein Surface Cysteines as Part of a Strategy for Selective Multiple Labeling[†]

Maik H. Jacob, Dan Amir, Vladimir Ratner, Eugene Gussakowsky, and Elisha Haas*

Department of Life Science, Building 212, 4th Floor, Bar-Ilan University, Ramat-Gan, Israel 52900

Received June 23, 2005; Revised Manuscript Received August 30, 2005

ABSTRACT: A variety of biophysical methods used to study proteins requires protein modification using conjugated molecular probes. Cysteine is the main residue that can be modified without the risk of altering other residues in the protein chain. It is possible to label several cysteines in a protein using highly selective labeling reactions, if the cysteines react at very different rates. The reactivity of a cysteine residue introduced into an exposed surface site depends on the fraction of cysteine in the deprotonated state. Here, it is shown that cysteine reactivity differences can be effectively predicted by an electrostatic model that yields site-specifically the fractions of cysteinate. The model accounts for electrostatic interactions between the cysteinyl anion and side chains, the local protein backbone, and water. The energies of interaction with side chains and the main chain are calculated by using the two different dielectric constants, 40 and 22, respectively. Twenty-six mutants of *Escherichia coli* adenylate kinase were produced, each containing a single cysteine at the protein surface, and the rates of the reaction with 5,5′-dithiobis(2-nitrobenzoic acid) (Ellman's reagent) were measured. Cysteine residues were chosen on the basis of locations that were expected to allow modification of the protein with minimal risk of perturbing its structure. The reaction rates spanned a range of 6 orders of magnitude. The correlation between predicted fractions of cysteinate and measured reaction rates was strong ($R = 92\%$) and especially high ($R = 97\%$) for cysteines at the helix termini. The approach developed here allows reasonably fast, automated screening of protein surfaces to identify sites that permit efficient preparations of double- or triple-labeled protein.

Spectroscopic measurements on proteins rely, whenever possible, on intrinsic probes that are naturally provided by the protein. However, a variety of biophysical methods ranging from site-specific infrared dichroism and site-specific NMR (*1*, *2*) to site-specific methods based on fluorescence depend on incorporation of artificial probes. Double-labeled protein is used to follow the distance-dependent Förster energy transfer between two probes (*3*) in single-molecule studies (*4*) and in steady-state and time-resolved ensemble measurements (*5*). These powerful techniques would profit immensely from methods facilitating the preparation of appropriately modified protein.

Modified protein can be obtained on a milligram scale by production and purification of the recombinant protein followed by probe conjugation to natural or engineered residues (*6*). When a labeling reaction is completed, both correctly and erroneously modified as well as unmodified protein molecules are present in the reaction mixture. Separation is possible if protein molecules in different modification states differ in their physical behavior. However, for most experiments to be meaningful, it is required that the original protein and the modified protein be comparable. The inserted probe should minimally perturb the structure, function, or stability of the protein even if this results in a lack of separability. This suggests the need for highly selective labeling reactions that almost exclusively yield the correctly modified protein.

Cysteine is the target of choice for selective reactions; other potential targets for modifications either are poorly reactive, react only with special reagents, or are too frequent in most proteins to allow selective modification of specific sites (*7*). Selectivity problems arise anew when two different probes are to be incorporated and two cysteines are to be modified. Chemical conjugation of the probes must be carried out in two steps, with purification of the desired species following each step. High losses of material can be avoided only when the correctly modified protein is the major reaction product in each of the labeling reactions. This can be achieved when the cysteine residues react with very different rates.

We have demonstrated previously that a 10-fold difference of reaction rates is sufficient to obtain high yields of pure double-labeled protein (*6*). If the reaction rates of two engineered surface cysteines differ by ∼3 orders of magnitude, pure double-labeled protein can be obtained even when the labels do not induce changes in any physical properties that allow separation of the multiple reaction products. The first molecular probe is added to the protein in slight excess (1.1:1) and modifies almost exclusively the more reactive cysteine. The small fraction of undesired double-labeled protein can be removed from the desired single-labeled

[†] This study was supported by grants from the Israel Science Foundation (Grant 621/03), the German Israeli Foundation (Grant 1-676-150.9/2000), the USA−Israel Binational Science Foundation (Grant 1998362), and the Damadian Center for Magnetic Resonance Research at Bar-Ilan University. M.H.J. was supported by fellowships from EMBO.

* Corresponding author (telephone ++972-3-531-8210; fax ++972-3-535-1824; e-mail haas@mail.biu.ac.il).

10.1021/bi051205t CCC: $30.25
Published on Web 10/04/2005

product by, for instance, disulfide-exchange chromatography. The second labeling reaction results in pure double-labeled protein.

Here we report a study of the question of whether the reactivity of surface-exposed cysteine residues can efficiently be predicted. The reactivity of a cysteine residue introduced into an exposed surface site depends on the fraction of the cysteine residue in the deprotonated state. We developed a model to compute the fraction of deprotonated cysteine based on the electrostatic interactions between the cysteinyl anion and its molecular environment. As a model protein we used *Escherichia coli* adenylate kinase—a ubiquitous 214-residue, single-chain, three-domain protein (*8*). We prepared 26 single-cysteine mutants of the cysteine-free C77S-adenylate kinase (AK) and measured the rates of the reactions with DTNB[1] (Ellman's reagent). The calculated fractions of deprotonated cysteine were highly correlated with the observed reaction rates. The predictive power of the model and its relative ease and robustness toward changes of parameter values suggest that it can generally be applied to single out site combinations that are suitable for multiple selective labeling by cysteine insertion and modification.

MATERIALS AND METHODS

Materials. E. coli adenylate kinase mutants were constructed by site-directed mutagenesis, overproduced and purified as described previously (*9*). All variants were derived from the cysteine-free C77S-adenylate kinase form (AK) (*5*, *6*).

Stopped-Flow Kinetic Experiments. The reaction rate of each cysteine mutant with Ellman's reagent was measured using a DX17MV sequential mixing stopped-flow spectrometer from Applied Photophysics (Leatherhead, U.K.). Conditions and procedures were as described (*6*). All measurements were performed in solutions containing 5 mM EDTA and 20 mM Hepes−HCl (pH 7.2) at 25 °C. The absorbance at 460 nm was recorded after stopped-flow mixing of equal volumes of 20 μM protein and 2 mM DTNB. Individual kinetic measurements were repeated at least 10 times, averaged, and analyzed as monoexponential functions.

Selecting Protein Sites Suitable for Minimally Perturbing Modifications. Every sequence position in the adenylate kinase backbone was computationally tested for the extent of rotational freedom available for a potential probe. Using the Swiss-PdbViewer (*10*) we computed a score that accounts for collisions and possible interactions between main-chain or side-chain atoms of the protein molecule and specific rotamers of the side chain of an amino acid inserted by virtual point mutagenesis. In this study, we used the tryptophan residue with its bulky and rigid indole ring to model sterically demanding probes. Averaging the score for 15 different tryptophan rotamers (*11*), we obtained for each position of the AK chain a *rotamer score* as a measure for the rotational restriction induced by a potential probe. Most of the cysteine sites that were selected had favorable rotamer scores, below 14.

Electrostatic Control of Cysteine Reactivity. Only the deprotonated and charged form of cysteine residues in protein molecules reacts with DTNB (*12*, *13*); therefore, the reaction rate *k* is expected to depend on the fraction of protein molecules, *f*, that contain a charged cysteine side chain (eq 1).

$$k \propto f(\mathrm{C}^-) \qquad (1)$$

The fraction of charged cysteine side chains, *f*, depends on the cysteine p*K* value (eq 2) and on the pH of the experimental solutions (pH 7.2).

$$f(\mathrm{C}^-) = 1/(10^{(\mathrm{p}K-\mathrm{pH})} + 1) \qquad (2)$$

Fractions of charged cysteine were obtained from two related models. The first model accounted for electrostatic interactions between the cysteinyl anion and the protein backbone and side chains; the second also included the interaction of the cysteinyl anion with water.

Model 1. The cysteine p*K* was calculated using eq 3.

$$\mathrm{p}K = \Delta pK_{\mathrm{MC,SC}} + \mathrm{p}K_1 \qquad (3)$$

The p*K* shift, $\Delta pK_{\mathrm{MC,SC}}$, is caused by partially charged main-chain atoms (MC) and fully charged side-chain atoms (SC). The constant pK_1 is the intrinsic p*K* of cysteine when electrostatic interactions are not effective and differences in solvation are not explicitly accounted for. It was set to 9.25, close to the p*K* value of 9.50 that was previously reported for a model compound (*14*). The parameter $\Delta pK_{\mathrm{MC,SC}}$ is a function of the electrostatic interaction energy, $W_{\mathrm{MC,SC}}$, between the charge on the cysteine sulfur atom and charges on the protein chain (eq 4).

$$\Delta \mathrm{p}K_{\mathrm{MC,SC}} = -\log[\exp(-W_{\mathrm{MC,SC}})/RT] \qquad (4)$$

The energy of electrostatic interactions between oriented backbone dipoles and the cysteinyl anion can change strongly in the course of rotation of the sulfur atom around the cysteine $\mathrm{C}_\alpha-\mathrm{C}_\beta$ bond. Rotameric states are described by the dihedral angle *i* ($\mathrm{N}-\mathrm{C}_\alpha-\mathrm{C}_\beta-\mathrm{S}$). $W_{\mathrm{MC,SC}}$ was thus obtained by averaging over all $W_{\mathrm{MC,SC}}(i)$ in rotameric states *i* (*i* = 1, 2, ..., 360°) according to the Boltzmann distribution (eq 5).

$$W_{\mathrm{MC,SC}} = \Sigma[f(i) \times W_{\mathrm{MC,SC}}(i)]/\Sigma f(i)$$

with

$$f(i) = \exp[-W_{\mathrm{MC,SC}}(i)/RT] \qquad (5)$$

Rotameric states in which atomic distances are <2.4 Å could not be populated. In a given state, *i*, the energy of electrostatic interaction $W_{\mathrm{MC,SC}}(i)$ considered all interactions of the cysteine charge with the main chain and side chains (eq 6).

$$W_{\mathrm{MC,SC}}(i) = \Sigma W_{\mathrm{MC}}(i) + \Sigma W_{\mathrm{SC}}(i) \qquad (6)$$

Permanent dipoles on five peptide groups that flank the cysteine C_α atom along the protein chain were taken into account. Charged side chains were included in the calculation without any limits. Coulomb's law was used to calculate individual interaction energies $W_{\mathrm{MC}}(i)$ (J/mol) between the cysteinyl anion and a partial charge on a main-chain atom

[1] Abbreviations: p*K*, pK_a; AK, C77S-adenylate kinase from *Escherichia coli*; DTNB, 5,5′-dithiobis(2-nitrobenzoic acid); dielectric constant, relative dielectric constant.

at a distance r (Å) (eq 7).

$$W_{\mathrm{MC}}(i) = 1.39 \times 10^{6} q/(\epsilon_{\mathrm{MC}} r) \qquad (7)$$

The dielectric constant, ϵ_{MC}, was set to 22. Partial charges of 0.55, −0.55, −0.35, and +0.35 were placed on peptide group atoms C, O, N, and H, respectively. By comparison, the CHARMM parameters (param22) for the C_α and the amide hydrogen atoms are 0.1 and 0.25, respectively; the other CHARMM parameters are identical (*15*). The partial charges of proline were set to 0.1 for C_α and C_δ atoms and to −0.2 for nitrogen (param22). The coordinates were taken from the pdb-file, 4AKE (*16*). Amide hydrogen atom positions were calculated by assuming a fixed N−H bond length of 1 Å. The direction $d_{\mathrm{N-H}}$ of the N−H bond vector was calculated from the unit vectors $u_{\mathrm{N-CO}}$ and $u_{\mathrm{N-C\alpha}}$ of bonds N−CO and N−C^α, respectively, according to the expression $d_{\mathrm{N-H}} = -(u_{\mathrm{N-CO}} + u_{\mathrm{N-C\alpha}})$.

Individual energies $W_{\mathrm{SC}}(i)$ of interactions between cysteine and positive or negative charges ($q = \pm 1$) on side chains were calculated via eq 8 with a dielectric constant ϵ_{SC} of 40.

$$W_{\mathrm{SC}}(i) = 1.39 \times 10^{6} q/(\epsilon_{\mathrm{SC}} r) \qquad (8)$$

Discrete side-chain charges were placed on atoms of titratable residues according to the following list: Asp, C^γ; Glu, C^δ; Gly214 (OXT), C; Arg, C^ζ; His, C^δ; Lys, N^ζ; Met1, N.

Model 2. Cysteinyl anions with different degrees of accessibility to water are to different extents stabilized by interactions with permanent and induced dipoles on water molecules. This effect was considered for calculation of the fractions of cysteinate (eq 2) based on cysteine p*K*, using eq 9.

$$\mathrm{p}K = \Delta \mathrm{p}K_{\mathrm{MC,SC}} + \Delta \mathrm{p}K_{\mathrm{B}} + \mathrm{p}K_2 \qquad (9)$$

The additional parameter $\Delta \mathrm{p}K_{\mathrm{B}}$ is the p*K* shift caused by the Born penalty, the additional solvation energy that is needed to accommodate a charge in an environment with a lower relative permittivity than water (*17*).

The constant $\mathrm{p}K_2$ is the p*K* value in the absence of charge−charge interactions and water-displacing atoms; it was set to 8.0, close to the p*K* value of 8.3 that was recently used in two computational studies on electrostatic interactions in unfolded proteins as a reference value for fully solvated cysteine (*18*, *19*). The dependence of $\Delta \mathrm{p}K_{\mathrm{B}}$ on the Born penalty W_{B} is analogous to eq 4 (eq 10).

$$\Delta \mathrm{p}K_{\mathrm{B}} = -\log[\exp(-W_{\mathrm{B}}/RT)] \qquad (10)$$

Charging an atom is energetically much less favorable in the core of a protein, where the local dielectric constant ϵ_{core} can adopt values between 2 and 5 (*20*), than in water with a dielectric constant $\epsilon_{\mathrm{water}} = 78.5$ at room temperature. The effective dielectric constants, ϵ_{EF}, at the surface locations of cysteine sulfur atoms were assumed to adopt intermediate values between ϵ_{core} and $\epsilon_{\mathrm{water}}$. The Born penalties, W_{B}, were accordingly calculated using eq 12 (*17*), where b is the radius of the charged sulfur atom ($b \approx 1$ Å).

$$W_{\mathrm{B}} = 1.39 \times 10^{6}/2b(1/\epsilon_{\mathrm{EF}} - 1/\epsilon_{\mathrm{water}}) \qquad (11)$$

To roughly estimate the value of the effective dielectric constant ϵ_{EF}, a spherical volume with radius $r = 7$ Å around the cysteine sulfur atom position was considered. Water molecules and protein atoms occupying this space were modeled as spheres of volume $4/3\pi \times 1.4^3$ Å^3 and were estimated to contribute to ϵ_{EF} according to their occupation numbers N_{water} and N_{NC} as expressed in eq 12.

$$\epsilon_{\mathrm{EF}} \approx (N_{\mathrm{water}}\epsilon_{\mathrm{water}} + N_{\mathrm{NC}}\epsilon_{\mathrm{core}})/(N_{\mathrm{water}} + N_{\mathrm{NC}}) \qquad (12)$$

The limiting cases are that of a cysteine sulfur atom, fully surrounded either by water molecules ($\epsilon_{\mathrm{EF}} = \epsilon_{\mathrm{water}}$) or by protein atoms ($\epsilon_{\mathrm{EF}} = \epsilon_{\mathrm{core}}$). Here we used $\epsilon_{\mathrm{core}} = 4$ (*21*). The number of water molecules in the spherical volume was approximated using eq 13

$$N_{\mathrm{water}} \approx 0.74 r^3/{r_{\mathrm{water}}}^3 - N_{\mathrm{NC}} \qquad (13)$$

where $r = 7.0$ Å and $r_{\mathrm{water}} = 1.4$ Å. Alternative approaches to modeling that take into account additional geometrical details of the systems are described in ref *17*.

RESULTS

Selection of Protein Sites Suitable for Minimally Perturbing Modifications. Sites suitable for modification are limited to those at which the available space is sufficient to host a molecular probe without inducing global or even local structural rearrangements of the natively folded protein. Ideally, even free rotation of the conjugated label would be possible; otherwise, the rotational entropy of the probe might be higher in the unfolded than in the folded state, and the conformational stability of the protein could be compromised. These conditions can be met only by solvent-exposed sites at the protein surface. When the local shape of the surface is convex, the risk of undesired probe−surface interactions is further reduced. Such geometrical “edge” locations are provided by sites in β turns or coils or at the ends of helices.

To identify suitable sites on the basis of the above considerations, we systematically tested every sequence position in the AK molecule for the extent of rotational freedom available for a potential probe by computing a *rotamer score* (Materials and Methods), which is a rough estimate of the conformational restriction of a potential probe. An alternative convenient measure of conformational restriction, suitable for fast and automated screening of positions, is the *neighbor count*, the number of atoms of the protein, N_{NC}, that are closer than 7 Å to the cysteine sulfur atom (Table 1, column 3).

On the basis of the results of the rotamer-score test, we selected 26 positions of C77S-adenylate kinase for cysteine insertion (Table 1) and prepared the corresponding single-cysteine mutants. The majority of the 26 sequence positions chosen for cysteine insertion provided high rotational freedom reflected by favorable rotamer scores below 14, but we also included sites that were less obvious candidates for modification (Table 1, group 3), as specific experiments can sometimes require placing a probe in protein regions where no highly exposed sites are available.

The selected mutations were situated in a broad variety of structural environments in the AK molecule (Supporting Information A), but more than half of the chosen cysteine positions were located at the terminal two positions of a helix

Table 1: Properties of Cysteine in the Single-Cysteine AK Variants

cysteine position[a]	rotamer score[b]	neighbor count[c] (N_{NC})	ΔT_m[d] (K)	reaction rate k ($mM^{-1}\ s^{-1}$)	accessibility[f] (%)	fraction cysteinate $f(C^-)$	
				$-\log k$ value[e]		$-\log f$ value[g]	$-\log f$ value[h]
Group 1: Cysteines at a Helix Terminus							
18	1	26	nd	−0.42	11.81	1.61	0.99
25	6	26	0.7	1.73	11.45	3.00	2.34
41	10	34	0.3	−0.14	6.35	1.51	1.12
42	7	28	−0.5	−0.26	7.08	1.31	0.77
55	29	36	1.0	0.26	5.90	2.13	1.81
73	33	35	nd	1.57	9.88	3.08	2.71
90	17	40	4.2	1.13	2.90	2.18	2.04
113	10	38	3.8	−0.90	0.81	0.95	0.77
162	3	25	3.1	−1.31	12.28	0.14	0.03
188	34	39	0.3	1.77	7.01	3.35	3.17
189	2	19	5.0	1.71	15.93	3.66	2.81
203	−1	23	0.9	−0.72	16.59	1.11	0.51
Group 2: All Other Accessible Cysteines with a Neighbor Count of <40							
28	9	39	−2.2	0.88	0.89	2.12	1.92
58	1	16	−2.1	−1.18	13.26	2.16	1.27
75	4	21	1.7	−0.62	13.88	0.48	0.12
102	5	30	6.4	1.19	3.50	1.83	1.32
138	11	35	1.1	−0.72	5.01	0.61	0.38
142	20	36	0.8	0.46	5.55	2.12	1.81
148	9	33	2.1	−0.16	7.15	0.26	0.11
154	13	33	1.7	−0.13	2.28	1.04	0.67
169	11	34	−0.9	2.17	7.97	3.52	2.67
214	2	19	1.3	−0.26	19.19	2.16	1.78
Group 3: Cysteines with Minimal DTNB Accessibility and a Neighbor Count of >50							
3	18	73	14.4	4.74	2.5×10^{-20}	2.49	6.02
24	28	53	6.7	3.14	9.2×10^{-3}	2.10	2.74
86	25	52	5.0	1.63	4.1×10^{-10}	2.14	2.66
109	29	59	7.0	2.19	1.0×10^{-32}	2.63	3.78

[a] Cysteines of group 1 occupy the two terminal positions of a helix and the two positions that precede and follow the helix. [b] The program Swiss-PdbViewer was used to test 15 tryptophan conformations inserted in each of the 26 chain positions. The program returns a score that accounts for collisions and favorable interactions of Trp with the protein in each conformation. The averaged rotamer score is listed. In rare cases, the average number of favorable interactions is higher than the collision count, resulting in a negative score. [c] The neighbor count at a cysteine location is measured by the number of protein atoms excluding hydrogen that are <7 Å removed from the cysteine sulfur atom. [d] Heat-induced equilibrium unfolding curves were measured by circular dichroism at 220 nm for AK (C77S-adenylate kinase) and for acetylated AK variants. Differences in midpoint temperatures $\Delta T_m = T_m(AK) - T_m(Ac\text{-}AK)$ are listed. [e] Rates of the reaction of 1 mM DTNB and 10 μM protein in 0.05 M Hepes−HCl, pH 7.2, were measured after stopped-flow mixing at 25 °C by following the absorption at 460 nm. Standard deviations were always <8% of the measured reaction rates (not shown). [f] Cysteine accessibility for DTNB was defined and calculated as described in Supporting Information B. [g,h] Fractions of cysteinate, $f(C^-)_{MC,SC}$ and $f(C^-)_{MC,SC,NC}$ (−log values), were calculated using models 1 (eqs 2−9) and 2 (eqs 10−14).

end or at the first two positions that precede or follow a helix. We grouped the cysteine residues of the 26 AK variants as follows (Table 1): Group 1 contains cysteines at the helix termini with a neighbor count below 50, group 2 consists of all other accessible cysteines with a neighbor count below 40, and group 3 contains cysteines with a neighbor count higher than 50.

To determine whether the substitutions altered the structure of AK, we recorded CD spectra and heat-induced unfolding transitions for the single-cysteine mutants containing free as well as alkylated cysteines (data not shown). Whereas no significant structural perturbations could be detected, some mutations led to a considerably decreased conformational stability of the AK molecule. In particular, for the variants of group 3, large decreases of the melting temperature (Table 1, column 6) were observed.

Rates of Reaction with DTNB. The rates of reaction of each of the 26 AK mutants with DTNB (Ellman's reagent) were measured to ascertain the relative reactivity of these sites. Time constants of the reaction spanned a range from 50 ms (E162C-AK) to 15 h (I3C-AK), covering 6 orders of magnitude (Figure 1). In Figure 1, the values of time constants are indicated by color (see the caption) ranging from red for an extremely fast reaction to black for extremely low reactivity. The most slowly reacting variant, I3C-AK, had the highest neighbor count value, 73, and was the most strongly destabilized mutant among the variants. It was prepared despite its poor rotamer score and neighbor count as a demonstration of an extreme case. Still, even the reaction times of cysteines with high rotational freedom spanned a range from about 50 ms (E162C-AK) to 150 000 ms (G214C-AK).

Factors Determining Cysteine Reactivity at Exposed Sites. In our search for a quantitative model explaining the observed differences in cysteine reaction rates, we considered electrostatic and steric effects. A cysteine residue in a protein or peptide is reactive toward most molecular probes only in its deprotonated charged state (*13*, *25*, *26*). This is why it was possible to obtain p*K* values of particular cysteine residues from the pH dependency of the rate of the reaction with DTNB (*12*, *13*) and with iodoacetamide (*22*). The fraction of cysteine in the deprotonated state depends on electrostatic interactions between the cysteinyl anion and its environment.

To a certain extent, the differences in cysteine reaction rates might be due to differences in the cysteine accessibility to DTNB. The more a cysteine is exposed to the solvent, the more easily it can be accessed by a probe. In the case of

Table 2: Regression Analysis of Rates and Calculated Fractions $f(C^-)$

model	Cys group[a]	R	RMSD	regression line
1[b]	1 (▲)	0.97	0.27	$y = (1.02 \pm 0.08) \times x - (1.68 \pm 0.19)$
	1 (▲) + 2 (○)[c]	0.90	0.44	$y = (0.88 \pm 0.10) \times x - (1.26 \pm 0.20)$
	all 26 variants	0.68	nd	nd
2[d]	1 (▲)	0.97	0.25	$y = (0.99 \pm 0.08) \times x - (1.36 \pm 0.16)$
	1 (▲) + 2 (○)[c]	0.90	0.45	$y = (0.96 \pm 0.11) \times x - (1.02 \pm 0.19)$
	all 26 variants	0.92	0.57	$y = (0.99 \pm 0.09) \times x - (1.08 \pm 0.20)$

[a] See Table 1 and Figure 2. [b] See Figure 2A, $y = -\log k$, $x = -\log f(C^-)_{MC,SC}$. [c] Correlation for cysteines of groups 1 and 2 without C58. [d] See Figure 2B, $y = -\log k$, $x = -\log f(C^-)_{MC,SC,NC}$.

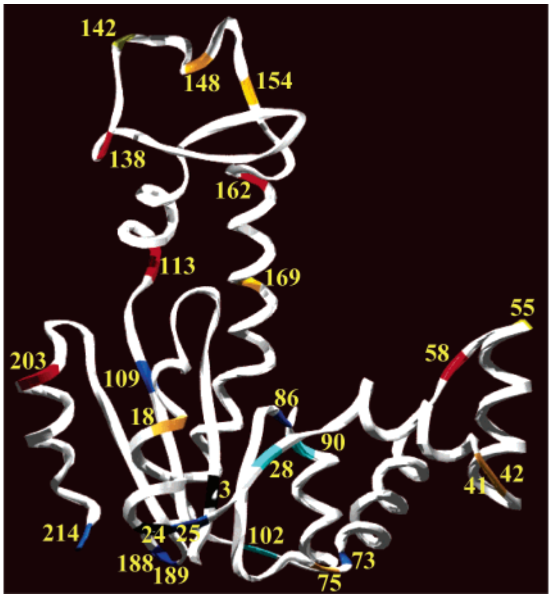

FIGURE 1: Color-coded AK sequence positions for which cysteine−DTNB reaction rates were determined: five cysteines reacted extremely quickly (red, 50−200 ms), seven very quickly (orange, 200−1000 ms), two quickly (yellow, 2 and 3 s), three slowly (cyan, 5−30 s), seven very slowly (blue, 30−200 s), and two extremely slowly (black, 20 min and 15 h).

surface cysteines, however, such excluded-volume effects might be negligible. For a test, we calculated for each cysteine an accessibility index, *A*, based on the crystal structure of adenylate kinase and the geometry of DTNB. The calculation is described in detail in Supporting Information B. It predicted that 22 of the 26 variants, when in conformations similar to those predicted by crystallography, are accessible to DTNB (Table 2, column 6). Thus, a quantitative analysis of cysteine reactivity can make use of the crystal structure in these cases.

The mutants of group 3 had accessibility values approaching zero (Table 1, column 6) and, thus, had to adopt different conformations when reacting with DTNB. These variants reacted most slowly, but their reaction rates were still much higher than expected on the basis of their accessibility indices alone. Protein surface regions are usually flexible and can adopt conformations quite different from that of the crystal structure. Furthermore, the flexibility of the local structure was probably enhanced when hydrophobic residues involved in packing interactions were replaced with cysteine.

Comparing rate constants and accessibility values, we could find no meaningful correlation or improvement of correlations when we combined the *A* values with the rate-determining factors described below. This suggests that excluded-volume effects are irrelevant for the reactivity prediction, probably because we confined our study to cysteines at the flexible protein surface—the different reaction rates are due to different fractions of cysteine in the deprotonated state.

Quantitative Models for Cysteine-Reactivity Prediction. Three primary environmental factors can shift the fraction of cysteinate, the p*K* value of a cysteine, by changing the stability of the thiolate form relative to that of the thiol form. First, amide dipoles of the protein backbone can favorably or unfavorably interact with the cysteinyl anion. They can be formally treated by assigning partial charges to electropositive or electronegative main-chain atoms. These charges are not evenly distributed around the sulfur atom, and their net effect is often substantial (*13*, *23*−*26*). Second, charged side chains can cause a cysteine p*K* shift. At a low salt concentration, the charge−charge interaction energy is inversely proportional to the interatomic distances (*12*, *27*). The third factor is the *Born penalty*, the additional solvation energy that is needed to accommodate a charge in an environment with a lower relative permittivity than water (*17*). Cysteine deprotonation increases with increasing solvation and accessibility to water. The Born penalties explain the low reactivities of the less solvated cysteines of group 3.

If a titratable residue is part of the active site of an enzyme, its p*K* value determination demands complicated approaches such as the PDLD (protein dipoles, Langevin dipoles) method (*28*), the Generalized-Born model (*17*, *29*−*31*), or calculations based on the Poisson−Boltzmann equation (*32*−*35*). However, in case of highly exposed surface residues, models based on additive energies of Coulomb-like interactions might yield no less accurate p*K* values (*36*). We pursued such a model to analyze our system, because it would enable reasonably fast automated screening of whole protein surfaces in the search for sites suitable for modification.

Because a cysteine residue reacts only in the thiolate form, the rate constant *k* is related to the fraction of charged cysteine, *f*, via $k \propto f(C^-)$ (eq 2). Therefore, an appropriate electrostatic model that yields the fraction of charged cysteine fulfills the condition that a logarithmic plot of the rate constant versus the fraction of cysteinate results in a straight line with a slope of 1.

We developed two related models. The first model accounts for the effect of main-chain dipoles (MC) and side-chain charges (SC) on cysteine deprotonation (Materials and Methods, Model 1, eqs 1−8). For each of the AK mutants, the calculated cysteinate fraction $f(C^-)_{MC,SC}$ was plotted against the measured rate of the reaction with DTNB as shown in Figure 2A. Negative log values of fractions of charged cysteine were compared with negative log values of reaction rates (Table 1, column 7). All helix-terminal

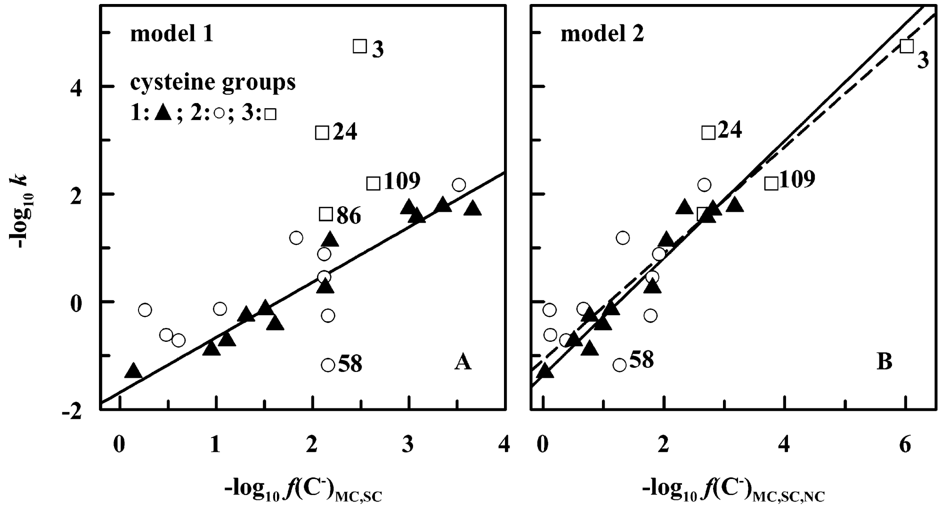

FIGURE 2: Reaction rates k compared to calculated fractions $f(C^-)$ of negatively charged cysteine. Values of $-\log k$ were plotted against values of $-\log f(C^-)$. The cysteine groups (Table 1) are indicated as follows: group 1, ▲; group 2, □; group 3, □. (A) The calculation of fractions $f(C^-)_{MC,SC}$ is based on electrostatic interactions between charged cysteine and main-chain partial charges and side-chain full charges according to model 1 (eqs 1−8). Linear regression (—) of data pairs for helix-terminal residues (▲) results in the first equation given in Table 2. (B) Fractions $f(C^-)_{MC,SC,NC}$ were calculated according to model 2 (eqs 1 and 9−13). Electrostatic interaction energies and, additionally, differences in solvation energies were accounted for. Linear regression (—) of data pairs for helix-terminal residues (▲) yields an unchanged correlation coefficient of $R = 0.97$ (Table 2). Linear regression (- - -) of data pairs for all 26 cysteines results in the last equation given in Table 2.

cysteines of group 1 (▲) closely approach a straight regression line with a slope of 1 (solid line). The correlation coefficient, R, was 0.97 (Table 2).

The four cysteines of group 3 (Table 1) in positions 3, 24, 86, and 109 and cysteine 58 have exceptionally low or high accessibilities to water as indicated by their neighbor counts—the Born penalties were crucial to explain their rates. With these sites excluded, the correlation coefficient would be 0.9 for all cysteines (Table 2).

Thus, the second more complete model included additionally the effect of solvation differences, quantified by using the neighbor counts (NC), on cysteine deprotonation (Model 2, eqs 9−13). In Figure 2B, the reaction rates were plotted as before against fractions of cysteinate $f(C^-)_{MC,SC,NC}$ (Table 1, column 8). Again, the helix-terminal cysteines (▲) in Figure 2B closely approach a straight regression line with a slope of 1 (solid line). The correlation coefficient was still 0.97 (Table 2), because the Born penalties for helix-terminal cysteines were small and comparable, but now, a significant correlation between rates and thiolate fractions ($R = 0.92$) was obtained that included all 26 AK mutants (dashed line, equation given in lower right and Table 2). Apparently, model 2 is superior to model 1 only then, when the set of regarded cysteine sites includes sites with neighbor counts outside the range of 20−40 (Table 2).

The L58C-AK variant reacted more rapidly than predicted by either model (Figure 2). In the crystal structure, the positive charge on the ϵ-amino group of K57 is separated by 10 Å from the sulfhydryl group of C58. However, in other possible rotameric states of K57 the charge−charge distance is reduced to 5 Å, the favorable charge−charge interaction energy is doubled, and cysteine deprotonation and reactivity are accordingly favored. Thus, improved correlations could be obtained by accounting for the dynamics of nearby charged side chains, for instance, by molecular dynamic simulation, which is, however, less suitable for fast automated screening of whole protein surfaces in the search for sites suitable for modification.

The proposed models have several critical parameters, the strong alteration of which would deteriorate the correlation. We added an analysis of the models' robustness toward changes of these parameters in Supporting Information C. Our results indicate that reaction rates and cysteinate fractions correlate best when the energies of cysteinate interaction with a side-chain charge and with a partial main-chain charge are calculated by using two different dielectric constants of 40 and 22, respectively. Note that the former constant should be about twice as high as the latter, whereas small variations of the absolute values have little consequence.

Can the approach be successfully applied to other protein systems? We expect that our models will be of general utility if the parameters that we have identified have similar values for most proteins. In the Discussion, we describe the comparison of our model's parameters with results published by other groups.

DISCUSSION

Selective Double Labeling of Globular Proteins Based on Cysteine Reactivity Differences. For decades the design of site-specifically labeled protein derivatives was guided by a trial and error approach. As a result, quite a few protein samples, which were produced at the cost of much time and effort, proved to be unsuitable for biophysical studies. This motivated an attempt to develop a rational approach for the selection of protein sites suitable for modification.

The procedure described here is designed to accommodate the often contradictory requirements of minimal perturbation and large-scale separation. Both can be achieved if cysteines with different reactivities toward the labels are chosen as modification targets. The risk of protein structure perturbation is minimal when the selected sites enable solvation and free rotation of the conjugated probes. The majority of the single-cysteine variants, which we selected for preparation on the basis of the rotamer score, had cysteine sites with convex surface geometries that provided optimal spatial separation of probe and protein atoms.

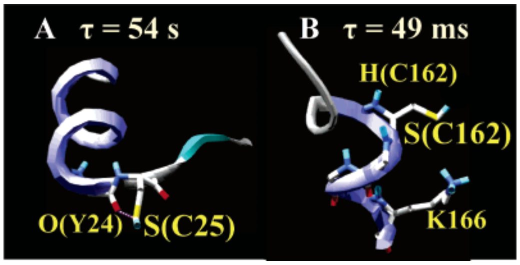

FIGURE 3: Examples of reaction constraints on cysteine residues at various positions. (A) Free rotation of the cysteine sulfur atom in position 25 is blocked by the carbonyl oxygen atom at position 24. (B) During the course of rotation, the sulfur atom S(C162) approaches either the amide hydrogen atom at position 162 and 163 or lysine K166.

The 26 variants reacted with DTNB with rates that spanned a range of 6 orders of magnitude. With this set of 26 sites, 325 different double-cysteine variants could be prepared. About 60% of these variants would have cysteine reaction rates that differ by >10-fold. For 42 variants, the cysteine reactivities would differ by >3 orders of magnitude. Thus, for a huge variety of experimental purposes, pairs of sites can be found that enable both modification with minimal structural perturbation and preparation of the desired doubly modified protein with high efficiency.

Factors Determining Cysteine Reactivity at Exposed Sites. The reactivity of a cysteine side chain depends on the fraction of deprotonated cysteine and therefore on the cysteine p*K*, which is controlled by electrostatic interactions between the cysteinyl anion and its microenvironment. As an example, we discuss why the G25C-AK variant reacted 1000-fold more slowly than the E162C-AK variant (Figure 3). The wild-type residue that was replaced by cysteine in variant G25C-AK (Figure 3A) was glycine, which is frequently found at the C-cap of an α-helix (*37*−*40*). In this position, glycine often adopts φ and ψ angles that are not easily accessible to other residues with space-requiring side chains. Accordingly, the sulfur atom of cysteine 25 approaches the electronegative carbonyl oxygen of cysteine 25 in the course of rotation and even collides with the carbonyl oxygen O(Y24) of tyrosine 24. A charge on S(C25) is expected to be highly unfavorable, explaining the very low reaction rate. Cysteine 162 (Figure 3B) is a helix−N-terminal residue. In all possible rotameric states, favorable interactions between the charge on the sulfur atom S(C162) and electropositive backbone atoms such as hydrogen H(C162) and H(T163) are stronger than unfavorable interactions with electronegative main-chain atoms. In addition, the charge on S(C162) is expected to be profoundly stabilized by the neighboring lysine (K166).

Depending on the application, the calculation of p*K* values requires different levels of sophistication. The charge of the cysteinyl anion can interact with discrete charges on side chains and with permanent dipoles of the main chain, of side chains, and surrounding water molecules. Furthermore, many modes of system polarizability, including induced dipoles and changes in rotamer populations of charged side chains, contribute to the electrostatic interaction energy that stabilizes the thiolate form of cysteine. Electrostatic models and their classification are reviewed in refs *41* and *42*.

In the present study we restricted ourselves to the question of the cysteine p*K* at exposed protein sites that most likely permit perturbation-free modification. This confinement enables a simplified treatment. For example, the cysteinyl anion interacts with dipoles of the whole protein main chain, but we took only the local main-chain segments into account. First, the energy of charge−dipole interaction decreases rapidly (with r^{-2}) with an increasing distance, *r*. Second, the higher the distance, the more effectively different dipole orientations are sampled in the flexible protein so that their net effect is neutralized. Third, most of the tested cysteine sites had a low rotamer score and neighbor count so that neighboring protein atoms belong predominantly to the local protein main chain.

Effective Dielectric Constants. The electrostatic energy of an interaction depends according to Coulomb's law on the interacting charges, their distance, and the dielectric constant of the medium between the charges. The upper limit is given by the dielectric constant of water, 78.5; this value was successfully applied to the interactions between fully solvated discrete charges in unfolded proteins modeled as Gaussian chains (*19*). The effective dielectric constant is smaller for hydrophobic than for hydrophilic regions of proteins, and, as a rule of thumb, it increases as one moves from the protein core toward solvent exposed regions (*41*).

Still, it is possible to use the same dielectric constant for pairs of interacting charges, if they have structurally comparable microenvironments. We found a very good agreement with the experimental data by using only two dielectric constants: a constant, ϵ_{SC}, with a value of 40 for cysteinyl−anion interactions with side-chain charges and a second dielectric constant, ϵ_{MC}, with a value of 22 for interactions with partial main-chain charges.

The ϵ_{SC} value of 40 is in good agreement with the results of other groups. On the basis of many experimental data sets, Warshel and co-workers recommended accounting for charge−charge interactions with a distance-dependent dielectric constant (*36*). For distances between 8 and 20 Å the constant they suggested adopts values between 34 and 53. Laurents et al. studied p*K* differences between the RNase Sa and a charge-reversed variant with five carboxyl to lysine substitutions (*27*). Data obtained by NMR and other spectroscopic methods showed convincingly that the structures of the wild-type protein and the mutant were almost identical despite the numerous replacements. This suggested that the measured p*K* differences were mainly due to different interaction patterns of side-chain charges. Theoretical p*K* differences obtained with a model based on Coulomb's law were found to be in good agreement with the experimental data. The plot of experimental against theoretical p*K* shifts had the desired slope of 1 when an effective dielectric constant of 45 was used.

Our analysis resulted in a 2-fold smaller dielectric constant, ϵ_{MC}, for interactions between the main-chain functional groups and deprotonated cysteine side chains. The low ϵ_{MC} value of 22 reflects the fact that the space surrounding a particular cysteine side chain and a backbone atom is on average less occupied by water molecules than the space that surrounds cysteine and a charge on an exposed side chain. With regard to the correlation, the choice of the ϵ_{MC} value of 22 was optimal for all cysteines, but it was in particular critical for the cysteines of group 1 (Supporting Information C). At the termini of α helices, permanent dipoles of the helix backbone are oriented and interact strongly with the charged form of nearby titratable residues. For example, the protonated form of a histidine residue at the C terminus of the small enzyme, barnase, is stabilized by ∼9 kJ/mol (*43*).

This charge-stabilizing effect was re-examined (*44*) using a method in which water molecules were explicitly modeled by Langevin dipoles (*28*) so that no effective dielectric constants were required. The effect was found to be chiefly due to the first one or two turns of the helix (four peptide groups are equal to a turn). This is in agreement with our study; we found the best correlation between reaction rates and fractions of cysteinate of AK variants when accounting for five flanking peptide groups on each site of the cysteine C_α atom (Supporting Information C).

Relative Influence of Main-Chain Dipoles and Side-Chain Charges on Cysteine Reactivity. According to our analysis, side-chain charges caused, on average, a 10-fold rate difference for any given cysteine pair. Main-chain dipoles induced on average only a 4-fold rate difference in cysteine pairs of group 2, but an 11-fold rate difference among helix-terminal cysteines (Supporting Information D). This emphasizes the role of main-chain electrostatics at helix-terminal positions. This is why the correlations would deteriorate if a dielectric constant ϵ_{MC} far from 22 would be chosen. This is also why about five adjacent peptide groups must be considered to interact with the cysteinyl anion. The interaction energy between a local permanent dipole and a cysteinyl anion depends strongly on the orientation of the anion. This is why Boltzmann averaging over the energies of possible cysteine rotamers was essential (eq 5). Helix-terminal positions are not rare; for example, one-third of all residues of AK belong to this group. These sites often provide optimal geometrical conditions for a molecular label and, in addition, are characterized by widely varying reaction rates of inserted cysteines with molecular probes.

CONCLUSIONS

When the cysteines of a double-cysteine protein react with strongly different rates, site-specifically double-labeled protein can be directly obtained via highly selective labeling reactions. The use of different reaction rates to control the specificity of the reaction is essential when no efficient methods are available to separate the multiple reaction products. A triple-labeled protein can be prepared by making use of both cysteine-reactivity differences and separability of protein species in different modification states.

The analysis presented here describing the labeling of 26 AK cysteine mutants contributes to our understanding of the parameters that determine the reactivity of cysteine side chains at protein positions suitable for modification. The proposed electrostatic models were tested against measured reactivity differences, not against measured p*K* values. The accuracy of the p*K* value predictions must still be tested, but clearly this approach enables the prediction of reactivity differences of protein cysteines for multiple selective modifications. It can be used for automated, rapid screening of protein surfaces to identify cysteine sites that permit economical preparations of double- or triple-labeled protein.

ACKNOWLEDGMENT

We thank Edith Kahana and Varda Ittah for organizational support and for providing an excellent atmosphere for research. We thank Yury Shapiro, Amit Fliess, Ron Unger, Arieh Warshel, and Ernst-Walter Knapp for fruitful discussions.

SUPPORTING INFORMATION AVAILABLE

Cysteine mutants of AK, cysteine accessibility to DTNB, electrostatic models, and electrostatic analysis of single-cysteine AK variants. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Torres, J., and Arkin, I. T. (2002) C-deuterated alanine: a new label to study membrane protein structure using site-specific infrared dichroism, *Biophys. J. 82*, 1068−1075.
2. Weigelt, J., Wikstrom, M., Schultz, J., and van Dongen, M. J. (2002) Site-selective labeling strategies for screening by NMR, *Comb. Chem. High Throughput Screen. 5*, 623−630.
3. Van Der Meer, B. W., Coker, G., III, and Simon Chen, S.-Y. (1994) *Resonance Energy Transfer: Theory and Data*, 1st ed., Wiley, New York.
4. Schuler, B., Lipman, E. A., and Eaton, W. A. (2002) Probing the free-energy surface for protein folding with single-molecule fluorescence spectroscopy, *Nature 419*, 743−747.
5. Ratner, V., Kahana, E., and Haas, E. (2002) The natively helical chain segment 169−188 of *Escherichia coli* adenylate kinase is formed in the latest phase of the refolding transition, *J. Mol. Biol. 320*, 1135−1145.
6. Ratner, V., Kahana, E., Eichler, M., and Haas, E. (2002) A general strategy for site-specific double labeling of globular proteins for kinetic FRET studies, *Bioconjugate Chem. 13*, 1163−1170.
7. Strauss, S., and Lundblad, R. L. (2004) *Chemical Reagents for Protein Modification*, CRC Press, Boca Raton, FL.
8. Gerstein, M., Schulz, G., and Chothia, C. (1993) Domain closure in adenylate kinase. Joints on either side of two helices close like neighboring fingers, *J. Mol. Biol. 229*, 494−501.
9. Sinev, M. A., Landsmann, P., Sineva, E., Ittah, V., and Haas, E. (1996) Domain closure in adenylate kinase, *Biochemistry 35*, 6425−6437.
10. Guex, N., and Peitsch, M. C. (1997) SWISS-MODEL and the Swiss-Pdb Viewer: an environment for comparative protein modeling., *Electrophoresis 18*, 2714−2723.
11. Bower, M. J., Cohen, F. E., and Dunbrack, R. L. (1997) Prediction of protein side-chain rotamers from a backbone-dependent rotamer library; a new homology modeling tool, *J. Mol. Biol. 267*, 1268−1282.
12. Snyder, G. H., Cennerazo, M. J., Karalis, A. J., and Field, D. (1981) Electrostatic influence of local cysteine environments on disulfide exchange kinetics, *Biochemistry 20*, 6509−6519.
13. Miranda, J. L. (2003) Position-dependent interactions between cysteine residues and the helix dipole, *Protein Sci. 12*, 73−81.
14. Antonsiewicz, J., McCammon, J. A., and Gilson, M. K. (1994) Prediction of pH-dependent properties of proteins, *J. Mol. Biol. 238*, 415−436.
15. Brooks, B. R., Bruccoleri, R. E., Olafsen, B. D., States, D. J., Swaminathan, S., and Karplus, M. (1983) CHARMM: a program for macromolecular energy, minimization, and dynamics calculations, *J. Comput. Chem. 4*, 187−217.
16. Muller, C. W., Schlauderer, G. J., Reinstein, J., and Schulz, G. E. (1996) Adenylate kinase motions during catalysis: an energetic counterweight balancing substrate binding, *Structure 4*, 147−156.
17. Pokala, N., and Handel, T. M. (2004) Energy functions for protein design I: efficient and accurate continuum electrostatics and solvation, *Protein Sci. 13*, 925−936.
18. Elcock, A. H. (1999) Realistic Modeling of the Denatured States of Proteins Allows Accurate Calculations of the pH Dependence of Protein Stability, *J. Mol. Biol. 294*, 1051−1062.
19. Zhou, H. X. (2002) A Gaussian-chain model for treating residual charge−charge interactions in the unfolded state of proteins, *Proc. Natl. Acad. Sci. U.S.A. 99*, 3569−3574.
20. Pitera, J. W., Falta, M., and van Gunsteren, W. F. (2001) Dielectric properties of proteins from simulation: the effects of solvent, ligands, pH, and temperature, *Biophys. J. 80*, 2546−2555.
21. Simonson, T., and Perahia, D. (1996) Polar fluctuations in proteins: molecular-dynamic studies of cytochrome *c* in aqueous solution, *Faraday Discuss.*, 71−90.

22. Krekel, F., Samland, A. K., Macheroux, P., Amrhein, N., and Evans, J. N. (2000) Determination of the pK_a value of C115 in MurA (UDP-*N*-acetylglucosamine enolpyruvyltransferase) from *Enterobacter cloacae*, *Biochemistry 39*, 12671−12677.
23. Hol, W. G., van Duijnen, P. T., and Berendsen, H. J. (1978) The alpha-helix dipole and the properties of proteins, *Nature 273*, 443−446.
24. Kortemme, T., and Creighton, T. E. (1995) Ionisation of cysteine residues at the termini of model alpha-helical peptides. Relevance to unusual thiol pK_a values in proteins of the thioredoxin family, *J. Mol. Biol. 253*, 799−812.
25. Lockhart, D. J., and Kim, P. S. (1992) Internal stark effect measurement of the electric field at the amino terminus of an alpha helix, *Science 257*, 974−951.
26. Sancho, J., Serrano, L., and Fersht, A. R. (1992) Histidine residues at the N- and C-termini of alpha-helices: perturbed pK_as and protein stability, *Biochemistry 31*, 2253−2258.
27. Laurents, D. V., Huyghues-Despointes, B. M. P., Bruix, M., Thurkill, R. L., Schell, D., Newsom, S., Grimsley, G. R., Shaw, K. L., Trevino, S., Rico, M., Briggs, J. M., Antonsiewicz, J. M., Scholtz, J. M., and Pace, C. N. (2003) Charge−Charge Interactions are Key Determinants of the p*K* values of ionizable Groups in Ribonuclease Sa (pI = 3.5) and a Basic Variant (pI = 10.2), *J. Mol. Biol. 325*, 1077−1092.
28. Sham, Y. Y., Chu, Z. T., and Warshel, A. (1997) Consistent Calculations of pK_a's of Ionizable Residues in Proteins: Semi-microscopic and Microscopic Approaches, *J. Phys. Chem. B 101*, 4458−4472.
29. Bashford, D., and Case, D. A. (2000) Generalized born models of macromolecular solvation effects, *Annu. Rev. Phys. Chem. 51*, 129−152.
30. Totrov, M. (2004) Accurate and efficient generalized born model based on solvent accessibility: derivation and application for LogP octanol/water prediction and flexible peptide docking, *J. Comput. Chem. 25*, 609−619.
31. Feig, M., Onufriev, A., Lee, M. S., Im, W., Case, D. A., and Brooks, C. L., 3rd. (2004) Performance comparison of generalized born and Poisson methods in the calculation of electrostatic solvation energies for protein structures, *J. Comput. Chem. 25*, 265−284.
32. Nicholls, A., and Honig, B. (1991) A Rapid Finite Difference Algorithm, Utilizing Successive Over-Relaxation to Solve the Poisson−Boltzmann Equation, *J. Comput. Chem. 12*, 435.
33. Antosiewicz, J., McCammon, J. A., and Gilson, M. K. (1994) Prediction of pH-dependent properties of proteins, *J. Mol. Biol. 238*, 415−436.
34. Rocchia, W., Sridharan, S., Nicholls, A., Alexov, E., Chiabrera, A., and Honig, B. (2002) Rapid grid-based construction of the molecular surface and the use of induced surface charge to calculate reaction field energies: applications to the molecular systems and geometric objects, *J. Comput. Chem. 23*, 128−137.
35. Neves-Petersen, M. T., and Petersen, S. B. (2003) Protein electrostatics: a review of the equations and methods used to model electrostatic equations in biomolecules−applications in biotechnology, *Biotechnol. Annu. Rev. 9*, 315−395.
36. Russell, S. T., and Warshel, A. (1985) Calculations of electrostatic energies in proteins. The energetics of ionized groups in bovine pancreatic trypsin inhibitor, *J. Mol. Biol. 185*, 389−404.
37. Richardson, J. S., and Richardson, D. C. (1988) Amino acid preferences for specific locations at the ends of alpha helices, *Science 240*, 1648−1652.
38. Chakrabartty, A., Doig, A. J., and Baldwin, R. L. (1993) Helix capping propensities in peptides parallel those in proteins, *Proc. Natl. Acad. Sci. U.S.A. 90*, 11332−11336.
39. Dasgupta, S., and Bell, J. A. (1993) Design of helix ends. Amino acid preferences, hydrogen bonding and electrostatic interactions, *Int. J. Pept. Protein Res. 41*, 499−511.
40. Kapp, G. T., Richardson, J. S., and Oas, T. G. (2004) Kinetic role of helix caps in protein folding is context-dependent, *Biochemistry 43*, 3814−3823.
41. Schutz, C. N., and Warshel, A. (2001) What Are the Dielectric "Constants" of Proteins and How To Validate Electrostatic Models?, *Proteins: Struct., Funct. Genet.* 400−417.
42. Ullmann, G. M., and Knapp, E.-W. (1999) Electrostatic models for computing protonation and redox equilibria in proteins, *Eur. Biophys. J. 28*, 533−551.
43. Serrano, L., and Fersht, A. R. (1989) Capping and alpha-helix stability, *Nature 342*, 296−299.
44. Aqvist, J., Luecke, H., Quiocho, F. A., and Warshel, A. (1991) Dipoles localized at helix termini of proteins stabilize charges, *Proc. Natl. Acad. Sci. U.S.A. 88*, 2026−2030.

BI051205T