J|A|C|S
ARTICLES
Published on Web 08/31/2002

Charged States of Proteins. Reactions of Doubly Protonated Alkyldiamines with NH_3: Solvation or Deprotonation. Extension of Two Proton Cases to Multiply Protonated Globular Proteins Observed in the Gas Phase

Michael Peschke, Arthur Blades, and Paul Kebarle*

Contribution from the Department of Chemistry, University of Alberta, Edmonton, Alberta, Canada T6G 2G2

Received November 26, 2001

Abstract: The apparent gas-phase basicities (GB^{app}'s) of basic sites in multiply protonated molecules, such as proteins, can be approximately predicted. An approach used by Williams and co-workers was to develop an equation for a diprotonated system, $NH_3(CH_2)_7NH_3^{2+}$, and then extend it with a summation of pairwise interactions to multiply protonated systems. Experimental determinations of the rates of deprotonation of $NH_3(CH_2)_7NH_3^{2+}$ by a variety of bases B, in the present work, showed that $GB^{app} = GB(NH_3) = 196$ kcal/mol. This result is supported also by determinations of the equilibria: $NH_3(CH_2)_pNH_3^{2+} + NH_3 = NH_3(CH_2)_pNH_3 \cdot NH_3^{2+}$, for $p = 7$, 8, 10, 12. The described experimental GB^{app} is 14 kcal/mol higher than the value predicted by the equation used by Williams and co-workers but in agreement with an ab initio result by Gronert. Equations based on electrostatics are developed for the two proton and multiproton systems which allow the evaluation of GB^{app} of the basic sites on proteins. These are applied for the evaluation of GB^{app} of the basic sites and of N_{SB}, the maximum number of protons that the nondenatured proteins, carbonic anhydrase (CAII), cytochrome *c* (CYC), and pepsin, can hold. The N_{SB} values are compared with the observed charges, Z_{obs}'s, when the nondenatured proteins are produced by electrospray and found in agreement with the proposal by de la Mora that Z_{obs} is determined by the number of charges provided by the droplet that contains the protein, according to the charge residue model (CRM). The GB^{app} values of proteins have many other applications. They can be compared with experimental measurements and are also needed for the understanding of the thermal denaturing of charged proteins and the thermal dissociation of charged protein complexes.

I. Introduction

The mass spectrometric study (ESIMS) of nondenatured proteins, protein−protein, and protein−substrate complexes produced by electrospray ionization (ESI) is at present a most active area of research,[1] which is making important contributions to biochemistry and biopharmacology (ref 1 represents only a small sampling of early and recent publications). Some of these studies attempt to correlate the noncovalent binding energy of the complex in the biological environment with the binding energy determined in the gas phase by mass spectrometric techniques.

Examples of such recent and very interesting work are experiments using the blackbody infrared radiative dissociation (BIRD) technique to dissociate multiply protonated protein complexes in an FTICR mass spectrometer.[2,3] For example, the multiply protonated protein pentamer of the Shiga-like toxin was found[3] to decompose predominantly to the tetramer and monomer. The multiple charges present on the pentamer and on the tetramer and monomer products were found to have a significant effect on the activation energies of the decomposition.[3] These results clearly demonstrate that it is very desirable to understand the origin of the charges on the proteins formed by ESI, know the positions on which the protons reside, and have an understanding of the ability of the protons to migrate to other basic groups, when the protein is heated to higher temperatures including temperatures that will lead to thermal decomposition. Such an understanding might allow the evaluation of the Coulombic energy terms due to the charges. Subtraction of the Coulombic energy terms from the observed activation energy might then lead to energy values which are representative of the bond energy of the complex due to noncovalent bonds, which hold it together in solution.

The origin of the charges on the proteins can be understood only from a consideration of the mechanism of electrospray. Recent research[4−11] on the mechanism of the generation of gas-

* To whom correspondence should be addressed. E-mail: Paul.Kebarle@ualberta.ca.

(1) (a) Ganem, B.; Li, Y.; Henion, J. D. *J. Am. Chem. Soc.* **1991**, *113*, 6294. (b) Ganem, B.; Li, Y. T.; Henion J. *J. Am. Chem. Soc.* **1991**, *113*, 7818. (c) Katta, V.; Chait, B. T. *J. Am. Chem. Soc.* **1991**, *113*, 8534. (d) Schwartz, B. S.; Light-Wahl, K. J.; Smith, R. D. *J. Am. Soc. Mass Spectrom.* **1994**, *5*, 201. (e) Light-Wahl, K. J.; Schwartz, B. L.; Smith, R. D. *J. Am. Chem. Soc.* **1994**, *116*, 5277. (f) Lafitte, D. *Eur. J. Biochem.* **1999**, *261*, 337.

(2) (a) Price, W. D.; Schnier, P. D.; Jockbush, R. H.; Strittmacher, E. F.; Williams, E. R. *J. Am. Chem. Soc.* **1996**, *118*, 10640. (b) Gross, B. S.; Zhao, Y.; Williams, E. R. *J. Am. Soc. Mass Spectrom.* **1997**, *8*, 519.

(3) Felitsyn, N.; Kitova, E. N.; Klassen, J. S. *Anal. Chem.* **2001**, *73*, 4647.

10.1021/ja012591e CCC: $22.00

phase ions by the ESI method indicates that the small ions (such as inorganic ions Na^+, NH_4^+, and so forth, or organic ions such as protonated organic bases BH^+) are produced by the ion evaporation model (IEM),[4−7,9−11] while large macro ions and typically the nondenatured globular (native) proteins are produced by the charge residue model (CRM).[4,6,8]

The most significant evidence that the multiply charged native proteins are produced by CRM was provided by de la Mora.[8] He showed that the experimentally observed number of charges, Z_{obs}, reported in the ESIMS literature was equal to Z_{CRM}, which is the charge at the surface of the precursor water droplets which contain the protein when the evaporating droplet has just reached the size of the protein. Z_{CRM} can be evaluated with the Rayleigh equation[8] and a value R for the radius of the protein (see Figure 1 in de la Mora[8] and Figure 2 and discussion in Felitsyn et al.[12]).

De la Mora[8] did not discuss the chemistry of protonation, that is, how exactly the charge at the surface of the disappearing droplets is converted to charge of the proteins. Recent work from this laboratory[12] has provided evidence that when the protein is sprayed from aqueous solution containing ammonium acetate as the major electrolyte, the charging is due to NH_4^+ ions. Ammonium containing buffer salts are the most frequently used buffers for nondenatured proteins. (When buffers are not used, the small electrolyte ions needed for electrospray must be provided by the ionization of the basic and acidic side chains. In that case, H_3O^+ ions are expected on the droplet surface. These will be in excess to the negative counterions in the droplet, and the charging will be due to the excess hydronium ions.)

As the last water evaporates, most of the ionized basic side chains can be expected to become neutralized by reacting with nearby counterions. Such a process will be fostered by prior ion pairing, because of the increasing electrolyte concentration in the evaporating droplets. In the last stage, the NH_4^+ ions at the surface of the droplet end up on the protein and can react with the basic side chains at the surface of the protein. The lowest energy products will be the proton-bridged adducts:

$$\text{protein}\sim\!\!\sim NH_2 + NH_4^+ = \text{protein}\sim\!\!\sim NH_2{-}H{-}NH_3^+ \quad (1)$$

The side chain in eq 1 models lysine but could be also any other basic side chain. Equation 1 indicates only the first attachment of an NH_4^+. However, there are many more NH_4^+ ions available such that ultimately Z_{CRM} ammonium ions can become attached to different basic side chains at the surface of the protein. A complete proton transfer leading to protonation of the side chain

$$\text{protein}\sim\!\!\sim NH_2{-}H{-}NH_3^+ \rightarrow \text{protein}\sim\!\!\sim NH_3^+ + NH_3 \quad (2a)$$

occurs later in the "desolvation" stage in the sampling system, either in the heated (100−200 °C) sampling capillary leading to the mass spectrometer or in the CID stage because of ion acceleration by the electric field applied between sampling capillary and skimmer electrodes. Charge loss via process 2b, driven by the Coulombic repulsion between the other charges and the leaving charged group, can also occur:

$$(\text{protein}\sim\!\!\sim NH_2{-}H{-}NH_3)^{Z+} \rightarrow (\text{protein}\sim\!\!\sim NH_2)^{(Z-1)+} + NH_4^+ \quad (2b)$$

Therefore, Z_{obs} will not depend only on Z_{CRM} but also on the ability of the protein to hold the charge when processed through the "desolvation" stage. This contingency was not considered by de la Mora. The number of basic sites, N_{SB}, of a given protein that can hold protons in the gas phase, depends on the apparent gas-phase basicities, GB^{app}'s, of the basic sites near the surface of the protein.

Williams and co-workers in a series of papers[13] were the first to try to determine GB^{app} and N_{SB}. Most of their work dealt with denatured proteins; however, calculations were made also for one globular protein (cytochrome *c*).[13d] The approach used was to develop a relationship based on electrostatics for GB^{app} of a diprotonated molecule and then generalize this relationship to more than two charged groups. More recently, an insightful analysis based on ab initio calculations for the two proton system

$$NH_3 + H_3N(CH_2)_7NH_3^{2+} = (NH_3(CH_2)_7NH_3{\cdot}NH_3)^{2+} = NH_3(CH_2)_7NH_2^+ + NH_4^+ \quad (3)$$

was published by Gronert.[14] This analysis predicts values for GB^{app} that differ significantly from those obtained by Williams et al.[13] for the same system.

Because of the importance of the two proton system as a test case, experimental measurements involving the alkyl diamines $NH_3(CH_2)_pNH_3^{2+}$ produced by ES and their solvation or deprotonation by NH_3 and other bases were performed (see sections a and b in Results and Discussion). The experimental results support Gronert's analysis for the two charge case. Therefore, improved equations based on electrostatics are developed for the doubly and then multiply protonated proteins (see sections c and d in Results and Discussion).

These equations are then used for the evaluation of GB^{app} and N_{SB} of three globular proteins: carbonic anhydrase, cytochrome *c*, and pepsin. The N_{SB} values obtained are compared with the experimentally observed charges, Z_{obs}, and Z_{CRM}, the predicted charges by CRM. This comparison provides an answer to the following question: Which determines Z_{obs}? Is it Z_{CRM} or N_{SB}? (See section e in Results and Discussion.)

The significance of the calculations predicting GB^{app} goes much beyond the answers concerning the mechanism leading to the observed charges on the proteins. The GB^{app} values can also be determined experimentally[15] (see section IIIf, in Results and Discussion).

The reactions 2a,b represent also the simplest prototype for the dissociation of protein−substrate complexes. Therefore, the

(4) Kebarle, P.; Ho, Y. In *Electrospray Ionization Mass Spectrometry*; Cole, R. B., Ed.; John Wiley & Sons: New York, 1997; p 55.
(5) Kebarle, P.; Peschke, M. *Anal. Chim. Acta* **2000**, *406*, 11.
(6) Cole, R. B. *J. Mass Spectrom.* **2000**, *35*, 763.
(7) Gamero-Castano, M.; de la Mora, J. F. *Anal. Chim. Acta* **2000**, *406*, 67.
(8) de la Mora, J. F. *Anal. Chim. Acta*. **2000**, *406*, 93.
(9) Labowsky, M.; Fenn, J. B.; de la Mora, J. F. *Anal. Chim. Acta*. **2000**, *406*, 105.
(10) Gamero-Castano, M.; de la Mora, J. F. *J. Mass Spectrom*. **2000**, *35*, 780.
(11) Kebarle, P. *J. Mass Spectrom*. **2000**, *35*, 804.
(12) Felitsyn, N.; Peschke, M.; Kebarle, P. *Int. J. Mass Spectrom.* **2002**, *219*(*1*), 39.
(13) (a) Gross, D. S.; Rodriguez-Cruz, S. E.; Bock, S.; Williams, E. R. *J. Phys. Chem.* **1995**, *99*, 4034. (b) Gross, D. S.; Williams, E. R. *J. Am. Chem. Soc.* **1995**, *117*, 883. (c) Schnier, P. D.; Gross, D. S.; Williams, E. R. *J. Am. Chem. Soc.* **1995**, *117*, 6747. (d) Schnier, P. D.; Gross, D. S.; Williams, E. R. *J. Am. Soc. Mass Spectrom.* **1995**, *6*, 1086. (e) Williams, E. R. *J. Mass Spectrom.* **1996**, *31*, 831.
(14) (a) Gronert, S. *J. Am. Chem. Soc.* **1998**, *118*, 3525. (b) Gronert, S. *J. Mass Spectrom.* **1999**, *34*, 787. (c) Gronert, S. *Int. J. Mass Spectrom.* **1999**, *185*/*186*/*187*, 351.

Table 1. Thermochemical Data[a] for $NH_3(CH_2)_pNH_3{\cdot}NH_3^{2+} = NH_3(CH_2)_pNH_3^{2+} + NH_3$

p	$\Delta H_{1,0}°$	$\Delta G_{1,0}°$	$\Delta S_{1,p}°$
10	20.3	13.1	24.3
12	19.5	12.7	22.6

[a] Energy values in kcal/mol. Entropy in cal/(degree·mol). Standard state 1 atm. $\Delta G_{1,0}°$ value at $T = 298$ K.

equations developed for the evaluation of the activation energies of these two reactions can be of significance to the development of equations for the protein−substrate complexes. The thermal denaturing of a multiply protonated native protein[16] is at least partially driven by the Coulombic repulsions between the charges and represents another related area to which the GB^{app} values are of significance.

II. Experimental Section

The experimental measurements relating to reaction 3 involving $NH_3(CH_2)_pNH_3^{2+}$ and NH_3 were performed with a reaction chamber sampled by a quadrupole mass spectrometer which has been described.[17] The same apparatus was used previously for determination of the hydration equilibria.[17]

$$NH_3(CH_2)_pNH_3^{2+} + H_2O = (NH_3(CH_2)_pNH_3{\cdot}H_2O)^{2+} \quad (4)$$

The reagent ion, $NH_3(CH_2)_pNH_3^{2+}$ ($p = 5-12$), was produced by electrospray, and the equilibria in eq 4 were determined by introducing these ions into the reaction chamber which contained also 10 Torr of N_2 as bath gas and known low partial pressures (1−100 mTorr) of H_2O. The intensities of the ionic products of the equilibria were determined with a quadrupole mass spectrometer.

Very similar techniques were used in the present work for the determination of solvation equilibria involving the diprotonated diamines and NH_3. The data obtained are presented in the Results and Discussion section (Table 1, Figures 2 and 3). However, for these systems, the reaction can involve not only solvation but also deprotonation by NH_3, particularly so for $p \leq 7$, see eq 3. Therefore, it was desirable also to determine the rates of deprotonation.

In the apparatus used,[17] the reactant ion, when inside the reaction chamber, is exposed to a weak drift field by applying a small voltage V_d between the ion entrance orifice IN and the ion exit orifice OR (see Figure 1 in Blades et al.[17]). This drift field and the pressure of the bath gas N_2 control the drift velocity and, thus, also the reaction time, t. The value of the drift voltage V_d is low so that the drifting ions have thermal internal energies. The drift times of the ions in the reaction chamber are in the 100−1000 μs range depending on the value of the drift field.

Relative rate constants for a given protonated ion, such as a given doubly protonated diamine, BH_2^{2+}, and different neutral bases, A, can be determined by working at constant drift voltage V_d, which leads to a constant drift and reaction time t when the same reagent BH_2^{2+} ion is involved. For the proton-transfer reaction eq 5, eq 6 is used.

$$BH_2^{2+} + A = BH^+ + AH^+ \quad (5)$$

$$\ln([BH_2^{2+}]_A/[BH_2^{2+}]_0) = kt[A] \quad (6)$$

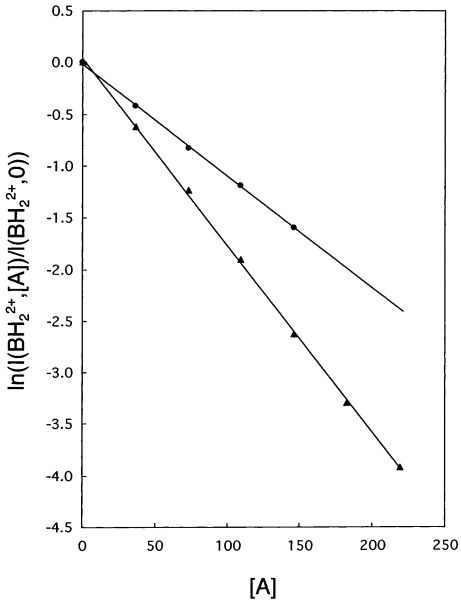


Figure 1. Results shown are for deprotonation reaction $BH_2^{2+} + A = BH^+ + AH^+$, where $BH_2^{2+} = NH_3(CH_2)_7NH_3^{2+}$. Logarithmic plot of ion intensity of BH_2^+, observed when a constant concentration of base A is present. Results are from several experiments with different [A] including [A] = 0. ● = NH_3, ▲ = CH_3NH_2.

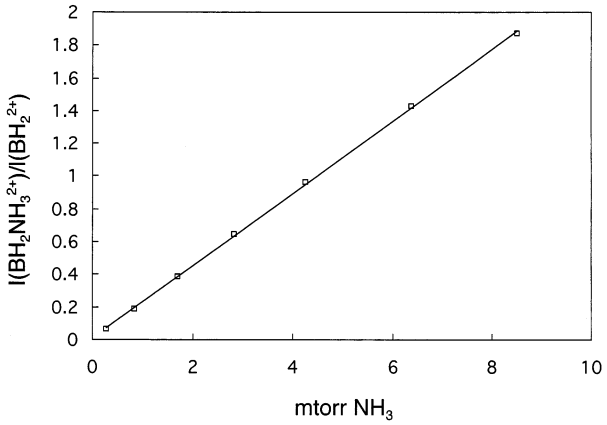


Figure 2. Plot of ion intensity ratio for ions $BH_2^{2+}{\cdot}NH_3$ and BH_2^+, where $BH_2 = NH_3(CH_2)_{12}NH_3^{2+}$, versus pressure of ammonia. Linearity of plot which goes through the origin meets equilibrium conditions, eq 12.

$[BH_2^{2+}]_A$ is the ion concentration after time t, that is, at the exit of the reaction chamber, when a constant concentration [A] is present in the reactor. $[BH_2^{2+}]_0$ is the concentration after time t, when [A] = 0. The ion concentration ratio in eq 6 is replaced with the corresponding ion intensity ratio $I(BH_2^{2+})_A/I(BH_2^{2+})_0$, observed with the mass spectrometer.

Determinations with a range of constant concentrations [A], at constant drift field, that is, constant t, lead to plots such as those shown in Figure 1, for $BH_2^{2+} = NH_3(CH_2)_7NH_3^{2+}$ and the two bases A, equal to NH_3 and CH_3NH_2. The plots are linear as expected from eq 6, and the slopes give the values for kt for each base A. The value of t is constant, because the same reagent ion is involved. Therefore, the slopes provide the relative values for the proton-transfer rate constants. Values of kt obtained for several bases are tabulated in the Results and Discussion section, see Table 2.

Equation 6 is obeyed even though there is also loss of the BH_2^{2+} ion, by diffusion to the wall followed by discharge of the ion. Such a loss is indicated by the observation that $I(BH_2^{2+})_0$ decreases when t is increased by lowering V_d. It can be readily derived that eq 6 will hold, if the ion loss by other parallel reactions is first order in the ion concentration. First order ion loss can be expected in the present

(15) (a) Cassady, C. J.; Carr, S. *J. Mass Spectrom.* **1996**, *31*, 247. (b) Cassady, C. J.; Wronka, J.; Kruppa, G. H.; Laukien, F. H. *Rapid Commun. Mass Spectrom.* **1994**, *8*, 394. (c) Ogorzalek Loo, R. R.; Winger, B. E.; Smith, R. D. *J. Am. Soc. Mass Spectrom.* **1994**, *5*, 1064. (d) Ogorzalek Loo, R. R.; Smith, R. D. *J. Mass Spectrom.* **1995**, *30*, 339. (e) McLuckey, S. A.; Glish, G. L.; van Berkel, G. J. *Anal. Chem.* **1991**, *63*, 1971.
(16) Mao, Y.; Ratner, M. A.; Jarrold, M. F. *J. Phys. Chem. B* **1999**, *103*, 10017.
(17) Blades, A. T.; Klassen, J. S.; Kebarle, P. *J. Am. Chem. Soc.* **1996**, *118*, 12437.

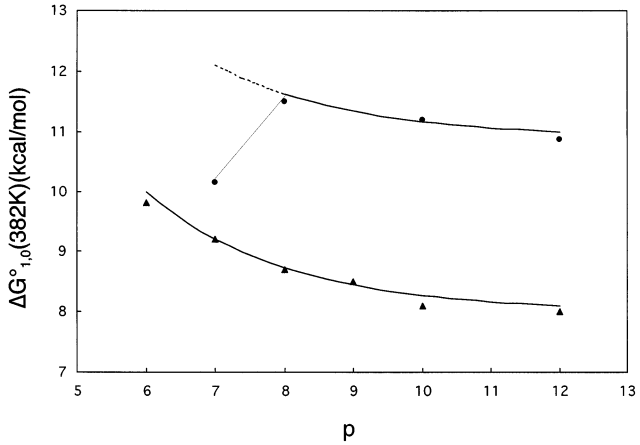


Figure 3. Plot of free energy values, $\Delta G_{1,0}°$, for the following reactions: $(NH_3(CH_2)_pNH_3{\cdot}NH_3)^{2+} = NH_3(CH_2)_pNH_3^{2+} + NH_3$ (●) and $((NH_3CH_2)_p$-$NH_3{\cdot}H_2O)^{2+} = NH_3(CH_2)_pNH_3^{2+} + H_2O$ (◆) for various values of p. Results for NH_3 are from the present work, results for H_2O, from previous work.[17] Temperature = 283 K. Sudden drop off for value for $p = 7$ (NH_3) indicates anomalous result, leading only to an apparent equilibrium. The true equilibrium could not be determined because of rapid deprotonation of the doubly protonated ions by NH_3.

Table 2. Reaction Rate Constants[a]

$NH_3(CH_2)_7NH_3^{2+} + A = NH_3(CH_2)_7NH_2^{+} + AH^{+}$

A	GB (A)[b]	α (A³)[c]	kt[d]	k[e]	k_c[f]
NH_3	196	2.2	4×10^{-13}	2.8×10^{-9}	1.8×10^{-9}
pyrrole	201.8	9.3	9.2×10^{-13}	6.6×10^{-9}	2.1×10^{-9}
aniline	203.5	(14.0)	9.1×10^{-13}	6.6×10^{-9}	2.4×10^{-9}
CH_3NH_2	206.8	(4.2)	8.0×10^{-13}	5.8×10^{-9}	1.9×10^{-9}
n-$C_3H_7NH_2$	211.5	(8.1)	8.2×10^{-13}	6.0×10^{-9}	2.0×10^{-9}
iso-$C_3H_7NH_2$	212.7	(8.1)	8.1×10^{-13}	6.0×10^{-9}	2.0×10^{-9}
$(C_2H_5)_3N$	227.5	(13.0)	8.2×10^{-13}	6.0×10^{-9}	2.2×10^{-9}
(2) n-$C_4H_9NH_3^{+} + (C_2H_5)_3N = n$-$C_4H_4NH_2 + (C_2H_5)_2NH^{+}$			3.6×10^{-13}	1.3×10^{-9}[g]	

[a] Bath gas N_2 at 10 Torr. Temperature ≈ 402 K. All measurements obtained with drift voltage 5 V. [b] Gas-phase basicities in kcal/mol. From Lias and Hunter.[21] [c] Polarizabilities of bases A in (Å)³. From: Miller, K. J. *J. Am. Chem. Soc.* **1990**, *112*, 8543. Values given in parentheses were estimated with additivity rules. [d] Experimentally determined product of rate constant k and ion drift time = reaction time, t, from plots such as in Figure 1 and eq 10. Units are $cm^3{\cdot}molecule^{-1}$. Time t is the same for all reactions. [e] Reaction rate constants, $cm^3{\cdot}molecules^{-1}{\cdot}s^{-1}$. These are rough estimated values based on a reaction time $t = 140$ μs. This time was estimated on the basis of the time for reaction 2. This reaction is expected to proceed at collision rates $k_c = k_{ADO} = 1.3 \times 10^{-9}$ $cm^3{\cdot}molecule^{-1}{\cdot}s^{-1}$. Using $kt = 3 \times 10^{-16}$ $cm^3{\cdot}molecule^{-1}$, one obtains $t = 280$ μs. The mobility of the $NH_3(CH_2)_7NH_3^{2+}$ ion is expected to be roughly twice that of the butylammonium ion.[19] [f] Rough estimates of collision limit rate constants, evaluated with the Langevin expression $k_L = 2\pi e(\alpha/\mu)^{1/2}$ (for singly charged ions) where e = charge of electron, α is the polarizability of A, and μ is the reduced mass of the colliding pair of reactant molecules. $k_c = 2k_L$ was assumed, treating $NH_3(CH_2)_7NH_3^{2+}$ as two singly charged reaction centers. Results illustrate that the collision rates do not change significantly for the different neutral reactants. [g] $k = k_c$.

apparatus, because the ion concentration is extremely low, so that space charge effects are negligible. Experimental proof that the ion loss is first order is given by the fact that eq 6 is obeyed. Additional evidence was obtained by determining kt for two different bases A, such as triethylamine (TEA) and NH_3, by plots such as those shown in Figure 1. The experiments were then repeated for three different times t (by using V_d values of 10, 5, and 3 V). It was found that the ratio k_{TEA}/k_{NH_3}, obtained from the kt values, remained the same.

Reaction rates involving singly charged ions were also studied. Of interest was the following reaction:

$$n\text{-}C_4H_9NH_3^{+} + TEA = n\text{-}C_4H_9NH_2 + TEAH^{+} \qquad (7)$$

The value, $kt = 3.6 \times 10^{-13}$ cm^3/molecule, was obtained with $V_d = 5$ V, $T = 418$ K. Because the gas-phase basicity GB(TEA) is very much higher than GB(n-$C_4H_9NH_2$), this reaction is expected to proceed at collision rates. Using the polarizability $\alpha = 13$ A^3 and dipole moment $\mu = 1$ D for TEA, one obtains the following collision rates:[18] $k_c = 1.28 \times 10^{-9}$ cm^3 $molecule^{-1}$ s^{-1} (Langevin) and 1.30×10^{-9} cm^3 $molecule^{-1}$ s^{-1} (ADO) for reaction 11, at 420 K. Combined with $kt = 3.6 \times 10^{-13}$ cm^3 $molecule^{-1}$, this leads to a time $t = 277$ μs ($V_d = 5$ V).

All the rate data with BH_2^{2+} were also obtained at $V_d = 5$ V. The $t = 277$ μs value obtained for $C_4H_9NH_3^{+}$ can be used for rough estimates of the rate constants, k, of the BH_2^{2+} ions, from the corresponding kt data, on the basis of mobility determinations[19] of singly and doubly charged ions. See footnotes in Table 2.

III. Results and Discussion

(a) Diprotonated Diamines: Solvation and Deprotonation by NH_3. Thermochemical Results. The equilibrium constants K for the solvation of diprotonated diamines by NH_3

$$NH_3(CH_2)_pNH_3^{2+} + NH_3 = NH_3(CH_2)_pNH_3{\cdot}NH_3^{2+} \quad (0,1) \qquad (8a)$$

were determined for $p = 7$, 8, 10, 12.

Shown in Figure 2 is a plot of the equilibrium condition (eq 8b):

$$K_{0,1} = \frac{[BH_2{\cdot}NH_3^{2+}]}{[BH_2^{2+}][NH_3]} \qquad (8b)$$

$$\Delta G_{0,1}° = RT \ln K_{0,1} \qquad (8c)$$

where BH_2 stands for the diprotonated diamine. The plot which is for $p = 12$ shows that the ratio $[BH_2{\cdot}NH_3^{2+}]/[BH_2^{2+}]$ increases linearly with the pressure (concentration) of NH_3 and goes through the origin as required by eq 8b. Similar plots were obtained also for $p = 8$, 10, 12. For $p = 7$, the plot was less satisfactory. An approximately linear relationship was observed but only at very low NH_3 pressures ($p < 0.5$ Torr). The $p = 7$ case is special as will be discussed.

Free energy values $\Delta G_{1,0}°$ for $p = 7$, 8, 10, 12 were obtained with eq 8c. (Note: $\Delta G_{0,1}° = -\Delta G_{1,0}°$.) A sufficiently wide range of temperatures required to obtain reliable values also for $\Delta H_{1,0}°$ and $\Delta S_{1,0}°$ could be achieved only for $p = 10$ and 12. The thermodynamic data obtained are summarized in Table 1.

Free energy values $\Delta G_{1,0}°$ at an intermediate temperature, 382 K, are shown in Figure 3. Also given in this figure are the free energy values for the hydration reaction eq 4 obtained in previous work.[17] The binding free energies $\Delta G_{1,0}°$ for NH_3 are seen to be about 3 kcal/mol higher than those for H_2O. This is an expected change because the hydrogen bonding is known to increase with increasing basicity of the proton accepting base,[20] and NH_3 is a much stronger base than H_2O.

The free energy values are seen also to increase as the length of the $(CH_2)_p$ chain is decreased. This trend is clearly observed for the hydrates and was attributed[17] to the increasing effect of the second charge. This charge exerts a polarizing effect on the

(18) Su, T.; Bowers, M. J. *Int. J. Mass Spectrom. Ion Phys.* **1975**, *7*, 211.
(19) Wu, C.; Siems, W. F.; Klasmaier, J.; Hill, H. H. *Anal. Chem.* **2000**, *72*, 391.
(20) Davidson, W. R.; Sunner, J.; Kebarle, P. *J. Am. Chem. Soc.* **1979**, *101*, 1675.

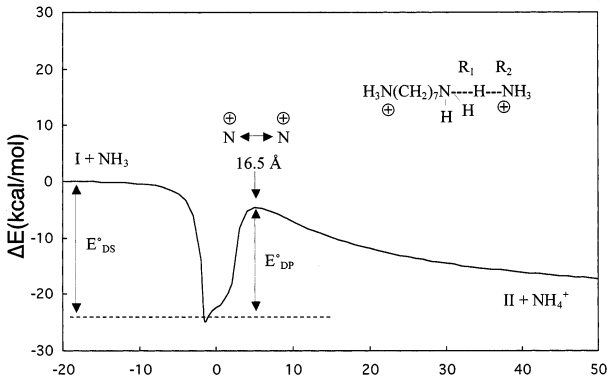


Figure 4. Potential energy surface obtained by ab initio calculations by Gronert[14a] for the following reaction: $NH_3(CH_2)_7NH_3^{2+} + NH_3 = NH_3(CH_2)_7NH_2^+ + NH_4^+$, where I in the figure is the doubly charged diamine, while II is the singly charged diamine in its extended linear configuration. E_{DS} and E_{DP} are the transition state energies for desolvation and deprotonation.

$(CH_2)_pNH_3^+$ group, making the N−H hydrogens more protic and thus more strongly hydrogen bonding. A similar effect is observed also with NH_3 as ligand. However, the free energy value for $p = 7$ does not fit the trend, being much lower than expected.

The potential energy diagram for the reaction of the $p = 7$ diprotonated diamine with NH_3, obtained by Gronert[14] from ab initio calculations, is shown in Figure 4. The exit channel to the right, which leads to deprotonation by NH_3, has an activation energy, $E_{DP}°$, which is lower than the activation energy, $E_{DS}°$, for the channel to the left, which leads to desolvation:

$$(E_{DS}° - E_{DP}°) \approx 4.6 \text{ (kcal/mol)} \quad \text{(Gronert)} \qquad (9a)$$

Thus, on the basis of these results, which correspond to the internal energy at 0 K, deprotonation is expected to dominate. For a more accurate prediction of the relative rates, one needs the free energies of activation, $\Delta G^{\ddagger}_{DP}$ and $\Delta G^{\ddagger}_{DS}$. Ab initio calculations of $\Delta G^{\ddagger}_{DP}$ can be expected to be very difficult. A qualitative estimate suggests that $\Delta S^{\ddagger}_{DS} > \Delta S^{\ddagger}_{DP}$. The transition state DS occurs at very large distances, $(R_1 - R_2) \approx -15$ Å, of the reactants (Figure 4). Both reactants will have three free external rotations each. The DP transition state occurs at a much shorter distance, $R_1 - R_2 \approx 5$ Å, where bonding between the reactants leads to a lowering of the energy. The bonding is largely due to attractive interactions between the ion (NH_4^+) and the dipole of the alkylamine group (see Figure 5 and its discussion in section c). This bonding leads to loose rocking vibrations whose entropy is lower than that of the free external rotations that they replace. The entropy difference $\Delta S^{\ddagger}_{DS} - \Delta S^{\ddagger}_{DP}$ could be as large as 10 cal/deg mol, or even larger. This would lead to $T\Delta S$ differences of some 3−5 kcal/mol:

$$\Delta S^{\ddagger}_{DS} - \Delta S^{\ddagger}_{DP} \approx 10 \text{ cal/deg·mol}$$

$$T(\Delta S^{\ddagger}_{DS} - \Delta S^{\ddagger}_{DP}) \approx 4 \text{ kcal/mol} \quad T = 420 \text{ K} \qquad (9b)$$

The observed anomalous, low value for $\Delta G_{1,0}°$ of BH_2^{2+}, $p = 7$, in Figure 3 must be a consequence of the rapid disappearance of $BH_2{\cdot}NH_3^{2+}$ by deprotonation, which is an irreversible reaction. The fact that some $BH_2NH_3^{2+}$ product was observed means that some collision complexes $(BH_2{\cdot}NH_3^{2+})^*$ are sufficiently long-lived to be collisionally stabilized by the third gas molecule. Some of the collisionally stabilized complexes, $BH_2{\cdot}NH_3^{2+}$, on thermal activation are expected to decompose via the irreversible DP channel. The rapid irreversible loss by deprotonation leads to a low concentration of observed $BH_2NH_3^{2+}$ and a low apparent $\Delta G_{1,0}°$.

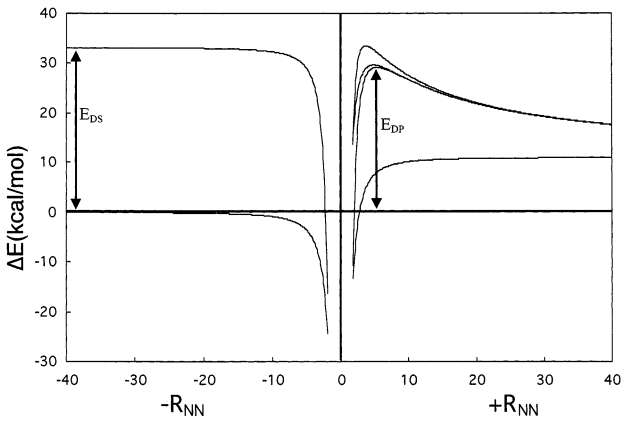


Figure 5. Potential energy diagrams and reactions associated with derived eqs 17−20. The highest maximum for the deprotonation reaction energy (lower right, curve) corresponds to only E_{POL} being considered in eq 17. The next lowest curve corresponds to only E_{DIP} being included, and lowest, to both E_{POL} and E_{DIP} being included. Energy scale on left side for upper plot; right side, lower plot.

(b) Diprotonated Diamines. Desolvation or Deprotonation. Results Based on Reaction Kinetics. As detailed in the Experimental Section, values for the product, *kt*, of the rate constant *k* and the time *t* that the reactant ion spends in the ion source, can be determined with the present apparatus. The time *t* is constant when the same reactant ion is involved.

Values for *kt*, obtained from plots such as that shown in Figure 1, for the $NH_3(CH_2)_pNH_3^{2+}$ ion (BH_2^{2+}, $p = 7$) reacting with different bases A at a constant $V_d = 5$ V, are given in Table 2 together with the gas-phase basicities, GB(A)'s, of the bases A. For all *kt* values given in Table 2, the only ionic product observed was that due to the deprotonation reaction; that is, no significant concentrations of stabilized adducts were present. All bases A which have a GB higher than that of NH_3, GB-(NH_3) = 196 kcal/mol, lead to a *kt* which is approximately equal, $kt = (8.6 \pm 0.5) \times 10^{-13}$ cm^3 molecule^{-1}. Such behavior is expected only when the GB(A) values are well above the GB^{app} of BH^+, where the rate constants *k* become equal to the collision rate constants k_c. The *kt* for NH_3 is about half as big as that for the other bases. This indicates that the GB^{app} is equal to $GB(NH_3)$:

$$GB^{app}(NH_3(CH_2)_7NH^{2+}) \approx GB(NH_3) = 196 \text{ kcal/mol} \qquad (10)$$

This is in good agreement with the result based on Gronert's calculations and entropy estimates which lead to $GB^{app} \approx 196$ kcal/mol (see eq 9b).

Attempts to determine the rates for proton transfer from BH_2^{2+}, $p = 7$, to bases A whose GB(A) < $GB(NH_3)$, were not successful. The available bases A in this range are ketones, such as acetone, methyl-ethyl ketone, and methyl-propyl ketone. These bases led to dominant formation of the (0,1) adducts, $BH_2^{2+}{\cdot}A$. Evidently, the much larger number of degrees of freedom in the collision complexes $(BH_2A)^{2+}$, relative to $(BH_2NH_3)^{2+}$, led to longer lifetimes and thus more efficient

collisional stabilization of the excited $(BH_2A)^{2+}$. Furthermore, the lower basicity of these bases A is also expected to lead to a longer lifetime of the collision complexes and thus also favor the stabilization of the complexes.

Williams and co-workers,[13] using an FTICR instrument, have also made determinations for the diprotonated diamines. The most extensive rate measurements were made for the $p = 7$ isomer.[13a] On the basis of such measurements, they chose the value

$$GB^{app}(BH^+, p = 7) \approx 182 \text{ kcal/mol}^{13a} \quad (11)$$

This is 14 kcal/mol lower than the value obtained from the kinetics determinations, eq 10, and the value expected from Gronert's results[16] (see Figure 4), which, with the reasonable assumptions about the entropies of activation (see eq 9b), indicates that GB^{app} should be close to $GB(NH_3) = 196$ kcal/mol.[21] The value of 182 kcal/mol was obtained[13a] by choosing bases A which led to rate constants $k \approx 0.02k_c$, a choice that is too far removed from the condition $k = 0.5k_c$.

(c) Generalization of Results for $NH_3(CH_2)_7NH_3^{2+}$ to Other Linear Diprotonated Ions and Polyprotonated Proteins. It is desirable to develop general and relatively simple methods with which the apparent gas-phase basicities of diamines with p other than $p = 7$, and other linear diprotonated systems, can be calculated and then to extend these to polyprotonated proteins.

Williams and co-workers, in a pioneering series of papers[13] involving both experimental and theoretical work, proposed the relationships

$$GB_{sp}{}^{app} \approx GB_{int} - \frac{q^2}{4\pi\epsilon_0\epsilon_r r} \quad (12a)$$

$$GB_{sp,}j^{app} \approx GB_{int,}j - \sum_{i=1}^{i=n} \frac{q^2}{4\pi\epsilon_0\epsilon_r r_{i,}j} \quad (12b)$$

Here, sp stands for the site of protonation, q is the elementary charge, ϵ_0 is the permittivity in a vacuum, and ϵ_r is the relative permittivity of the medium in which the charges interact. The distance between the charges is r. Equation 12a is for two charges while 12b applies to multiple charges with distances $r_{i,}j$ between the sites of protonation. GB_{int} stands for the (intrinsic) basicities of the sites of protonation, in the absence of all charges.

The evaluation of the apparent proton affinity PA^{app}, which corresponds to the proton affinity of the base that leads to the condition $E_{DP} = E_{DS}$, does not require a complete knowledge of the potential energy surface; only the energies of activation, E_{DS} and E_{DP}, are needed (see Figure 4). These energies are in regions where the bonding and repulsive forces can be evaluated by electrostatic equations. The approach used is illustrated in Figure 5. The upper curve gives the energy changes for the proton transfer in the singly protonated system.

$$H^+ + NH_2(CH_2)_7NH_3{}^+ + NH_3 = H^+ + (NH_2(CH_2)_7NH_3{\cdot}NH_3)^+ = H^+ + NH_2(CH_2)_7NH_2 + NH_4{}^+ \quad (13)$$

This potential energy curve needs not to be known and is used only to illustrate the model. The lower potential energy curve is for the reaction of interest:

$$NH_3(CH_2)_7NH_3{}^{2+} + NH_3 = (NH_3(CH_2)_7NH_3{\cdot}NH_3)^{2+} = NH_3(CH_2)_7NH_2{}^+ + NH_4{}^+ \quad (14)$$

Calling the distance between the two N atoms on each side of the proton R_{NN}, we note that for large absolute values of $(R_1 - R_2)$, in Figure 4, R_{NN} is close to equal to the absolute value of $(R_1 - R_2)$. The energy at the two end points of eq 14 relative to the zero level, see Figure 5, can be readily obtained. The end point on the right side of the diagram

$$R_1 = \infty \quad R_{NN} = \infty$$

corresponds to the energy change for the reaction

$$(NH_3(CH_2)_7NH_2)^+ + NH_3 + H^+ = (NH_3(CH_2)_7NH_2)^+ + NH_4{}^+ \quad (15)$$

$$E_{15} = PA(NH_2(CH_2)_7NH_2) - PA(NH_3) - PA(NH_2(CH_2)_7NH_2)$$

$$E_{15} = -PA(NH_3)$$

For the left side of the diagram

$$R_2 = \infty \quad R_{NN} = \infty$$

the energy corresponds to the energy change for the reaction

$$NH_2(CH_2)_7NH_3{}^+ + NH_3 + H^+ = NH_3(CH_2)_7NH_3{}^{2+} + NH_3$$

$$E_{16} = -PA(NH_2(CH_2)_7NH_3)^+ \quad (16a)$$

$$E_{16} = -PA(NH_2(CH_2)_7NH_2) + E_{REP} + E_{CD} \quad (16b)$$

The equality between eqs 16a and 16b was proposed by Gronert.[14c] The coulomb repulsion energy term E_{REP}

$$E_{REP} = \frac{q^2}{4\pi\epsilon_0 r}$$

makes the major contribution. For $r = 10$ Å, which corresponds to the distance between the two N atoms in $(NH_3(CH_2)_7NH_3)^{2+}$, $E_{REP} = 33$ kcal/mol. E_{CD} is a smaller correction and is due to a charge delocalization caused by the presence of the two charges. $E_{CD} \approx 7$ kcal/mol was obtained[14c] by a comparison with the ab initio results.[14a] Gronert did not consider in detail the actual charge distribution leading to the term E_{CD}. The charge distribution becomes even more relevant in the presence of more than two charges. Therefore, an analysis of the causes for this term is given in section d.

E_{16} provides a value for the desolvation energy E_{DS} relative to the zero level used in Figure 5. At values $R_1 - R_2 > 5$ Å, the reaction complex consists essentially of $^+NH_3(CH_2)_7NH_2$ and $NH_4{}^+$, and the energy due to the attractive interactions between these reactants can be approximated by the electrostatic bond energy between a charge located on the N atom of the $NH_4{}^+$ nitrogen and a dipole μ and polarizability α located on

(21) Hunter, E. P.; Lias, S. G. *J. Phys. Chem. Ref. Data* **1998**, *27*, 3.

the neutral amino group of the $^{+}NH_3(CH_2)_7NH_2$ reactant. The distance between the two N atoms is R_{NN}. Inclusion of this bond energy leads to

$$R_1 - R_2 > 5\ \text{Å} \quad (R_1 - R_2) \approx R_{NN}$$

$$E_{17} = -\text{PA}(NH_3) + E_{REP}(r + R_{NN}) + E_{POL}(R_{NN}) + E_{DIP}(R_{NN}) \quad (17)$$

where $E_{REP}(r + R_{NN})$ represents the Coulombic repulsion energy between the charges on $^{+}H_3N(CH_2)_7NH_3$ and NH_4^{+}

$$E_{REP}(r + R_{NN}) = \frac{q^2}{4\pi\epsilon_0(R_{NN} + r)} \quad (18)$$

and $E_{POL} + E_{DIP}$ are the energies due to the charge on NH_4^{+} and the polarizability and dipole on the NH_2 group of the protonated diamine.

The curve shown on the lower right side of Figure 5 was obtained with eq 17. The dipole moment used, $\mu = 1.3$ D, to model the dipole on the C−NH_2 group on the diamine corresponds to the literature value for the dipole of CH_3NH_2. This dipole was assumed to be located on the N atom, because most of the dipole is due to the NH_2 group, as evident from the value of $\mu = 1.4$ for NH_3. The polarizability used for the C−NH_2 group, $\alpha = 4 \times 10^{-24}$ cm^3, corresponds to the literature value for CH_3NH_2. The polarizability was located in the middle of the C−N bond, because the CH_3 group makes a substantial contribution to the total polarizability.

$$E_{DS} - E_{DP} = E_{17} - E_{16} = \text{PA}(NH_3) - \text{PA}(NH_2(CH_2)_7NH_2) + E_{REP} + E_{CD} - E_{REP}(r + R_{NN}\text{max}) - E_{POL}(R_{NN}\text{max}) - E_{DIP}(R_{NN}\text{max}) \quad (19)$$

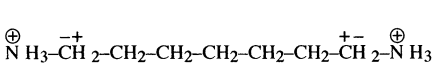
PA^{app} is obtained from eq 19 by setting $E_{DS} = E_{DP}$. The PA of the neutral base, that can either deprotonate or desolvate, is equal to PA^{app}. Setting $\text{PA}(NH_3) = PA^{app}(NH_3(CH_2)_7NH_2^{+})$ and approximately converting the equation into a free energy relationship by replacing the PA with the GB values and including a $T(\Delta S^{\ddagger}_{DS} - \Delta S^{\ddagger}_{DP})$ term, one obtains the final, general equation for a diprotonated molecule:

$$GB^{app} = GB_{int} - \frac{q^2}{4\pi\epsilon_0 r} + \frac{q^2}{4\pi\epsilon_0(R_{NN}\text{max} + r)} + E_{DIP}(R_{NN}\text{max}) + E_{POL}(R_{NN}\text{max}) + -E_{CD} + T(\Delta S^{\ddagger}_{DS} - \Delta S^{\ddagger}_{DP}) \quad (20)$$

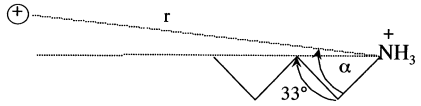
Using $GB_{int} = GB(NH_2(CH_2)_7NH_2) = 214$ kcal/mol,[22] $T(\Delta S^{\ddagger}_{DS} - \Delta S^{\ddagger}_{DP}) = 4$ kcal/mol (at ~420 K) (see eq 9b), and $E_{CD} = 7$ kcal/mol, the result, $GB^{app}(NH_3(CH_2)_7NH_2^{+}) \approx 196$, is obtained with eq 20 (for values of R_{NN}max, see Figure 5), in complete agreement with the value of the experimental determinations (see Table 2 and eq 8).

Only the first two terms in eq 20 are the same as those of eq 12a, which is from Williams and co-workers. These authors had qualitatively deduced (see Figures 2 and 3)[13c] a shape of the potential surface somewhat similar to that obtained in Gronert's work[14a] and Figure 5. However, they made the assumption that the much simpler eq 12a is sufficiently accurate for the purpose at hand.

Scheme 1



Scheme 2



The equation for polyprotonated proteins is obtained with eq 20, in a manner analogous to the approach used to obtain eq 12b from eq 12a. This leads to

$$GB^{app}_j = GB_{int}j + E_{DIP}j(R_{NN}j\text{max}) + E_{POL}j(R_{NN}j\text{max}) + T(\Delta Sj^{\ddagger}_{DS,} - \Delta Sj^{\ddagger}_{DP,}) - \int_{i=1}^{z-1} E_{CD}ij - \int_{i=1}^{z} \frac{q^2}{4\pi\epsilon_0 r_i j} + \int_{i=1}^{z} \frac{q^2}{4\pi\epsilon_0(r_i j + R_{NN}j\text{max})} \quad (21)$$

(d) Use of ab Initio Calculations Involving a Positive Point Charge for the Analysis and Evaluation of the Charge Delocalization Terms $E_{CD}ij$ and $\sum E_{CD}ij$ as well as R_{NN}max. For the $p = 7$ alkyldiamine, the correction term $E_{CD} = 7$ kcal/mol was needed. The correction depends also on the distances between the charges (see Gronert[14c] where that correction is included via an $\epsilon_r = 0.82$ in the electronic repulsion term). This correction cannot be applied directly when more than two charges are present. A more detailed examination of the two charge case proves useful.

The E_{CD} term should be mostly due to a destabilization whose origin is illustrated by Scheme 1.

The two dipoles, induced by the two proton charges, cause a destabilization, mainly because of repulsion between the charge on the one nitrogen and the opposing dipole near the other nitrogen. Such ion−dipole interactions act at relatively long distances.

The validity of this model was explored on the basis of ab initio calculated energies (see Appendix I) for the system shown in Scheme 2, involving a protonated *n*-butylamine and a ⊕ point charge. The butyl group was chosen to represent the molecular environment near one of the charges (that due to the protonation).

Both the distance r between the charges and the angle α relative to the N−C bond were varied. The results are given in Table 3. The ab initio evaluated energy, $\Delta E(\text{ab initio}) = E(\oplus\cdots C_4H_9NH_3^{+}) - E(C_4H_9NH_3^{+})$, was found to be greater than the charge repulsion coulomb energy, $E_{REP} = q^2/(4\pi\epsilon_0 r)$, at constant $r = 10$ Å for all angles α where the (presumed) induced dipole by the NH_3^{+} group was opposing the point charge. The maximum of $\Delta E(\text{ab initio}) - E_{CR}$ was found for $\alpha = 21.5°$, indicating that the polarizability of the N−C−C region is the major contributor to the induced dipole. The directly opposing position at $\alpha \approx 201.5°$ leads to a negative value (see Table 3), whose absolute value is somewhat smaller than that for $\alpha = 21.5°$, because the induced dipole is now somewhat

(22) $GB_{intr} \approx GB(H_2N(CH_2)_7NH_2)$, obtained from experimental GB determinations, is not available, because cyclization of the (singly) protonated diamine by intramolecular strong hydrogen bonding leads to higher values for GB and PA, see Yamdagni and Kebarle.[23] The value of $GB_{int} = 214$ kcal/mol is based on GB values for $C_nH_{2n+1}NH_2$, compounds for $n = 4-8$ available in Hunter and Lias.[21]

Table 3. Results from Positive Point Charge Calculations for Evaluation of E_{CD}

distance and angle[a]	(ΔE(ab initio) − E_{REP})[b]	ΔE(dipole model)[c]
10 Å; 10°	2.70	2.84
10 Å; 21.5°[d]	2.90	2.95
10 Å; 33° ($p = 7$)[e]	2.70	2.85
10 Å; 111.5°	−0.59	−0.31
10 Å; 201.5°	−1.60	−1.59
10 Å; 291.5°	−0.50	−0.31
12.7 Å; 33° ($p = 9$)[e]	1.80	1.69
15.2 Å; 33° ($p = 11$)[e]	1.20	1.12

[a] Butylamine with a point charge at the indicated distance from the nitrogen atom and angle with respect to the N−C bond. See Scheme 2 and text following Scheme 2. [b] Relative energies in kcal/mol at the B3LYP/6-311++G(d,p) level without zero point correction minus Coulombic repulsion using the point charge−nitrogen distance. [c] Ion dipole model fitted to computed values using an induced dipole with strength $\mu = 3.11$ D at a distance of 1.54 Å from the inducing charge on the nitrogen. The point charge, however, would experience it primarily as a permanent dipole because of its longer distance. For these values, a dipole axis was used that is 21.5° to the N−C bond. [d] The angle of the dipole axis was evaluated with several additional bracketing calculations (not shown). [e] The placement of the point charge corresponds to the position of the second nitrogen in $NH_3(CH_2)_pNH_3^{2+}$ ($p = 7, 9, 11$).

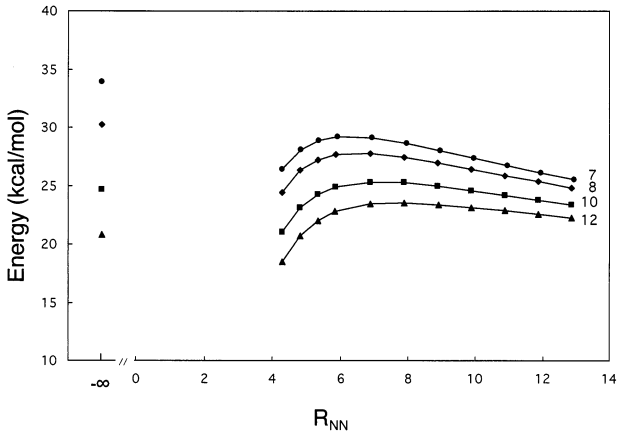


Figure 6. Potential energy surfaces from ab initio calculations for the model system $NH_3(CH_2)_pNH_2^+$, $p = 7$, 8, 10, 12, interacting with NH_4^+. $NH_3(CH_2)_pNH_2^+$ was modeled using methylamine and a point charge at the distance where the protonated nitrogen would be. Only a single point is given for E_{DS}, at $R_{NN} = -8$. The maxima of the curves shown correspond to E_{DP}.

farther away from the charge. The value is negative because now an attraction between the dipole and the charge is present.

The ΔE(ab initio) − E_{REP} values at $r = 10$ Å could be reproduced by an induced dipole with magnitude $\mu = 3.11$ D located at a distance of 1.54 Å from the nitrogen with the inducing charge (see electrostatic values ΔE (dipole model) in Table 3). For this distance, the induced dipole corresponds to a polarizability of 1.5 Å^3. This is a physically realistic value, because the polarizability of a C atom is of a similar magnitude, if one considers that the polarizability values in the literature are valid only for much larger distances between the inducing charge and the polarizable medium.

The angle $\alpha = 33°$ when the ⊕ charge would be aligned with the two nitrogens in a $NH_2(CH_2)NH_3^+$ molecule. The value of ΔE(dipole model) = 2.7 kcal/mol for this angle. Because there are two charges and two dipoles (see Scheme 1), the total correction term, E_{CD}, in eq 20 is predicted to be 5.4 kcal/mol, which is close to the more accurate $E_{CD} = 7$ kcal/mol value based on Gronert's results.[14] Thus, the model based on Schemes 1 and 2 recovers most of the expected correction.

Gronert[14c] has pointed out that the choice $\epsilon_r = 0.82$ is equivalent to $\epsilon_r = 1$ and a distance between the two N atoms $r = 8.2$ Å rather than the actual distance of 10 Å. Scheme 1 provides a rationale for the shorter distance because the two dipoles lead to a shift of the net positive charge corresponding to a smaller distance.

The cause for the apparent paradox of having to use $\epsilon_r < 1$ rather than $\epsilon_r \geq 1$ derives from eq 16b, which includes the intrinsic proton affinity based on experimental data. The experimental value includes the energy release due to the stabilization of the charge by the charge delocalization near the N atom. Using $\epsilon_r > 1$, as in some calculations[13,24] involving the experimental proton affinities, leads to an overcounting of this stabilization.

In the alkyldiamine, the induced dipoles are forced to align against the distant charges. In multiply protonated proteins, that is not necessarily the case. A full treatment would involve the determination of such an induced dipole on each molecular group close to each charge, including also neighboring side chains or backbone carbonyls which are hydrogen-bonded to the protonated side chain. The approach used in the present work, which includes some approximations, is described in Appendix II.

Equations 18−21 contain the term R_{NN}max. The evaluation of this term requires finding the maximum of the activation energy for the deprotonation, E_{DP}. Considering the large number of such terms (see eq 21), it is desirable to simplify the process. Shown in Figure 6 are results from positive point charge calculations where the approach was the same as that used to obtain the data for Table 3. However, butylamine was replaced with methylamine to simplify the calculations. These data demonstrate that a choice of $R_{NN} \approx 6.5$ Å provides a good approximation of E_{DP} for all four alkyl diamines, which extend from a distance $r = 10$ Å ($p = 7$) to $r = 16.5$ Å ($p = 12$). The approximation R_{NN}max = 6.5 Å was made in the calculations for the proteins (see Appendix II).

(e) Results for GB_j^{app} for Different Charged States of Three Proteins: Carbonic Anhydrase, Cytochrome *c*, and Pepsin. Significance of Results to Observed Charged States in ESI. Determinations for three selected proteins, carbonic anhydrase, CAII (MW = 29200), cytochrome *c* (MW = 12400), and pepsin (MW = 34600) were made. The gas-phase basicities GB_{intr} of the basic side chains selected, Arg, His, Lys, Trp, Pro, are given in Table 4. Justification for the values used is given in Appendix II. The values for GB_j^{app} of each protein were obtained with eq 21 (see also Appendix II).

The results obtained are summarized in Tables 5 and 6 for CAII and CYC (for data on pepsin, see Felitsyn[12]). The tables give GB^{app} for the side chain which is easiest to deprotonate when Z protons (charges) are present. When $GB_j^{app} - GB(NH_3) > 0$, the protein will be able to hold the assigned Z charges when NH_4^+ is the protonating reagent (as would be the case when ammonium acetate was used as buffer) or (which is equivalent) when the protonated site is solvated by a base such as NH_3. The charge Z, where $GB_j^{app} - GB(NH_3)$ is just above zero, corresponds to N_{SB}.

(23) Yamdagni, R.; Kebarle, P. *J. Am. Chem. Soc.* **1973**, *95*, 3504.
(24) Miteva, M.; Demirev, P. A.; Karshikoff, D. *J. Phys. Chem. B* **1997**, *101*, 9645.

Table 4. Gas-Phase Basicities[a]

	GB_{exp}[b]	$GB_{int,protein}$[c]
arginine	241	251
histidine	227	237
(imidazole)	(217)	
lysine	227	237
(*n*-butylamine)	(212)	
tryptophan	219	234
(indole)	(216)	
proline	212	227
ammonia	196	

[a] All values are in kcal/mol. [b] All values are from NIST database.[21] Compounds which model the amino acid side chains (in parentheses) are shown below the corresponding amino acid. The GB values for suitable models were found only for histidine, lysine, and tryptophan. [c] Estimated average intrinsic GB of amino acid in protein. The deviation for specific amino acids in the protein is estimated to be on the order of ±5 kcal/mol. Unusual protein arrangements may lead to larger deviations, see Appendix II.

Table 5. Carbonic Anhydrase (MW 29200)[a,b]

charges *z*	GB_j^{app}	$GB_j^{app} - GB_{NH_3}$	residue[c,d]
2	247	51	ARG182
3	243	47	ARG58
4	240	44	ARG89
5	228	32	HIS36
6	224	28	HIS3
7	220	24	LYS252
8	218	22	LYS45
9	215	19	LYS168
10	212	16	LYS154
Z_{obs}[e] 11	208	12	LYS113
Z_{CRM}[f] 12	205	9	LYS133
13	202	6	LYS39
N_{SB}[g] 14	200	4	LYS9
15	195	−1	LYS127
16	193	−3	LYS225
17	187	−9	LYS159
18	185	−11	LYS18
19	181	−15	LYS112
20	179	−17	LYS261

[a] All energies are given in kcal/mol. [b] Total of 36 basic sites considered. [c] Residue that is easiest to be deprotonated in the presence of a base such as NH_3. [d] The zinc center is considered to be $Zn(OH)^+$. [e] Z_{obs}, observed charge with highest intensity. [f] Charge provided by CRM from precursor droplet of same radius as protein, $R = 25$ Å (based on X-ray data). [g] Largest number of charges that available basic sites can hold.

As discussed in the Introduction, de la Mora[8] found that for the majority of globular proteins

$$Z_{obs} \approx Z_{CRM} \quad \text{(de la Mora}^{8}\text{)} \tag{22}$$

which leads to the requirement

$$N_{SB} \geq Z_{CRM} \tag{23}$$

It is important to examine whether the condition, eq 23, holds, because the assumption that it is N_{SB} that determines Z_{obs}, that is, $Z_{obs} \approx N_{SB}$, has also been made often. See, for example, ref 13.

The results obtained for carbonic anhydrase, Table 5, predict $N_{SB} = 15$, while $Z_{obs} \approx 11$ and $Z_{CRM} \approx 12$. Thus, in this case, de la Mora's proposal holds. For cytochrome *c*, Table 6, $N_{SB} \approx 10$, $Z_{obs} = 8$, and $Z_{CRM} = 9$; thus, the de la Mora condition holds also for this system.

Pepsin is an interesting case. It exhibits a large deviation to low Z_{obs} in the plot used by de la Mora.[8] It has only four strongly basic side chains, Lys 319, Arg 315, Arg 307, and His 53, at its surface and is not expected to be able to accommodate all the protons provided by the CRM process. The results (Table 4)[12] confirmed this expectation. It was estimated[12] that $Z_{CRM} \approx 15$, and $N_{SB} = 9$ was evaluated[12] with the equations derived in the present work. For this protein, N_{SB} would be limiting. Apparently, pepsin is very difficult to produce by electrospray,[25] and the result used,[12] $Z_{obs} \approx 10$ from Standing and co-workers,[25] was obtained at pH ≈ 3.5. Information on the solvent used and the possible presence of other small electrolyte ions such as NH_4^+ is not available.[25] This and the low ion intensities observed reduce the value of these results. We hope to be able to report in the future determinations of Z_{obs} under controlled solution and other experimental conditions.

Table 6. Cytochrome *c* (MW-12400)[a,b]

charges *z*	GB_j^{app}	$GB_j^{app} - GB_{NH_3}$	residue[c]
1	251.0	55.0	ARG38
2	247.2	51.2	ARG91
3	228.6	32.6	LYS7
4	225.1	29.1	LYS55
5	220.6	24.6	LYS13
6	215.6	19.6	LYS72
7	211.5	15.5	HIS26
Z_{obs}[d] 8	208.5	12.5	LYS53
Z_{CRM}[e] 9	203.3	7.3	LYS27
N_{SB}[f] 10	198.2	2.2	LYS99
11	194.6	−1.4	LYS22
12	191.6	−4.4	LYS87
13	184.8	−11.2	LYS79
14	178.4	−17.6	LYS60
15	175.5	−20.5	LYS8
16	169.6	−26.4	LYS39
17	163.7	−32.3	LYS100

[a] All energies are given in kcal/mol. [b] Total of 29 basic sites considered. [c] Residue that is easiest to be deprotonated in the presence of a base such as NH_3. [d] Z_{obs} = observed charge with highest intensity; however, $Z_{obs} = 9$ was also present, see Figure 9 in ref 12. A value of $Z_{obs} = 8$ was also obtained when pure water, without buffer electrolytes, was used (present work). [e] Charge provided by CRM from precursor droplet of same radius as protein, $R = 21$ Å (based on X-ray data). [f] Largest number of charges that available basic sites can hold.

Proteins produced by ESI in the negative ion mode lead more often to the condition $Z_{obs} < Z_{CRM}$. Recent work from this laboratory provides evidence that the condition, $N_{SB} < Z_{CRM}$, occurs more frequently in the negative ion mode.[26]

(f) Comparison with Experimental Determinations Involving Proton Transfer from Protonated Proteins. Unfortunately, most of the determinations of proton-transfer rates to bases of known GB involve denatured proteins.[13c,15a,b,e] A study by Smith and co-workers[15c] includes nondenatured CYC. The major purpose of this study was to compare the extent of deprotonation of a nondenatured and a denatured protein by a strong base. The denatured protein is expected to lose fewer protons because of the higher expected GB^{app} for the basic sites. In the looser, denatured structure, the protonated basic sites are farther apart because of the charge repulsion and their increased mobility. This is an interesting application of the GB^{app} concept; however, here we consider only the result[15c] for the nondenatured CYC. When the nondenatured CYC, $Z_{obs} \approx 8$, was exposed to triethylamine (TEA), the charge reduced to $Z_{obs} \approx 4$, corresponding to $N_{SB} \approx 4$ (TEA). A similar value is predicted

(25) Chernushevich, I. V.; Ens, W.; Standing, K. G. In *Electrospray Ionization Mass Spectrometry*; Cole, R. D., Ed.; Wiley: New York, 1997.
(26) Blades, A. T.; Peschke, M.; Verkerk, U.; Kebarle, P. *J. Phys. Chem. A*, in press.

by the present equations (see Table 6) when the literature GB(TEA) = 227.5 kcal/mol[21] is used. The side chains that are probably protonated are Arg38, Arg91, Lys7, and Lys55.

Actually, there is a paucity of reasonably accurate proton-transfer rate measurements involving nondenatured proteins. This is regrettable considering the importance of such data to the development of a gas-phase ion chemistry of proteins. For example, Williams and co-workers[13d] observed a maximum charge state of $Z = 11$ for nondenatured CYC sprayed from pure water and predicted $GB^{app} \geq GB(H_2O) = 158$ kcal/mol. This result could be checked using bases with GB higher than water. Similar checks of the very much higher values predicted by the present results, such as $GB^{app} = 208$ kcal/mol for $N_{SB} = 8$ (Table 6), would also be very useful.

Conclusions

1. Results predicting the activation energies E_{DS} and E_{DP} for $NH_3(CH_2)_7NH_3^{2+}$ can be obtained from simple and fast calculations based on electrostatics (Figure 5 and eqs 17−20). These lead to good agreement with the result from the ab initio calculation.[14] Furthermore, these equations can be used for any other two proton systems.

2. The ab initio calculations involving a positive point charge (see Scheme 2 and Table 3) represent a very convenient method for obtaining energy changes due to the position of the charge and then finding the electrostatic counterparts. This approach makes the use of an arbitrarily chosen relative permittivity ϵ_r unnecessary.

3. The electrostatic equations can be extended to multiply protonated globular proteins. This allows determinations of GB^{app} of the basic side chains and N_{SB}, the maximum number of protons that a globular protein can hold when the protonating agent is NH_4^+.

4. Comparison of N_{SB} for carbonic anhydrase (CAII) and cytochrome *c* (CYC), with the observed number of charges (protons) Z_{obs} and with Z_{CRM}, the number of charges that are expected to be provided by the water droplet that generates the charged protein, leads to $Z_{CRM} \approx Z_{obs} < N_{SB}$. This means that there are enough basic sites N_{SB}, and therefore, Z_{CRM} is the charge limiting factor. For pepsin, which has an unusually small number of basic sites, the finding is $Z_{obs} \approx N_{SB} < Z_{CRM}$. In this case, the protein cannot accommodate the available charge, Z_{CRM}, such that N_{SB} is the limiting factor. However, the Z_{obs} value was obtained from the literature, and the experimental conditions are not well-known.

5. The values for N_{SB} and Z_{CRM} given in Tables 5 and 6 are not as reliable as one would wish. Difficulties in the evaluation of GB^{app} are the following: (a) Values of intrinsic gas-phase basicities, GB_{int}, of the side chains of the proteins are somewhat arbitrary. These are not available and have to be modeled on the experimentally determined GB of the corresponding amino acids and assumed additional stabilization by neighboring groups (see Table 4). (b) The electrostatic calculations, while completely satisfactory for two proton systems, become complex for the polyprotonated proteins. For example, the reaction coordinate at a given basic site *j* depends on a vector superposition of repulsion forces from all protonated sites (see Appendix II). These problems become much more tractable for lower charge states. Experimentally determined GB^{app} values of such low charge states of proteins would be of great value for the further development of the modeling calculations.

6. One can expect that the evaluation of GB^{app} values could soon be made via MD simulations using force fields, such as GROMOS-87,[27] to model the proton transfer from the protonated site to an attached base, such as ammonia. The electrostatic equations for proton transfer, developed in this work, might prove useful for the inclusion of a force field for proton transfer in MD simulations.

Appendix I. Ab Initio Calculations for Alkyl Diamines Model, Equation 20. All calculations were performed using Gaussian94[29] with the B3LYP density functionals as the method of choice. This method is fast with relatively small computational hardware demands and, on the basis of previous experience with similar ion−molecule systems,[30] provides a relative energy accuracy of better than 5 kcal/mol. Because Gaussian94 does not, to our knowledge, allow geometry optimizations of molecular systems that include point charges, a point charge was fabricated using a "hydrogen atom" with a modified basis set description that made it energetically prohibitive to put an electron into the resulting orbitals. For all other atomic centers, the 6-311++G(d,p) basis set was used.

Appendix II. Extension of the Electrostatic Model to Multiply Charged Proteins. The basis for the evaluation of N_{SB} is eq 21. However, a number of approximations were made in order to obtain an easily workable model for the evaluation of GB_j^{app}.

(a) Values of $GB_{int,j}$. The values of $GB_{int,j}$ for each basic site are difficult to estimate. Experimental GB values for the basic amino acids are available in the literature and are listed in Table 4. However, these values have to be used with caution. Amino acids with long side chains such as lysine are able to stabilize a proton further by forming a hydrogen bond between the protonated side chain and the carbonyl group, and the experimental GB would include that stabilization. For example, the experimental GB for lysine is 227.3 kcal/mol. However, the side chain for lysine is butylamine with an experimental GB of 211.9 kcal/mol. The larger bulk of lysine would increase the GB of the NH_2 side chain analogous to octylamine, and a value of 214 kcal/mol might be appropriate. That leaves 13.3 kcal/mol in the measured GB for lysine for the expected hydrogen-bond stabilization. In a protein environment, additional stabilization from other nearby groups can be expected. Such stabilization can be anywhere from 5 to 15 kcal/mol over that from the first hydrogen-bond stabilization.

The GB_{int} values for arginine, histidine, lysine, tryptophan, and proline that have been used in the model are detailed in Table 4. For histidine and lysine, whose measured GBs of the amino acid are known to include one hydrogen-bond stabiliza-

(27) van Gunsteren, W. F.; Berendsen, H. J. C. *Groningen Molecular Simulation (GROMOS) Library Manual*; Biomos: Gronigen, The Netherlands. For applications of GROMOS to proteins in the gas phase, see also Miteva et al.[24] and Reitman et al.[28]
(28) Reitman, C. G.; Velasquez, I.; Tapia, O. *J. Phys. Chem. B* **1998**, *102*, 9344.
(29) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Gill, P. M. W.; Johnson, B. G.; Robb, M. A.; Cheeseman, J. R.; Keith, T.; Petersson, G. A.; Montgomery, J. A.; Raghavachari, K.; Al-Laham, M. A.; Zakrzewski, V. G.; Ortiz, J. V.; Foresman, J. B.; Cioslowski, J.; Stefanov, B. B.; Nanayakkara, A.; Challacombe, M.; Peng, C. Y.; Ayala, P. Y.; Chen, W.; Wong, M. W.; Andres, J. L.; Replogle, E. S.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Binkley, J. S.; Defrees, D. J.; Baker, J.; Stewart, J. P.; Head-Gordon, M.; Gonzalez, C.; Pople, J. A. *Gaussian 94*, revision D.3; Gaussian, Inc.: Pittsburgh, PA, 1995.
(30) (a) Smith, B. J.; Radom, L. *Chem. Phys. Lett.* **1994**, *231*, 231; 345. (b) Soliva, R.; Orozco, M.; Luque, F. J. *J. Comput. Chem.* **1997**, *18*, 980. (c) Bogdanov, B.; Peschke, M.; Tonner, D. S.; Szulejko, J. E.; McMahon, T. B. *Int. J. Mass Spectrom.* **1999**, *187*, 707.

tion, only 10 kcal/mol has been added to account for additional stabilizations. For arginine, to our knowledge, no GB has been measured for the side chain. However, because the side chain in arginine is also fairly long and flexible, it is highly likely that the amino acid GB measurement includes one hydrogen-bond stabilization. Therefore, again only 10 kcal/mol has been added.

For tryptophan and proline, the GB assignment is different. Both amino acids have inflexible or nonexistent side chains. The measured GB therefore would not include any hydrogen-bond stabilization. That can be shown by comparing the GB of indole, the side chain of tryptophan, with tryptophan itself. The increase of 3.4 kcal/mol is likely due to the increased charge delocalization in the amino acid. It certainly lacks the approximately 10 kcal/mol stabilization energy that a hydrogen bond would yield. Because both amino acids would be depending on other groups to be flexible enough to stabilize the charge, it is likely that on average both tryptophan and proline would be less stabilized than arginine, histidine, and lysine whose flexible side chains have a longer reach to find a more suitable protein environment. Consequently, as a reasonable guess, 15 kcal/mol was added to the measured GB to account for the first hydrogen bond and additional stabilization. All assigned GB_{int} values are shown in Table 4.

(b) Evaluation of $E_{DIP}j$ (R_{NN} max), $E_{POL}j$ (R_{NN}max), and $T(\Delta S^{\ddagger}_{DS} - \Delta S^{\ddagger}_{DP})$. For each protein considered, the X-ray crystal structure was taken as an approximation of its gas-phase structure. Both r_ij and $R_{NN}j$max are vector quantities. The R_ij direction is determined by the position of the charged nitrogens of each pair of amino acids considered. The $R_{NN}j$max direction was taken as the direction from the center of charge to the charged nitrogen of the protonated base j. The center of charge was based on all protonated side chains that were considered. For carbonic anhydrase and cytochrome *c*, all arginine, histidine, and lysine side chains were considered as sites that could be protonated, while for pepsin tryptophan and proline were also included. A more accurate approach toward finding the $R_{NN}j$ direction would be to use a superposition of force vectors from the individual electrostatic repulsions, but that would necessitate a trajectory calculation with many small steps because the initial direction of the force superposition does not generally hold once the charge has been moved any appreciable distance. The direction approximation using a center of charge, on the other hand, will yield good results if the charges are spread evenly across the surface of the protein. For the three proteins considered, the basic sites are spread out enough to allow such an approximation. Given the other uncertainties and approximations, the elaborate trajectory calculation is not justified at this stage.

Figure 6 shows that the energy near the maximum for E_{DP} does not vary greatly with R_{NN}, so that using a constant $R_{NN}j$max of 6.5 Å would be a reasonable approximation. $E_{DIP}(R_{NN}\text{max}) + E_{POL}(R_{NN}\text{ max}) \approx -4$ kcal/mol for the alkyl diamine (see Figure 5). The alkyl diamines model the situation where both side chains i and j are lysines. It was assumed that the same value holds also for all other pairs of side chains.

The term $T(\Delta S^{\ddagger}_{DS} - \Delta S^{\ddagger}_{DP}) \approx +4$ kcal/mol (at ~420 K) obtained for the diamine (see eq 9b) was assumed to be the same for all basic side chains. The terms for $E_{DIP} + E_{POL}$ and the term due to $T\Delta S$ are seen to cancel.

(c) Evaluation of the Terms $E_{CD}i,j$ and $\Sigma E_{CD}i,j$. In the alkyl diamine, the induced dipoles are forced to align against the distant charges. In proteins, that is not necessarily the case. A correct treatment would involve the determination of such an induced dipole on each molecular group close to each charge, including also neighboring side chains or backbone carbonyls which are hydrogen-bonded to the protonated side chain. However, several general observations can be made that facilitate estimating such complex energetic contributions. Each interaction with neighboring side chains increases the $GB_{int}j$ of the protonated side chain j while on the other hand also increasing $E_{CD}j$ because the positive end of the induced dipole will generally point into the protein toward the other charges. The correction would not be large because the dipole interactions fall off rapidly with $1/r^2$ and depend on charges being aligned with the dipole axis. For example, in cytochrome *c* where lysine predominates, calculating the induced dipole repulsions (μ = 3.11 D) centered on the first carbon next to nitrogen of the side chain of lysine using an NMR structure, without taking the dipole alignment ($\theta = 0$ in all cases) into account, the overall correction, $E_{CD}j = \Sigma E_{CD}i,j$, for 13 charges, varies from 9 to 13 kcal/mol depending on each site j. Inclusion of the dipole alignment by using a $\cos(\theta)$ term (θ is the angle between C−N bond axis and N−N line) reduces the correction to 1−6 kcal/mol with 4 kcal/mol as the average.

A similar correction would be expected for other nearby polarizable groups such as hydrogen-bonded carbonyl groups**.** Using the polarizability of acetone, the induced dipole at hydrogen-bond distances is 2.3 D, weaker than the induced dipole on the adjacent carbon atom. However, the polarizability of backbone carbonyls tends to be higher than that of acetone so that the induced dipole might be expected to add corrections of 2−3 kcal/mol per hydrogen-bonded interaction on average if 13 charges are present. If 3 hydrogen-bonded interactions are assumed then for 13 charges, $E_{CD}j = 12$ kcal/mol might be a reasonable estimate. However, for only 2 charges, the E_{CDj} correction would be much lower because of the longer average distance between the two charges and the probability that the dipoles are not aligned. Using the CYC NMR structure for 2 charges, corrections between 0.4 and 0.9 kcal/mol are obtained for the carbon induced dipole interaction. So, an estimate of 1 kcal/mol for 2 charges might be reasonable if the polarizable carbonyl group interactions are included. Using a linear relationship for $\Sigma E_{CD}i_{,j}$ leads to $E_{CD}i,j = 1$ kcal/mol beyond the first charge. This approximation assumes similar induced dipoles for each basic site so that on average each charged basic site induces identical dipoles. Given the small magnitude of the correction, treating each basic site in an identical manner for this correction is not expected to lead to significant errors.

Including the above-mentioned cancellations in eq 21 leads to eq 24, which forms the basis of the model. The model starts

$$GB^{app} = GB_{int} - \sum \frac{q^2}{4\pi\epsilon_0 r_{i,}j} + \sum \frac{q^2}{4\pi\epsilon_0 (r_{i,}j + R_{NN,}j\text{max}) - \sum E_{CD}i,j} \quad (24)$$

with all basic sites that have been included as charged. The number of bases used for each protein is given in Tables 5 and

6. Next, a visual inspection of the nitrogen that carries the charge is done on the X-ray or NMR structure to see if all charges are on the surface of the protein. Charges that are imbedded in the protein are excluded (His 64 in carbonic anhydrase and Pro 41, Trp 39, Trp 190 for pepsin). Using the assigned GB_{int} values from Table 4, the value for $\sum q^2/4\pi\epsilon_0 r_i j - GB_{int}$ is calculated. Each charge is then moved 6.5 Å away from the center of charge to calculate the term $\sum q^2/4\pi\epsilon_0(r_i j + R_{NN,j}\mathrm{max})$. The site with the lowest GB^{app} is then assumed to deprotonate. This cycle is then repeated for the remaining charges. The results are shown in Tables 5 and 6. If ammonia is the deprotonating agent, the number of charges where $GB^{app} - GB(NH_3)$ becomes positive corresponds to N_{SB}.

A global search that considers all possible combinations of protonated sites for each charge state would be preferable. However, such a search is likely to improve the gas-phase basicity of the most basic protein sites for each charge state by only a few kilocalories per mole over the current method because such an optimization would only change long distance electrostatic interactions. Also, the number of permutations of possible protonated sites at charge state 10 for cytochrome *c*, where 29 bases are considered, is over 20 million. Given that the expected improvement is less than the error bars, the additional time needed to perform a global search is currently not warranted.

JA012591E