Theoretical and Experimental Studies of the Zwitterion ⇌ Neutral Form Equilibrium of Ampholytes in Pure Solvents and Mixtures

Peter I. Nagy*,[†] and Krisztina Takács-Novák[‡]

Contribution from the Department of Medicinal and Biological Chemistry, The University of Toledo, Toledo, Ohio 43606-3390, and Institute of Pharmaceutical Chemistry, Semmelweis University of Medicine, H-1092, Budapest, Hungary

Abstract: The equilibrium ratios for the zwitterionic and neutral forms of the nicotinic and isonicotinic acids (3- and 4-COOH pyridines) have been determined by spectrophotometry in mixtures of water with methanol, tetrahydrofuran and dioxane. The zwitterionic form of the solute is the prevalent structure in mixtures with at least 60% water. In pure organic solvents, the neutral form was found in almost 100%. Ab initio calculations at the MP2/6-311++G**//MP2/6-31G* level predict that the zwitterion is higher in free energy by 33−35 kcal/mol than the neutral tautomer at $T = 298$ K and $p = 1$ atm in the gas phase. In contrast, relative solvation free energies, as calculated by the free energy perturbation method in Monte Carlo simulations, are −36, −27, and −15 kcal/mol in water, methanol, and tetrahydrofuran, respectively. The solvent effects for overbalancing the relative internal free energy of the zwitterion and stabilizing this tautomer are large enough only in water and in aqueous solution with water fraction above 60%. Predominance of the neutral form in pure organic solvents is due to considerably lower internal energy of this structure as compared to that of the zwitterion.

Introduction

The outstanding role of lipophilicity in biological activity of drug molecules is well-known and has attracted much interest.[1] Lipophilicity is usually characterized by the partition coefficient (log *P*) related to the partitioning equilibrium of a solute molecule between water and an immiscible lipid-like organic solvent, most frequently *n*-octanol.[1a]

The definition of log *P* referring to the concentration ratio of the "same molecular form" (nonionized and monomer species) of the drug in the two solvents is evident for simple acids and bases.[1b] However, definition of lipophilicity becomes more complicated when the solute can exist in two or more interconverting species. Takács-Novák et al. have recently considered the case of ampholyte molecules[2] with overlapping protonation.[3] Ampholytes, having two protonation sites in the simplest case, can exist in four microforms presenting cation, anion, and forms of zero net-charge (referred to as neutral (n) and zwitterionic (zw)) in aqueous solution. For many of ampholytes where the two pK_a = log *K* values differ by less than 4 units, acid−base characterization by protonation macroconstants (*K*) does not suffice. These constants describe the overall basicity of the molecule but cannot be assigned to the individual functional groups. In this type of ampholytes, knowledge of protonation microconstants is required for the
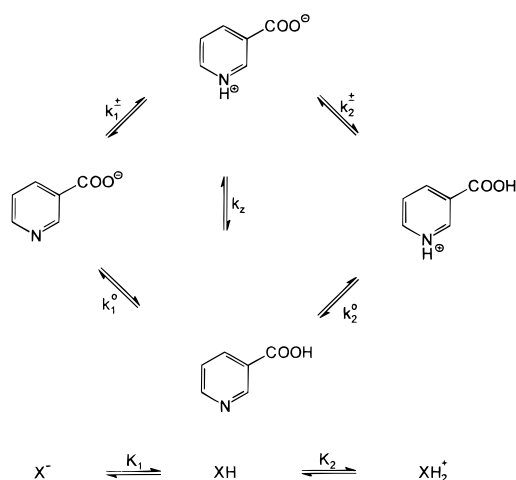

Figure 1. Protonation macro- and microequilibria of NA.

complete characterization of the protonation equilibria[4] (Figure 1). These microconstants play important role in other physicochemical processes, too. A formula has been recently derived[2b] relating log *P*, the true or micropartition coefficient of the neutral form to the experimentally obtainable log P_{app} and protonation microconstants

$$\log P = \log P_{app} + \log\left(1 + \frac{1}{k_1{}^0[\text{H}^+]} + \frac{k_2{}^0}{k_2{}^\pm} + k_2{}^0[\text{H}^+]\right) \quad (1)$$

where $k_1{}^0$, $k_2{}^0$, and $k_2{}^\pm$ are the protonation microconstants (see Figure 1).

The derivation is based on the assumption that only the neutral species is the partitioning tautomer even when the zwitterionic

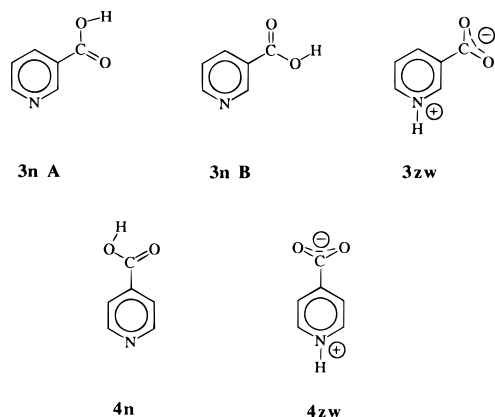[†] The University of Toledo.
[‡] Semmelweis University of Medicine.
[⊗] Abstract published in *Advance ACS Abstracts,* May 15, 1997.
(1) (a) Leo, A.; Hansch, C.; Elkins, D. *Chem. Rev.* **1971**, *71*, 525 and references therein. (b) Van der Waterbeemd, H.; Testa, B. In *Advances in Drug Research*; Testa, B., Ed.; Academic Press: New York, 1987; Vol. 16, p 85.
(2) (a) Takács-Novák, K.; Józan, M.; Hermecz, I.; Szász, G. *Int. J. Pharm.* **1992**, *79*, 89. (b) Takács-Novák, K.; Józan, M.; Szász, G. *Int. J. Pharm.* **1995**, *113*, 47. (c) Takács-Novák, K.; Avdeef, A.; Box, K. J.; Podányi, B.; Szász, G. *J. Pharm. Biomed. Anal.* **1994**, *12*, 1369. (d) Takács-Novák, K.; Szász, G.; Budvári-Bárány, Zs.; Józan, M.; Löre, A. *J. Liquid Chromatogr.* **1995**, *18*, 807.
(3) (a) Takács-Novák, K.; Tökés-Kövesdi, M.; Szász, G.; Noszál, B. *Magy. Kém. Foly.* **1994**, *100*, 538. (b) Takács-Novák, K.; Noszál, B.; Tökés-Kövesdi, M.; Szász, G. *J. Pharm.Pharmacol.* **1995**, *47*, 431.

(4) (a) Bjerrum, J. *Z. Phys. Chem.* **1923**, *106*, 209. (b) Noszal, B. *J. Phys. Chem.* **1986**, *90*, 4104. (c) Takács-Novák, K.; Noszál, B.; Hermecz, I.; Kereszturi, G.; Podanyi, B.; Szász, G. *J. Pharm. Sci.* **1990**, *79*, 2023.

Scheme 1

3n A   **3n B**   **3zw**

4n   **4zw**

Figure 2. UV spectra of NA and INA in methanol/water (Britton−Robinson buffer) solvent system.

form is the predominant one in the aqueous phase. Applications of this formula for investigating the lipophilicity of antibacterial fluoroquinolones[2a] and other ampholytes[2b,d] supported this idea. Nonetheless, a theoretical explanation for the success of this model has not been provided yet. Our goal in this paper is to perform calculations for studying the thermodynamics of the zwitterion ⇌ neutral form equilibrium in pure solvents and mixtures and comparing theoretical results with experiment. The results may give a rationale for the assumption above. Two simple ampholytes as nicotinic acid (NA) and isonicotinic acid (INA) (Scheme 1) have been selected for testing.

Despite the importance of the problem, there have been relatively few theoretical calculations regarding zwitterionic systems. Tautomerism of the neutral and zwitterionic forms and their hydrates containing a small number of water molecules has been investigated for several amino acids in the gas and/or condensed phases with different methods.[5] Aqueous solution of glycine was studied in Monte Carlo simulations[6a−c] and with the supermolecule/continuum approach.[6d] Energy of the zwitterion formation in the gas phase and in solution for the $^{+}H_3N\cdots SO_3^{-}$ system was calculated by Wong et al.[7] using ab initio methods and a continuum solvent model.

When this work was finished, Halle et al.[8] published their results for the protonation microconstants of NA and INA determined in water/dimethyl sulfoxide mixtures. Thus, our experiments in mixtures of water with methanol, tetrahydrofuran, and dioxane provide possibility to estimate the effect of the organic solvent in determining these physicochemical parameters. Also, this is the first attempt to calculate tautomeric equilibrium in mixed solvents by combining ab initio and Monte Carlo methods.

(5) For example, see: (a) Voogd, J.; Derissen, J. L.; van Duijneveldt, F. B. *J. Am. Chem. Soc.* **1981**, *103*, 7701. (b) Bonaccorsi, R.; Palla, P.; Tomasi, J. *J. Am. Chem. Soc.* **1984**, *106*, 1945. (c) Langlet, J.; Caillet, J.; Evleth, E.; Kassab, E. In *Studies in Physical and Theoretical Chemistry*; Revail, J.-E., Ed.; Elsevier: Amsterdam, 1990; Vol. 71, p 345. (d) Ramek, M.; Cheng, V. K.; Frey, R. F.; Newton, S. Q.; Schäfer, L. *J. Mol. Struct., (THEOCHEM)* **1991**, *245*, 1. (e) Yu, D.; Armstrong, D. A.; Rauk, A. *Can. J. Chem.* **1992**, *70*, 1762. (f) Hu, C.-H.; Shen, M.; Schaefer, H. F., III. *J. Am. Chem. Soc.* **1993**, *115*, 2923. (g) Kikuchi, O.; Matsuoka, T.; Sawahata, H.; Takahashi, O. *J. Mol. Struct. (THEOCHEM)* **1994**, *305*, 79. (h) Csaszar, A. *J. Mol. Struct. (THEOCHEM)* **1995**, *346*, 141. (i) Jensen, H. J.; Gordon, M. S. *J. Am. Chem. Soc.* **1995**, *117*, 8159 (j) Gronert, S.; O'Hair, R. A. *J. Am. Chem. Soc.* **1995**, *117*, 2071. (k) Ding, Y.; Krogh-Jespersen, K. *J. Comput. Chem.* **1996**, *17*, 338.

(6) (a) Mezei, M.; Mehrotra, P. K.; Beveridge, D. L. *J. Biomol. Struct. Dyn.* **1984**, *2*, 1. (b) Alagona, G.; Ghio, C.; Kollman, P. A. *J. Mol. Struct. (THEOCHEM)* **1986**, *166*, 385. (c) Alagona, G.; Ghio, C. *J. Mol. Liquids* **1990**, *47*, 139. (d) Rzepa, H. S.; Yi, M. Y. *J. Chem. Soc., Perkin Trans. 2* **1991**, 531.

(7) Wong, M. W.; Wiberg, K. B.; Frisch, M. J. *J. Am. Chem. Soc.* **1992**, *114*, 523.

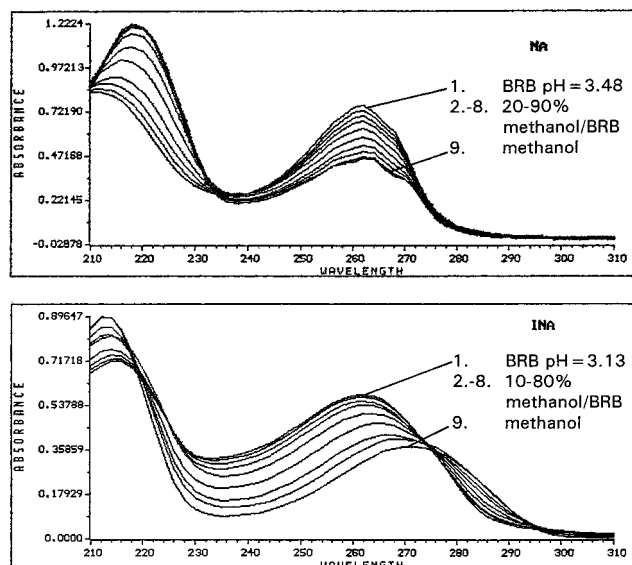(8) Halle, J.-C.; Lelievre, J.; Terrier, F. *Can. J. Chem.* **1996**, *74*, 613.

Experimental Section

Materials. Nicotinic acid was of pharmacopeial grade (Pharmacopoea Hungarica VII), and the sample of isonicotinic acid was prepared from isonicotinyl hydrazide by alkaline hydrolyses and then crystallized from hot water. The purity of the product was controlled by melting point measurement (319 °C) and TLC method. *n*-Octanol, methanol, and 1,4-dioxane were of HPLC grade (Aldrich, 99%), and tetrahydrofuran was of LiChrosolv, Merck. The Britton−Robinson buffer (BRB) (acetic, phosphoric, and boric acids, each at 0.04 M, treated with 0.2 M sodium hydroxide) was used to adjust the pH to the isoelectric point (iep) of the examined molecules. All other reagents were of analytical grade.

Methods. Protonation macroconstants (K_1 and K_2) were taken from previous work.[9] The protonation microconstants, were calculated using macroconstants, and the k_z tautomerization microconstant was determined by UV spectroscopy for the equilibrium XH^0(neutral) ⇌ $XH^{\pm}$-(zwitterion). The experimental method is based on the spectral differences between the zwitterionic form of the molecule (found as predominant in aqueous solution at the pH of the iep) and the neutral form (predominant in organic solvents of lower dielectric constant, such as methanol, THF, and dioxane). The spectra can be converted from one to the other by changing the composition of the mixture. The approach was originally described by Metzler and Snell[10] and has been developed and refined by Takács-Novák et al. recently.[2c,3a,9] A stock solution (∼3 mM) of the sample (NA or INA) was prepared in BRB solution at the relevant iep pH. This solution (1.00 mL) was diluted to 25.00 mL with BRB solution or with organic solvent (methanol, THF, or dioxane) or with various mixtures of the buffer and organic solvent. The concentration of the organic solvent was changed from 0% to 100%. The k_z constants were calculated based on data from absorption spectra for these solutions (Figure 2, spectra with the THF solvent are provided as Supporting Information).

The apparent partition coefficient has been determined in octanol/BRB solution at the iep pH of NA (3.48) and INA (3.13) at 25.0 ± 0.01°C, using shake flask technique.[2b] The log P_{app} value is an average of 12 parallel measurements. Experimental log P_{app} values were converted to the true partition coefficient using eq 1. (Table 1).

Calculations

Ab initio calculations have been carried out for the neutral and zwitterionic forms of NA and INA (structures **3** and **4**, respectively, in Scheme 1) considering two neutral conformers for the 3-COOH isomer. All structures were assumed to be planar on the basis of previous studies

(9) Takács-Novák, K.; Kökösi, J.; Podányi, B.; Noszál, B.; Tsai, R. S.; Lisa, G.; Carrupt, P. A.; Testa B. *Helv. Chim. Acta* **1995**, *78*, 553.

(10) Metzler, D.; Snell, E. E. *J. Am. Chem. Soc.* **1955**, *77*, 2431.

Table 1. Protonation Macro- and Microconstants, k_z Tautomerization Microconstant, and Apparent and True Partition Coefficients of NA and INA

compd	log K_1	log K_2	log k_z [a]	log k_1^0	log $k_1^\pm$	log k_2^0	log $k_2^\pm$	log P_{app} [b]	log P [b]
NA	4.82	2.14	1.47 ± 0.05	3.3	4.8	3.6	2.1	−0.67 ± 0.03	1.04
INA	4.38	1.88	1.90 ± 0.04	2.4	4.3	3.8	1.8	−1.00 ± 0.08	1.00

[a] Average log k_z value obtained from water/methanol and water/dioxane mixtures. [b] In octanol/water system.

Table 2. Relative Energies and Free Energies in the Gas Phase[a]

	E(zw) − E(n)		
		NA	
	INA	A	B
HF/6-31G*//HF/6-31G*	36.86	37.56	37.80
HF/6-31G*//MP2/6-31G*	36.75	37.35	37.59
HF/6-31+G**//HF/6-31+G**	35.73		
HF/6-31+G**//MP2/6-31G*	35.26		
HF/6-311++G**//HF/6-311++G**	35.74		
HF/6-311++G**//MP2/6-31G*	35.43	36.02	36.31
MP2/6-31G*//HF/6-31G*	34.80	33.39	33.60
MP2/6-31G*//MP2/6-31G*	34.85	33.55	33.75
TS1-zw	5.69	7.45	
TS2-neu	11.37	12.73	
MP2/6-31+G**//HF/6-31+G**	32.53		
MP2/6-31+G**//MP2/6-31G*	32.57		
MP2/6-311++G**//HF/6-31G*		33.14	33.39
MP2/6-311++G**//HF/6-311++G**	34.51		
MP2/6-311++G**//MP2/6-31G*	34.56	33.10	33.35
$0.9\Delta ZPE + \Delta H(298) - 298\,\Delta S(298)$			
HF/6-31G*	0.36	0.11	0.10
MP2/6-31G*	0.50	0.21	0.24
$\Delta G_{gas}(298)$	34.92	33.21	33.45

[a] Values in kcal/mol.

for the benzoic acid.[11] Ab initio calculations were performed using the Gaussian 92[12] package running on a Cray Y-M8 computer at the Ohio Supercomputer Center. The 6-31G*, 6-31+G**, and 6-311++G** basis sets were used at the Hartree-Fock and MP2 levels[13a] in geometry optimizations and single-point energy calculations (Table 2). Normal frequencies were calculated for optimized structures in Scheme 1 at the HF/6-31G* and MP2/6-31G* levels. All **3** and **4** tautomers correspond to local energy minima at both levels. Zero-point energies (ZPE, with HF/6-31G* and MP2/6-31G* frequencies scaled by factors 0.9[13a] and 0.97,[13b,c] respectively) and thermal corrections at 298 K and 1 atm were calculated in the rigid rotator−harmonic oscillator approximation.[14] Relative free energies were calculated as ΔG_{gas} (298, 1 atm) $= \Delta E(0) + 0.9\Delta ZPE + \Delta H(298) - 298\Delta S(298)$.

Transition states were identified at the MP/6-31G* level. Starting from optimized geometries for **3** and **4**, the carboxylate and the carboxylic groups (taking the anti form for the latter) were rotated by 90° to become perpendicular to the ring plane. Removing the ring planarity constraint two first-order saddle point structures (TS1 and TS2) were determined for each of INA and NA. A TS structure was identified upon its one imaginary frequency.

Solvent effects on the tautomeric equilibria were considered by Monte Carlo simulations using the free energy perturbation method.[15] Jorgensen's BOSS program[16] versions 3.1 and 3.4 were ported to DEC

Alpha OSF/1 and SUN SPARC II computers running at the University of Toledo. Isothermal−isobaric (NpT) ensembles were studied at $T = 298$ K and $p = 1$ atm as described by Jorgensen et al.[17] Simulations were performed in pure water, methanol, and THF, as well as in water/methanol and water/THF mixtures.

The single solute molecule was placed in the center of the solvent box containing 502 water, 262 methanol, or 262 THF solvents. The TIP4P water model[18] and the three- and five-point models for methanol and THF, respectively, were provided by the BOSS program. The 12-6-1 OPLS (optimized potential for liquid simulations)[19] potential was applied for calculating solvent−solvent and solute−solvent interactions. Corresponding cutoff radii were set to 8.5, 11, and 14 Å and 12, 13, and 16 Å in water, methanol, and THF, respectively, and intermediate values were taken in mixtures. Simulations were subject to periodic boundary conditions and preferential sampling. Solute move and volume change were attempted after every 50 and 1000 steps, respectively. Configurations at 4500 and 6000 K were considered in the equilibration and averaging phases, respectively.

Steric (12-6) parameters, referring to structurally similar atoms, were taken from the program's library. Hydrogens connected to the aromatic ring were considered in the CH united atom model. Different sets of the atomic charge parameters were considered in order to reproduce experimental equilibrium constant for the zwitterion ⇌ neutral form tautomerization of INA in water. Atomic charges developed in this manner take into consideration, at least partially and on average, the effects of polarization, pH, ionic strengths, etc., not considered explicitly in the simulations. Charges in the first two sets (Table 5, Supporting Information) are generally based on OPLS values, e.g., charges for the neutral and protonated ring nitrogen were set to the corresponding OPLS values for histidine. Quantum chemical study based parameters were obtained with the CHELPG procedure[20a] of Gaussian 92 by fitting atomic charges to the HF/6-31G* or MP2/6-31G* electrostatic potentials. Electrostatic interactions beyond the solute−solvent cutoff radius were accounted for by the Onsager reaction field contribution.[21] In obtaining this term, we used dielectric constants 78.3, 32.66, and 7.58 for water, methanol, and THF, as provided by the BOSS program. For mixtures the components contributed to the final ϵ value in proportion to their volume fraction in the mixture. In transforming the geometric and simulation parameters, λ, the linear interpolation of the form $\lambda(x) = x\lambda(1) + (1 - x)\lambda(0)$ was used with $x = 0,1$ at the endpoints, i.e., with the real molecular structures. Double-wide sampling was applied with λ changes assuring free energy increments of about 1 kcal/mol or less.

Simulations with mixed solvents have been performed for the 80% water/20% methanol, 60% water/40% methanol, and 80% water/20% THF systems (concentration in v mL/100 mL solution). Our experiments showed that volume decrease upon mixing is up to 2.6%. The theoretical calculations are not supposed to reach this accuracy, and the volume contraction was disregarded in the simulations. Models

(11) Nagy, P. I.; Smith, D. A.; Alagona, G.; Ghio, C. *J. Phys. Chem.* **1994**, *98*, 486.

(12) Gaussian 92, Revision A. Frisch, M. J.; Trucks, G. W.; Head-Gordon, M.; Gill, P. M. W.; Wong, M. W.; Foresman, J. B.; Johnson, B. G.; Schlegel, H. B.; Robb, M. A.; Replogle, E. S.; Gomperts, R.; Andres, J. L.; Raghavachari, K.; Binkley, J. S.; Gonzalez, C.; Martin, R. L.; Fox, D. J.; Defrees, D. J.; Baker, J.; Stewart, J. J. P.; Pople, J. A.; Gaussian Inc.: Pittsburgh, PA, 1992.

(13) (a) Hehre, W. J.; Radom, L.; Schleyer, P. v. R.; Pople, J. A. *Ab Initio Molecular Orbital Theory*; John Wiley and Sons: New York, 1986. (b) Nagy, P. I.; Durant, G. J.; Smith, D. A. *J. Am. Chem. Soc.* **1993**, *115*, 2912. (c) Nagy, P. I.; Ulmer, C. W., II; Smith, D. A. *J. Chem. Phys.* **1995**, *102*, 6812.

(14) McQuerrie, D. *Statistical Mechanics*; Harper and Row: New York, 1976.

(15) Jorgensen, W. L.; Ravimohan, C. *J. Chem. Phys.* **1985**, *83*, 3050.

(16) Jorgensen, W. L. BOSS, Version 3.4; Biochemical and Organic Simulation System User's Manual, Yale University, New Haven, CT, 1993.

(17) (a) Jorgensen, W. L.; Madura, J. D. *J. Am. Chem. Soc.* **1983**, *105*, 1407. (b) Jorgensen, W. L.; Swenson, C. *J. Am. Chem. Soc.* **1985**, *107*, 1489. (c) Jorgensen, W. L.; Gao, J. *J. Phys. Chem.* **1986**, *90*, 2174.

(18) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926.

(19) Jorgensen, W. L.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1988**, *110*, 1657.

(20) (a) Breneman, C. M.; Wiberg, K. *J. Comput. Chem.* **1990**, *11*, 361. (b) Besler, B. H.; Merz, K. M.; Kollman, P. A. *J. Comput. Chem.* **1990**, *11*, 431.

(21) (a) Onsager, L. *J. Am. Chem. Soc.* **1936**, *58*, 1486. (b) Beveridge, D. L.; Schnuelle, G. W. *J. Phys. Chem.* **1975**, *79*, 2562.

(22) Kassimi, N. E. B.; Doerksen, R. J.; Thakkar, A. J. *J. Phys. Chem.* **1995**, *99*, 12790.

for mixed solvents were comprised of 452 water + 50 MeOH, 387 water + 115 MeOH, and 476 water + 26 THF molecules. Monte Carlo simulations with these solvents followed those described above for the pure solvents.

Results and Discussion

Gas Phase. Optimized geometric parameters are provided as Supporting Information. Geometry optimizations for the zwitterion at the HF/6-31+G** and HF/6-311++G** levels led to changes, as compared to those obtained with the HF/6-31G* basis set, in bond lengths and bond angles of the heavy atoms up to 0.005 Å and 0.5°, respectively. The dipole moment for this species calculated with the three basis sets varies between 15.1 and 15.7 D. The MP2/6-31G*-optimized bond lengths are generally larger than those at the HF/6-31G* level with the exception of the $C-C_c$ (carboxylic) bond. Bond angles agree within some tenths of a degree. The only remarkable difference, almost 3°, was found for the COH angle. All of these results are in line with previous experiences when using the MP2 level and the 6-31G* basis set.[13a] The larger N−C bond length at the MP2 level was already found for the pyrrole and imidazole rings.[13b] Since experimental results favored the MP2-optimized geometries for those aromatics, the MP2 level has also been considered relevant for the present systems.

The $C-C_c$ bond length and the CNC angle increase upon zwitterion formation by about 0.07 Å and 6−7°, respectively, and the shortening for the C_2-C_3 bond is 0.01−0.02 Å. These geometric changes can affect the π electron delocalization justifying development of specific charge parameters for these systems.

Transition state structures have the COO plane (both in the zwitterion and the neutral form) perpendicular to the ring plane. The zwitterionic TS structures for INA and NA save the C_{2v} and C_s symmetry, respectively, similarly to those in the local minima. $C-C_c$ bond lengths decrease by 0.01−0.03 Å and increase by 0.01−0.02 Å in the zwitterionic and neutral forms, respectively.

Regarding the tautomerization energy, $E_t = E(zw) - E(n)$ (Table 2), a general conclusion may be that E_t is much more sensitive to the basis set and the level used in the single-point calculations than those used in the optimization of the geometries. For example, HF/6-31G* and MP2/6-311++G** single-point calculations using the MP2/6-31G*-optimized geometry give a difference of 2.19 kcal/mol for the $E(zw) - E(n)$ energy of the 4-COOH derivative. On the other hand, E_t changes little even with quite different levels of the optimization, e.g., compare MP2/6-311++G**//MP2/6-31G* and MP2/6-311++G**//HF/6-311+G** results. The NA and INA molecules present overall neutral systems in any tautomeric forms. Addition of diffuse functions to the basis set, which has large effect on the calculated protonation energy[23a] or binding energies of van der Waals complexes[23b] for negatively charged systems, results in only limited changes for E_t, even when the zwitterion was optimized in such augmented basis sets.

Overall, the tautomerization energy is 32−35 kcal/mol for the **3** and **4** structures at the MP2 level. Transition states are higher in energy than the closer minimum by 5.4−7.4 kcal/mol for the zwitterionic and by 11.4−12.7 kcal/mol for the neutral form at the MP2/6-31G* level. Relative ZPE and thermal corrections amount to 0.5 kcal/mol. The activation energy from the neutral form to zwitterion along an intramolecular pathway was calculated as at least 41 kcal/mol for both INA and NA. The HF calculations give tautomerization

(23) (a) Chandrasekhar, J.; Andrade, J. G.; Schleyer, P. v. R. *J. Am. Chem. Soc.* **1981**, *103*, 5609. (b) Del Bene, J. E. *J. Comput. Chem.* **1989**, *10*, 603.

Table 3. Charge Parameters for the Monte Carlo Simulations[a]

	INA		NA		
			zw[b]	neu[b]	
	zw[b]	neu[b]		A	B
N	−0.356	−0.655	−0.284	−0.657	−0.667
C2	0.283	0.490	0.285	0.478	0.497
C6	0.283	0.471	0.222	0.510	0.500
C3	−0.097	−0.278	−0.078	−0.328	−0.328
C5	−0.097	−0.256	−0.062	−0.297	−0.266
C4	0.323	0.276	0.258	0.313	0.280
Cc	0.714	0.711	0.759	0.779	0.798
—O	−0.707	−0.646	−0.717[c]	−0.650	−0.667
=O	−0.707	−0.578	−0.731[d]	−0.602	−0.607
H	0.361	0.464	0.349	0.454	0.461

[a] Rounded values. [b] Obtained with the CHELPG procedure[19a] using the HF/6-31G* wave function. [c,d] The oxygen atom is *cis* and *trans* to the nitrogen, respectively.

energies in the range of 35−38 kcal/mol. These values are close to those obtained for amino acids using the DZP basis set[5a] meaning that the proton transfer energy is fairly large with either aliphatic or aromatic "spacer". For structures with a missing spacer, thus with directly interacting negative and positive sites, the tautomerization may lead to energy changes of only some kcal/mol, as for the H_2N-SO_3H molecule.[7] Considering the $\Delta G_{gas}(298)$ values, the two conformers of the 3-COOH structure comprise an equilibrium mixture in the gas phase with 60% B and 40% A form.

In Solution Energetics. In order to reproduce the experimental tautomerization constant for INA, solvent effects of about −37 kcal/mol should be found if taking the MP2/6-311++G**//MP2/6-31G*-based values for the difference in the internal free energy. Monte Carlo simulations with the first two sets of the charge parameters (Supporting Information) indicate much larger solvent effects. Simulations utilizing the MP2/6-31G* electrostatic potential fitted charges underestimate the relative solvation free energy by about 8 kcal/mol. Best accord with the experimental results was obtained when charges fitted to the HF/6-31G* electrostatic potential were used (Table 3). This supports the finding of Orozco et al.[24] who first proposed the procedure. In the present application, however, the MP2/6-31G*-optimized geometries were used that are superior to those at the HF/6-31G* level for heterocycles.[13b,22]

Table 4 contains a summary of the relative solvation free energies with pure solvents and in mixtures. Onsager correction of −0.3 to −1.0 kcal/mol are considered in the ΔG_{sol} values. Theoretical estimate for the zw/n (zwitterion/neutral form) equilibrium constant of the 3-COOH isomer is 32.6 in good agreement with the experimental values: 31.7, 27.0, and 23.4 extrapolated from measurements in water/methanol, water/dioxane, and water/THF mixtures, respectively. Experimental tautomer ratios for the 80% water/20% THF and 80% water/20% dioxane mixtures are close to each other. The zw/n values, not indicated in the table, are similar to each other for the corresponding 60/40, 40/60, and 20/80 mixtures of water and organic solvents, as well. Calculations with methanol and THF solvents predict practically no zwitterion in the solution. This result agrees well with the experiment where the zwitterionic form was found only in 2−4%. The ΔG_{sol} values are −27 and −15 kcal/mol in the organic solvents, meaning strong solvation preference of the zwitterionic form even in the nonaqueous phase. These values (mainly that in THF) are, however, not enough for compensating the increase in the internal free energy ($\Delta G_{gas} = 33−35$ kcal/mol) upon the proton transfer. The ΔG_{tot}

(24) Orozco, M.; Jorgensen, W. L.; Luque, F. J. *J. Comput. Chem.* **1993**, *14*, 1498.

Table 4. Zwitterion ⇌ Neutral Form Equilibrium in Solution[a]

	ΔG_{gas}	ΔG_{sol}	ΔG_{tot}	zw/A	zw/B	zw/n	zw/n (exp)
			NA				
water							
A	33.21	-36.17 ± 0.16	-2.96 ± 0.16	148.41			
B	33.45	-35.66 ± 0.11	-2.21 ± 0.11		41.81	32.62	23.4−31.7
W8/M2							
A	33.21	-34.27 ± 0.15	-1.06 ± 0.15	5.99			
B	33.45	-33.81 ± 0.12	-0.36 ± 0.12		1.84	1.41	8.7
W6/M4							
A	33.21	-33.52 ± 0.17	-0.31 ± 0.17	1.68			
B	33.45	-33.32 ± 0.14	0.13 ± 0.14		0.80	0.54	2.3
MeOH							
A	33.21	-27.13 ± 0.16	6.08 ± 0.16	∼0			
B	33.45	-27.44 ± 0.12	6.01 ± 0.12		∼0	∼0	0.04
THF							
B	33.45	-14.60 ± 0.11	18.85 ± 0.11		∼0		0.02
W8/THF2							
B	33.45	-33.54 ± 0.12	-0.09 ± 0.12		1.16		5.6
			INA				
water	34.92	-36.33 ± 0.35	-1.41 ± 0.35		10.82		61.7−79.4
MeOH	34.92	-27.51 ± 0.13	7.41 ± 0.13		∼0		0.02
THF	34.92	-14.37 ± 0.11	20.55 ± 0.11		∼0		0.01

[a] $\Delta G = G$ (zwitterion) $- G$ (neutral) free energy differences in kcal/mol. ΔG_{sol} and ΔG_{tot} values are given with standard deviation. The zw/A, zw/B, and zw/n values were calculated by taking the average value for ΔG_{tot}.

$= \Delta G_{gas} + \Delta G_{sol}$ values are negative and positive in water and in the organic solvents, respectively, corresponding to the overall preference of the zwitterion and the neutral form accordingly. Thus, the almost exclusive existence of the solute in the neutral form with pure organic solvents is not due to a less favorable solvation thermodynamics of the zwitterion here, as might be expected based on the low dielectric constants for THF and dioxane. Instead, the reason is the large internal energy required for the formation of the zwitterion.

Calculations provide information also for the A ⇌ B conformational equilibrium in the neutral form. The MP2/6-31G* dipole moments for A and B are 3.24 and 0.79 D, respectively. B is more stable than A in the gas by 0.24 kcal/mol comprising 60% of the mixture. Solvation in water shifts the equilibrium further toward the B form that has a 80% population in aqueous solution. In solutions with water/methanol solvents, A becomes gradually more preferred. In pure methanol, A has a population of 53%. Stabilization of B in water is interesting because it is in contrast to the general idea that the structure with larger dipole moment is more readily stabilized in a polar solvent.

To explain this result the components of the relative free energy terms must be studied. $\Delta G_{sol}(T) = \Delta H_{sol}(T) - T\Delta S_{sol}(T)$ at $T = 298$ is $-0.54 = -3.38 + 2.84$ for the A to B transformation in water and is $+0.29 = -3.52 + 3.81$ in methanol. The conformational change is enthalpy dominated in water but is entropy dominated in methanol.[25] The solute−solvent interaction energy, E_{SX} is more negative by about 2.5 kcal/mol for B than A both in water and methanol. Since the N−C and O−H bond polarities are larger for B than for A (Table 3), more negative E_{SX} values for B suggest larger energy contributions from interactions of the solute with solvent molecules in the first solvation shell than with more remote ones, with which the interactions correspond basically to the dipole−dipole type. Since the preference for the conformer in water is enthalpy dominated (see above), it explains the more favorable solvation free energy for the B conformer. The E_{SX} is more negative for B than for A still in methanol. It means that methanol molecules in the first solvation shell are more

strongly bound, and correspondingly, localized to the B rather than to the A conformer. It leads to the decrease of the solvent entropy with the B form and an overall stabilization of the A conformer in methanol. (Because of the negative sign for the $T\Delta S_{sol}$ term, decrease of entropy with B results in a positive contributions to ΔG_{sol} throughout the A to B transformation.) A larger entropy decrease with methanol as solvent rather than with water (-3.81 vs -2.84 kcal/mol) upon strong localization of solvent molecules is not unexpected. The extensive and strong hydrogen-bond network in water considerably hinders the disorder of the water molecules even in the pure liquid. A more disordered structure is expected for methanol which will then react more sensitively for stronger localization of some solvent molecules in the first solvation shell.

Calculated zw/n equilibrium constants in mixtures (water/methanol 80/20 and 60/40, W8M2 and W6M4, respectively, water/THF 80/20, W8THF2 in Table 4) underestimate the zwitterionic fraction. The preference for the neutral form emerges more quickly than found experimentally. The experimental equilibrium mixture consists of 90% zwitterion and 10% neutral form in W8M2, considering only these two forms. The calculated values are 59% and 41%, respectively. The ratio is 85/15, as found experimentally in the mixture of 80% water/20% THF, while the calculated ratio for the W8THF2 model is 54/46. Thus, in solvents with 80% water, the calculated values still predict the prevalence of the zwitterionic form and reflect the reduction of the equilibrium constant if the organic component changes from methanol to THF. In the mixture W6M4, the calculated fraction of the zwitterion (35%) is inferior compared to that of the neutral form, in contrast to the experimental values in methanol of 70% and 30%, respectively, and in dioxane of 55% and 45%, respectively. The calculated equilibrium constant for W6M4 is, however, fairly sensitive to the energy values considered in obtaining zw/n, because standard deviations are close to the absolute ΔG_{tot} values in this case. Numerical deviations from experimental values above are mainly attributed to the use of a uniform set of atomic charges which, seemingly, do not reflect subtle changes in mixtures as compared to pure solvents. In general, on the basis of the course of the calculated equilibrium constants, proper prediction of the dominating form would be expected for mixtures W4M6 and

(25) Nagy, P. I.; Ulmer, C. W., II; Smith, D. A. In *Modeling Hydrogen Bond*; ACS Symposium Series 569; Smith, D. A., Ed.; American Chemical Society: Washington, DC, 1994; p 60.
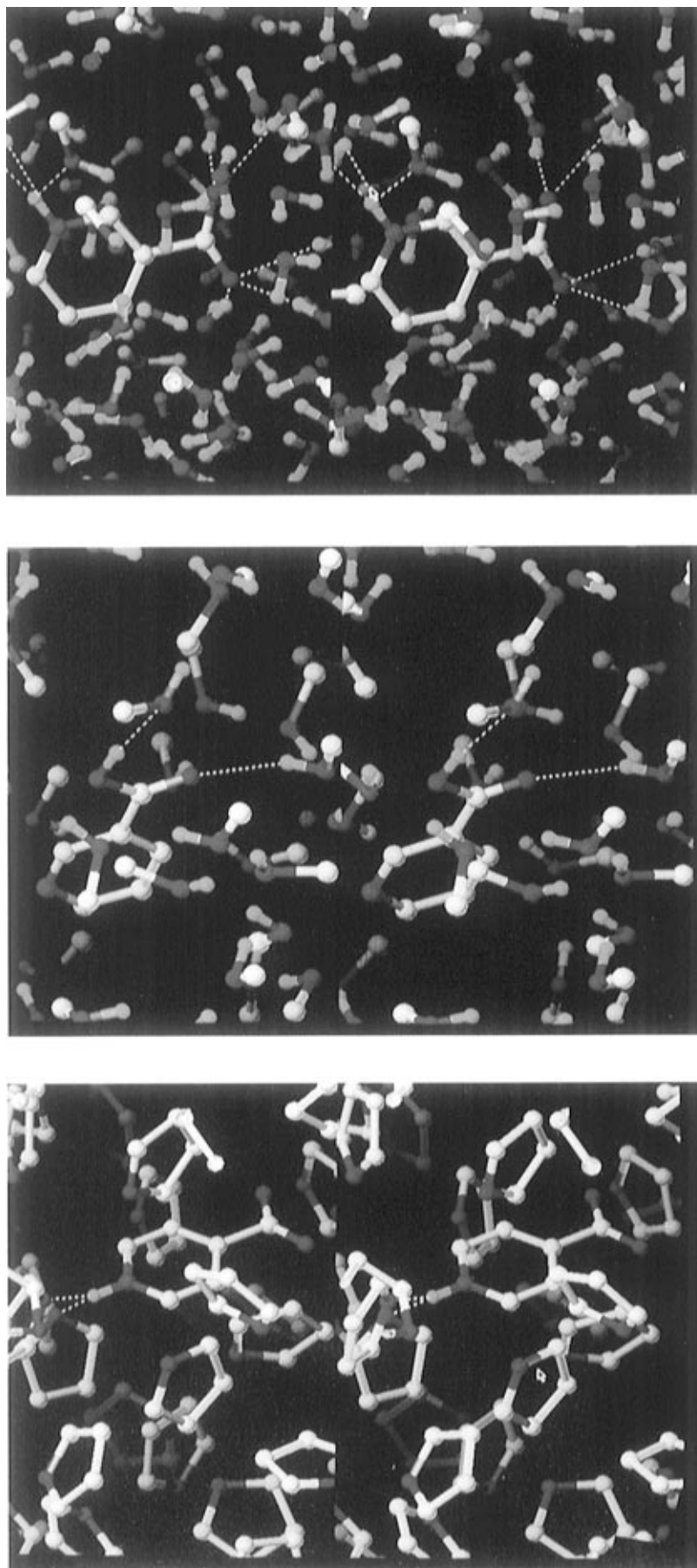
Figure 3. Stereoviews for snapshots from Monte Carlo simulations. Coloring: C (white), O (red), H (cyan), N (blue). Hydrogen bonds are designated with yellow dashed lines. (A) Zwitterionic nicotinic acid in mixed solvent with 60% water and 40% methanol. (B) Neutral nicotinic acid in methanol. (C) Zwitterionic isonicotinic acid in tetrahydrofuran.

Ampholyte Equilibrium in Solvents and Mixtures

J. Am. Chem. Soc., Vol. 119, No. 21, 1997 5005
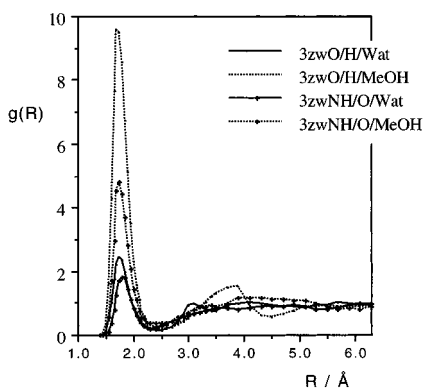
Figure 4. NH$^+$/O and O$^-$/H radial distribution functions for the zwitterionic NA (3zw) in water and methanol.
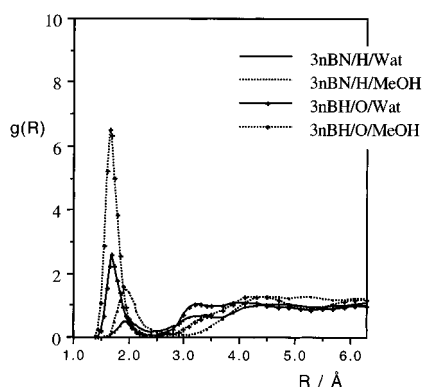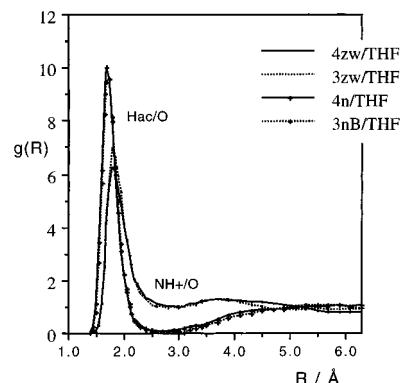
Figure 5. OH/O and N/H radial distribution functions for the neutral NA (3nB) in water and methanol.

W2M8. The experimental values for these mixtures are 0.33 − 0.65 and 0.14, respectively, indicating the prevalence of the neutral form. The two tautomeric forms must be of equal population in mixtures with water fraction between 40% and 60%. Recently Halle et al.[8] has predicted an equilibrium constant of 1 in the mixture of 62% water/38% dimethyl sulfoxide. Equilibrium constans of 2.3, 1.2 and about 1.6 for mixtures with 40% methanol, dioxane, and dimethyl sulfoxide, respectively, indicate slight dependence on the organic component throughout the tautomerization process.

Calculations for the 4-COOH derivative have been made only in pure solvents. The experimental results are similar to those for the 3-substituted isomer. The zwitterionic form is dominant in water, the neutral one in the organic solvents. Although the calculated zw/n constant, 10.82 in water, is considerably smaller than the experimental value of 61.7−79.4, the fractions for the zwitterion, 91.5% vs 98.6%, are relatively close.

Solution Structure. Solution structures will be analyzed by virtue of snapshots taken after considering 6×10^6 configurations in the simulations (Figure 3) and interpreting radial distribution functions (rdfs) in Figures 4−6. Figure 3A shows the neighborhood of 3zw in the W6M4 mixture. A water and a methanol molecule form O···H−N hydrogen bonds to the protonated site of the ring. The carboxylate group forms altogether five hydrogen bonds both to water and methanol molecules. It indicates mixed solvation in favor, however, of hydration in accord with the larger water fraction in the solution. H$^+$/O coordination numbers (see Supporting Information) calculated from rdfs in Figure 4 are around 1 in both pure water and methanol. Although Figure 3A shows two molecules

involved in hydrogen bonds to H$^+$N, the CH$_3$HO···H$^+$N distance of 2.43 Å is just within the minimum of the H$^+$/O rdfs of about 2.5 Å. This molecule can move out and into the first solvation shell, and only the water molecule should be considered as strongly bound on the basis of the H$_2$O···H$^+$N distance of 1.66 Å in Figure 3A. The O$^-$/H coordination numbers summed up for the two oxygens are 5.2 in water and 3.8 in methanol. Figure 3A shows two strong hydrogen bonds for each of the carboxylate oxygens. There is also a fifth and a weaker H-bond by a methanol molecule. This solvation pattern nicely fits to the average values. The total number of hydrogen bonds, calculated by integration up to the first minima of the pair energy distribution functions,[26] is 6.5 in water and 4.4 in methanol. Five strong and one or two weak hydrogen bonds in the configuration shown in Figure 3A are in reasonable agreement with the average values.

Characteristic hydrogen bonds to the neutral B conformer of the NA solute (3nB) in methanol are shown in Figure 3B. The solute is a proton donor in the COOH···OHCH$_3$ hydrogen bond and is an acceptor when the carbonyl oxygen is bound to the alcoholic hydrogen. Solvation of the carboxylic OH and =O by one methanol molecule for each corresponds nearly to the average case with H$_{ac}$/O and =O/H coordination numbers of 1.0 and 1.3, respectively. Figure 3B does not show any hydrogen bonds to the basic N. The closest CH$_3$OH···N distance is 3−4 Å, far beyond the limit for a hydrogen bond. The snapshot partially reflects the general finding: coordination of the pyridine nitrogen by the alcoholic hydrogen is relatively weak, giving 0.5 units for the coordination number on averge. Total number of hydrogen bonds to the 3nB solute by methanol was calculated at 2.6 on average. Considering the two strong hydrogen bonds with the carboxylic group, not even one stable hydrogen bond is expected at the N-site. Analysis of the corresponding values for the aqueous solution leads basically to similar conclusions (Figure 5).

The THF solvent can form strong hydrogen bond only by its oxygen as an acceptor. Figure 3C shows two THF molecules around the H$^+$N site with O···H separations of 2.2−2.3 Å. The average H$^+$/O coordination numbers for NA and INA are 1.3 and 1.4, respectively. However, the average number of the hydrogen bonds is around 1 with both solutes, indicating that not each THF molecule in the first solvation forms a strong hydrogen bond. Indeed, O···H separations in Figure 3C are larger than 1.80 Å, the location of the first maximum of the corresponding rdf (Figure 6). More separated solvent molecules

Figure 6. NH$^+$/O and OH/O radial distribution functions for INA (4) and NA (3) in tetrahydrofuran solvent.

(26) (a) Nagy, P. I.; Dunn, W. J., III; Alagona, G.; Ghio, C. *J. Phys. Chem.* **1993**, *97*, 4628. (b) Dunn, W. J., III; Nagy, P. I. *J. Comput. Chem.* **1992**, *13*, 468. (c) Nagy, P. *Acta Chim. Hung.* **1992**, *129*, 429.

should form weaker hydrogen bonds. Analysis of rdfs and pair energy distribution functions for the neutral form indicate a single $O(THF)\cdots HOOC$ hydrogen bond both with NA and INA solute.

A problem, frequently addressed for interactions of ring systems, is the so-called stacking effect. Figure 3C shows that the most THF molecules are *not* stacked with the ring of solute. This may be the general case for THF−(substituted)pyridine systems, because an ordered, stacked arrangement should result in remarkable peaks in the $CH(pyr)/CH_2(THF)$ rdfs, which, however, have not been found.

Conclusions

The equilibrium ratios for the zwitterionic and neutral forms of the ampholyte nicotinic and isonicotinic acids have been determined by spectrophotometry in different water/methanol, water/THF, and water/dioxane mixtures. The zwitterionic form of the solute is the prevalent structure in mixtures with at least 60% water. The neutral form is taken in almost 100% in pure organic solvents.

Ab initio calculations at the MP2/6-311++G**//MP2/6-31G* level predict that the neutral form is more stable by 33−35 kcal/mol than the zwitterionic tautomer at $T = 298$ K and $p = 1$ atm in the gas phase. In contrast, Monte Carlo simulations using the free energy perturbation method prefer the zwitterionic form by about 36, 27, and 15 kcal/mol in water, methanol, and THF, respectively. Taking both effects into account, the solvation term for overbalancing the increased internal free energy of the zwitterion is large enough only in water. Predominance of the neutral form in pure organic solvents is thus due to the considerably lower internal energy of this structure as compared to the zwitterion. The dominant tautomeric form in mixed solvents has been generally predicted in agreement with the experiment.

Solution structure analysis shows comparable role of the components in the solvation process with mixed solvents. Strong hydrogen bonds are formed both to the carboxylate and N−H groups in any solvent. In the neutral form, weaker and less-specified hydrogen bonds are formed with the solute.

A hypothesis has been recently proposed that only the neutral form enters the organic phase at the partitioning of tautomers of an ampholyte solute between immiscible solvents, even though the zwitterionic form is the dominant one in the aqueous solution. Because of the two-phase system, theoretical studies of the tautomeric equilibrium in the constituent pure solvents (with a wide range of dielectric constants and with different characters for forming hydrogen bonds) may be useful for characterizing ampholytes in the separate phases. The present results are considered as a first step toward the better understanding of partitioning between immiscible solvents and interpreting the true log P values useful in drug design.

Supporting Information Available: Spectrophotometry curves for NA and INA in water/THF mixtures, optimized geometric parameters, all sets of atomic charges used in four series of MC simulations, summary of coordination numbers, number of hydrogen bonds, and solute-solvent energies, and pair energy distribution functions (7 pages). See any current masthead page for ordering and Internet access instructions.

JA963512F