Quantitative Structure−Activity (Affinity) Relationship (QSAR) Study on Protonation and Cationization of α-Amino Acids

Fung-Ming Siu* and Chi-Ming Che

Department of Chemistry, Open Laboratory of Chemical Biology of The Institute of Molecular Technology for Drug Discovery and Synthesis, The University of Hong Kong, Pokfulam Road, Hong Kong SAR, China

Received: July 10, 2006; In Final Form: September 3, 2006

A quantitative structure−activity (affinity) relationship (QSAR) study is carried out to model the proton, sodium, copper, and silver cation affinities of α-amino acids (AA). Stepping multiple linear regression (MLR), partial least squares (PLS), and artificial neural network (ANN) approaches are applied to elucidate the multiple factors affecting these affinities. The MLR and PLS models reveal that the variation in proton affinity is attributed to the highest electrophilic superdelocalizability of nitrogen (major) and the number of rotatable bonds (minor) in AA. The noncovalent interactions, especially ion−dipole interactions, are responsible for the changes in Na^+ affinity. The ionization potential, dipole moment of the side chain, and degree of linearity are the properties of AA that give the best correlation with the Cu^+ and Ag^+ affinities. The ANN models are developed to study the relationships (linear or nonlinear) between the molecular descriptors and binding affinities. The ANN models show higher predictive power. The QSAR models are used to study the binding forms of AA (neutral vs zwitterionic) upon protonation/cationization. To our knowledge, this is the first attempt to carry out a QSAR study on protonated/cationized AA to elucidate their binding properties. In virtue of the Na^+ affinity ANN model, the Na^+ affinities of dihydroxyphenylalanine (DOPA) were predicted. This work may pave the way for the success of applying similar approaches to peptides or proteins (with AA as the building blocks) in the future.

1. Introduction

Proton, alkali metal, and transition metal cations play essential roles in various biological processes. Proton (H^+) is ubiquitous in biological systems; H^+ plays a key role in determining the structures and energetics of proteins.[1] Sodium cation (Na^+) is one of the most abundant metal cations in biological systems and is involved in various processes, including osmotic balance, stabilization of biomolecular conformations, and information transfer via ion pumps and ion channels.[1] Copper cation (Cu^+) plays an essential role in oxidation, dioxygen transport, and electron transfer.[2] Silver cation (Ag^+) does not appear to have a role in natural biological systems; it is used as an antimicrobial agent.[3] Studies by Adams and other researchers have demonstrated that, in many cases, the gas-phase binding parallels that encountered inside proteins in aqueous solution chemistry.[4] Thus, thermodynamic data obtained in the gas phase are of particular value for understanding the nature of cation−protein interactions in the aqueous environment.

There are considerable evidences that the site of protonation[5] and cationization[6−8] in peptides influences the fragmentation reactions observed in tandem mass spectra. For instance, protonated peptide is invariably observed in the positive ion mass spectrometric analysis. The fragmentation reaction chemistry of protonated peptide is adequately described in terms of a "mobile proton model".[9] The fragmentation reactions observed depend in part on the relative proton affinities of various AA. In fact, that a variety of cations can easily bind to peptides has been exploited in the field of mass spectrometry, where proton and metal cations (such as Na^+, Cu^+, and Ag^+) have been used as ionizing agents for peptide sequencing.[5−8]

Due to the biological importance and practical applications in mass spectrometric research fields, the intrinsic nature of binding between a proton/metal cation and α-amino acids (AA, as the building blocks of peptides and proteins) has been a hot topic in recent years. Several quantitative studies on H^+, Na^+, Cu^+, and Ag^+ affinities of AA have been reported. Interestingly, although the ionic radii of Na^+ and Cu^+ are similar (0.97 Å) and the outer electronic configurations are the same for Cu^+ and Ag^+ (d^{10}), their biological roles and binding properties are different.[1−3,5−8]

Quantitative structure−activity relationship (QSAR) studies have been demonstrated to be an effective computational tool in understanding the relationships between the structures of molecules and their properties, e.g., biological activity, boiling point, etc. In this work, the QSAR models developed using multiple linear regression (MLR), partial least squares (PLS), and artificial neural network (ANN) approaches were used to study (i) the binding nature of protonated and cationized (Na^+, Cu^+, Ag^+) AA, (ii) the relationships (linear or nonlinear) between molecular descriptors and binding affinities, (iii) the binding forms of AA (neutral (CS) vs zwitterionic (ZW)) upon protonation/cationization, and (iv) the noncovalent interaction between Na^+ and dihyroxyphenylalanine.

2. Methodology

2.1. Data Set. The QSAR analysis was performed on 20 common AA. The proton affinity reviewed and finally recommended by Harrison[10] was used. The Na^+,[11] Cu^+,[12−14] and Ag^+ [15,16] binding energies determined by a kinetic method were chosen for this study.

* Corresponding author. Fax: +(852)-2857-1586. E-mail: fmsiu@hkucc.hku.hk.

10.1021/jp064332n CCC: $33.50
Published on Web 10/12/2006

TABLE 1: Molecular Descriptors Used in the Final QSAR Models

notation	molecular descriptors	refs
DM	dipole moment (entire AA)	20, 21
DM_S	dipole moment (side chain)	20
ES_N	highest electrophilic superdelocalizability on nitrogen (site of protonation)	20, 21
ES_S	sum of the E-state index (side chain)	32−35
IP	ionization potential (entire AA)	20, 21
KA	KAlpha2 index (entire AA)	42−45
RB	rotatable bonds (entire AA)	20

2.2. Molecular Modeling and Molecular Descriptors. All structures were generated within the CACHE WorkSystem Pro (version 6.1.12.33) package (Fujitsu Limited, Oxford Molecular Ltd.), in which the subroutine MM3/PM3 calculation was used. The same approach has been used in other QSAR studies.[17,18] In brief, the chemical structures were initially refined by performing a preoptimized calculation of molecular mechanics. The CACHE MM3 augmented mode implements the Allinger's MM3 force field was used.[19] The structures were further refined with CACHE MOPAC2002, which is a semiempirical quantum mechanics package that solves the Schrödinger equation by using the PM3 semiempirical Hamiltonian algorithm.

Molecular descriptor values derived from the CS and ZW forms of AA were calculated with the TSAR 3D (version 3.3) and CACHE WorkSystem Pro (version 6.1.12.33) packages. Descriptors for the entire molecule, side chain, and nitrogen of the AA were calculated. TSAR afforded calculation of the following descriptors: molecular mass, surface area, volume, ellipsoidal volume, dipole/lipole moment, Verloop parameter, Wiener, Randic, Balaban, electrotopological state index, molecular connectivity index, Kier and Hall shape index, number of rotatable bonds/H-bond donors/acceptors, VAMP energy and VAMP polarizability, and dipole and octupole components. Partial charge, HOMO/LUMO density, electrophilic/nucleophilic/radical frontier density, and superdelocalizability of nitrogen were calculated using the CACHE program. These descriptors encode topologic, geometric, and electronic information about AA. The definitions of these descriptors are available in the literature;[20,21] thus, no details are repeated here, unless they are used in the final QSAR models. Details will be discussed in section 3.1 and Table 1.

2.3. MLR and PLS Models. On the basis of identical tests, those descriptors with identical values for more than 90% of the AA were discarded. This is important because if there was no discrimination between descriptor values, little useful information would be provided by that descriptor. Pairwise correlation analysis of the remaining descriptors was performed; one of the descriptors, with correlation greater than 0.9, was removed.[22]

In brief, the MLR and PLS models were developed with the following steps (Figure 1): (i) A stepping MLR procedure, with leave-one-out cross-validation, forward selection, and backward elimination, was used to select the important molecular descriptor values derived from the CS and ZW form of AA. The predictive power of a model is described with r^2(CV), the cross-validated equivalent of the correlation coefficient (r^2). In general, models developed with those descriptor values derived from the CS forms of AA give higher cross-validation coefficients (r^2(CV)). However, two exceptions were observed, i.e., Na^+/Ag^+−Pro. In other words, the final MLR models on H^+ and Cu^+ affinities were developed with those descriptor values derived from the CS form of AA. For the final Na^+ and Ag^+ affinities models, the descriptor values were derived from the ZW form of Pro but the CS form of other AA. (ii) For a well-defined problem, MLR and PLS models should give similar predictions. PLS analysis with leave-one-out cross-validation was performed using the same set of molecular descriptors, to check if the predictions (affinities) are consistent. The coefficients were expressed relative to their standardized form to give an immediate indication of the relative importance of each descriptor in the final models.[23] (iii) Final models with high cross-validation coefficients r^2(CV) and number of descriptors not more than three are discussed in details.

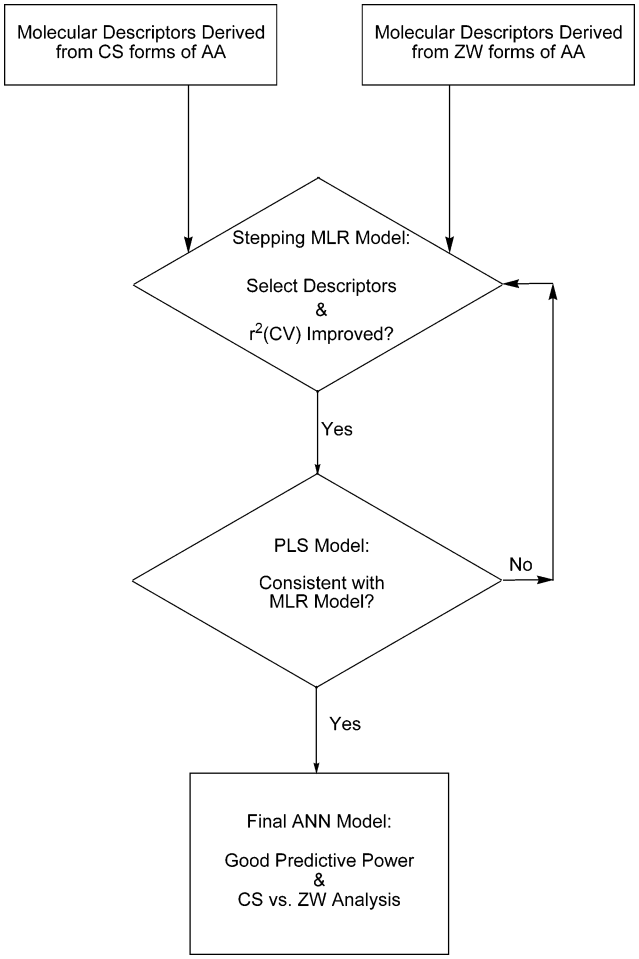

Figure 1. Workflow used to develop the QSAR models.

2.4. ANN Models. ANN is a layered system of processing units that are interconnected to facilitate the ordered transfer and data processing. It is intended to simulate the interpretative capacity of the brain. It was reported that ANN is superior to MLR in providing accurate predictions.[24−27] By definition, MLR assumes a linear relationship between binding affinities and molecular descriptors, or incorporated explicitly.[24−26] In contrast, ANN makes no assumption about the linearity of a problem. The major advantage of ANN lies in the fact that QSAR can be developed without having to specify the analytical form of a particular correlation model.

In this work, the neurons were arranged in a three-layered forward feed ANN model: an input layer (molecular descriptor values used in the final MLR and PLS models), one hidden layer, and an output layer (affinities). The Monte Carlo algorithm was used to select a better set of starting weights within the default constrained limits. A proportion of the input data (30%) was excluded from the training set and used as a test set. The ratio (ρ) between the number of input variables and the number of hidden neurons, which is critical to the predictive power of

the ANN, was set to close to 2 to prevent the problems of overfitting or memorizing data.

The ANN models were used to study the (linear or nonlinear) relationships between the molecular descriptors and binding affinities (Table 1). To display the dependency of each molecular descriptor (in a qualitative manner), a constant value was fed into all input nodes, except for the molecular descriptor in question, which was varied over a range of 0.1−1.0. The results were visualized on a 2D plot of output node against input (dependency graph).

3. Results and Discussion

3.1. Final MLR and PLS Affinity Models. According to the procedures mentioned above, the final MLR and PLS models on binding affinities were developed.

3.1.1. Proton Affinity. Harrison reviewed the literature to derive a consistent set of proton affinities (PA) for AA.[10] This set of data was chosen for our study. The final MLR model is presented in eq 1 (Table 1S in the Supporting Information):

$$\mathrm{PA} = 162\mathrm{ES_N} + 19\mathrm{RB} + 672 \quad (1)$$

$r^2(\mathrm{CV}) = 0.8$ where $\mathrm{ES_N}$ and RB stand for the highest electrophilic superdelocalizability of nitrogen and the number of rotatable bonds in AA, respectively; $r^2(\mathrm{CV})$ is the cross-validated equivalent of the correlation coefficient (r^2) that describes the predictive power of the model.

The final PLS model developed with the same molecular descriptors is represented by eq 2 (Table 1S):

$$\mathrm{PA} = 180\mathrm{ES_N} + 17\mathrm{RB} + 653 \quad (2)$$

The statistical significance = 0.4, $r^2(\mathrm{CV}) = 0.8$.

Consistent results were obtained from the MLR and PLS models. Predicted affinities derived from the MLR model (eq 1) versus those of the PLS model (eq 2) give an equation of the form $y = x$ and an r^2 of 0.996. The statistical significance value less than 1.0 indicates that the PLS components are significant.[23] The coefficients of the PLS models are shown relative to the standardized descriptors values; their relative magnitudes give an immediate indication of the relative importance of each molecular descriptor in the final models.[23] They are discussed below:

The electrophilic superdelocalizability of an atom is defined as:

$$\text{electrophilic superdelocalizability} = 2\sum_{j=1,m}\sum_{\alpha=1,q}(c_{\alpha j}{}^2/\lambda_j) \quad (3)$$

where $C_{\alpha j}$, λ_j, α, and j stand for the eigenvectors and the eigenvalues, atomic orbitals (q), and occupied molecular orbitals (m), respectively. It is a measure of the availability of electrons. It was reported that experimental hydrogen binding strength increases as electrophilic superdelocalizability increases, and a good correlation was reported for small organic molecules.[28] However, no correlation has been reported for AA, to the best of our knowledge.

It is generally believed that nitrogen is the most favorable protonation site for AA. The symbol $\mathrm{ES_N}$ (eqs 1 and 2) is defined as the highest electrophilic superdelocalizability on nitrogen. Most of the AA show the highest electrophilic superdelocalizabilty on the α-NH_2 nitrogens, with some exceptions: the nitrogens in the side chain of Arg, His, and Lys have the highest electrophilic superdelocalizability instead. Consistently, those nitrogens which we found to have the highest electrophilic superdelocalizability were also found to be the sites of protonation by Maksić et al.[29] In this work, we found the PLS coefficient (eq 2) for $\mathrm{ES_N}$ is the largest. Our results further confirmed that the variation in PA is mainly attributed to the availability of electrons in the major protonation sites (α-NH_2 or the side chain of AA).

A rotatable bond is defined as any single-order bond that is not a terminal bond, not a ring bond, and not an amide bond.[20] In eq 2, the number of rotatable bonds (RB) shows a minor contribution to the variation in proton affinity. Neutral and protonated AA are stabilized by networks of intramolecular hydrogen bonds (H-bond); their stabilities are determined by the interplay between the intramolecular H-bond and internal strain. Upon protonation, the original networks of the intramolecular H-bonds are disturbed. Those AA with more rotatable bonds tend to have lower internal strain. They have a higher tendency to retain the original networks of the intramolecular H-bonds or change to another binding network with comparable stabilities. As a result, PA increases as the RB increases.

3.1.2. Sodium Affinity. Kish et al. determined the Na^+ affinities of AA by a kinetic method;[11] this set of data was chosen for our studies. According to the procedure mentioned in section 2.3, a high predictive power model was developed when the molecular descriptor values were derived from the neutral form of AA and the ZW form of Pro. For example, using the molecular descriptors in eq 4, but replacing the molecular descriptor values derived from the ZW form of Pro with those of the CS form, the $r^2(\mathrm{CV})$ decreased by 0.1. Both the MLR and PLS models give similar predictions (Table 2S in the Supporting Information); the final PLS model is shown in eq 4:

$$Na^+\ \text{affinity} = 6\mathrm{DM_S} + 2\mathrm{DM} + 2\mathrm{ES_S} + 165 \quad (4)$$

The statistical significance = 0.4, $r^2(\mathrm{CV}) = 0.8$.

The dipole moment of the side chain ($\mathrm{DM_S}$) and that of the entire molecule (DM) of AA increase as the Na^+ affinities increase. This implies that ion−dipole interaction plays an important role in Na^+ binding; this result is consistent with previous findings.[30,31]

The E-state index is a descriptor that represents the electron density and the accessibility of those electrons to participate in noncovalent intermolecular interactions.[32,33] The index also takes into account the structural configuration of the nearest neighbors surrounding the atom and thus contains some shape information, although in a secondary fashion. The E-state value of the side chain ($\mathrm{ES_S}$) is equal to the sum of states values (S) for all atoms in the side chain. The states value for atom i in the side chain is defined as S_i:[34,35]

$$S_i = I_i + \Delta I_{ij} \quad (5)$$

The perturbation term is defined as

$$\Delta I_{ij} = \sum (I_i - I_j)/r_{ij}{}^2 \quad (6)$$

in which the separation, r_{ij}, is given as the number of atoms in the shortest path between atoms i and j. The intrinsic state (I_i) for an atom i is obtained from the ratio of its valence state electronegativity to the number of skeletal bonds, that is, the avenues over which electron density may be distributed, and is given as follows:

$$I_i = ((2/N_i)^2\delta_i^v + 1)/\delta_i \quad (7)$$

where N_i is the principal quantum number for the valence electrons, δ_i is the number of connections in the skeleton, i.e., the number of electrons in the σ orbital (σ) minus the number of hydrogens (h) bonded to the atom i, δ_i^v is the molecular connectivity valence, i.e., $\sigma + \pi + n - h$, where π and n are the number of electrons in the π orbital and the lone pair, respectively, in atom i. As a consequence of this definition, atoms that possess π and lone pair electrons or are terminal atoms or lie on the mantle of the molecule tend to have large positive values for S_i. Atoms which do not have π and lone pair electrons and/or are buried in the interior of the molecule tend to have small or negative E-state values.[34] To the best of our knowledge, this is the first time Na^+ affinity was reported to increase with E-state value increases. This could be attributed to those favorable cation−dipole and cation−π interactions.

It is noted that two out of the three molecular descriptors used in the final Na^+ models describe the properties of side chains (ES_S and DM_S). This reflects the important role of the side chain in Na^+ binding. It is consistent with previous theoretical results for sodiated Ser, Cys, Phe, Tyr, Trp, etc. that their functional side chains also participate in Na^+ bindings.[36−39]

3.1.3. Copper Affinity. Hoyau and Ohanessian[13] combined their best calculated Cu^+−Gly affinity with Cerda and Wesdemiotis's experimental relative scale to obtain absolute affinities for Cu^+−AA.[12] The Cu^+−Val, Lys, and Arg affinities were further determined by an advanced kinetic method.[14] These reported affinities were used to develop the MLR and PLS models (Table 3S in the Supporting Information); the final PLS model is depicted in eq 8:

$$Cu^+ \text{ affinity} = -21\text{IP} + 6\text{DM}_S + 12\text{KA} + 449 \quad (8)$$

The statistical significance = 0.5, r^2(CV) = 0.8.

3.1.4. Silver Affinity. Lee and co-workers determined the relative silver ion binding energies by a kinetic method.[15] This relative scale was anchored to Shoeib et al.'s reported theoretical Ag^+−Gly affinity (at the B3-LYP/DZVP level of theory),[16] to obtain the data set for this work. The final MLR and PLS models were developed with the molecular descriptor values derived from the CS form of AA but the ZW form of proline. Both MLR and PLS models give similar predictions (Table 4S in the Supporting Information). The final PLS model is shown in eq 9:

$$Ag^+ \text{ affinity} = -29\text{IP} + 8\text{DM}_S + 14\text{KA} + 455 \quad (9)$$

The statistical significance = 0.4, r^2(CV) = 0.8.

It is of interest to note that the same molecular descriptors with similar coefficients (same order of magnitude) were used in the final Cu^+ and Ag^+ affinities models (eqs 8 and 9). Thus, the variations in Cu^+ and Ag^+ affinities are attributed to the similar factors as discussed below:

Both Cu^+ and Ag^+ are soft acids according to the hard−soft acid−base (HSAB) principle. The definition of the absolute softness of an AA is given by 1/(ionization potential − electron affinity). Since all AA have very small electron affinities,[40] one can reasonably assume that the absolute softness increases as the ionization potential decreases. The negative coefficient of the IP (eqs 8 and 9) indicates the applicability of the HSAB principle to these bindings, i.e., Cu^+ and Ag^+ affinities increase as the softness of AA increases.

It is generally believed that the binding of Cu^+ and Ag^+ to AA involves a sum of charge-transfer and electrostatic interactions.[40,41] It is not a surprise that the dipole moment of the side chain (DM_S), also contributed to the variations in Cu^+ and Ag^+ affinities.

Kappa Alpha index (KA) is defined as

$$\text{KA} = ((A + \alpha - 1)(A + \alpha - 2)^2)/({}^2P_i + \alpha)^2 \quad (10)$$

where A is the number of atoms in the molecule, ${}^2P_i$ is the total number of paths along adjacent bonds with two bond lengths, and α accounts for the contribution of each atom to the overall shape of a molecule based on a comparison with a carbon sp^3 atom. Thus, α is defined as the ratio of radii of carbon and each individual atom.

$$\alpha = (r_x/r_{\text{Csp}^3}) - 1 \quad (11)$$

As a consequence of this definition, KA indicates the degree of linearity of bonding patterns for a molecule.[42−44] The KA value is large for linear structures but small for branching structures.[44] In general, the flexibility of a molecule is directly related to the degree of linearity.[45] All AA have a common backbone structure; thus, the variation in KA is mainly attributed to the linear flexible side chain, such as Arg, whose KA value is largest. Positive coefficients are observed for KA in Cu^+ and Ag^+ affinities models (eqs 8 and 9). This implies that the high Cu^+ and Ag^+−Arg affinities originate from the flexible side chains enabling multidentate coordination.

3.1.5. General Remarks on MLR/PLS Models. Together with the measurement and computation of affinities, discussion and interpretation of trends has been reported over the years. Based on chemical intuition and by interpreting the correlations among various binding affinities, researchers believe that protonation and Na^+ binding are covalent and electrostatic in nature, respectively,[11,46] while Cu^+ and Ag^+ are soft acids.[12,15] In this work, the final MLR and PLS models on proton affinity show that H^+ favors interaction with those AA with a high electrophilic superdelocalizability on nitrogen (major factor) and a large number of rotatable bonds (minor factor). Noncovalent interactions, especially ion−dipole interactions, contribute significantly to the Na^+ binding. The HSAB principle can be used to explain the Cu^+ and Ag^+ binding affinities nicely. In addition, ion−dipole interactions and the degree of linearity of AA are responsible for the variations in Cu^+ and Ag^+ binding affinities. In general, these conclusions are consistent with the literature; hence, they prove the validity of the QSAR models developed in this work.

Despite what has been studied in the past, no good correlation between affinity and any single physical/chemical property of AA has been reported. Although such a trend is present in the data, the correlation is not strong. For instance, a poor correlation between hardness and Cu^+ affinity was reported for small ligands (e.g., NH_3, MeCN, Me_2SO, etc.). This is because hardness only accounts for one component of the binding.[47] There is no doubt that for complex systems such as AA, a single property is not enough to explain the trend in a quantitative manner. For example, correlations between affinities and the descriptors with the largest coefficients in their corresponding PLS models (i.e., ES_N, DM_S, and IP for H^+, Na^+, and Cu^+/Ag^+ affinities, respectively) give an average r^2 of 0.4 only. Thus, a single property could only account for part of the variations in affinities. The QSAR models described herein show better correlations and higher predictive powers (average r^2(CV) = 0.8) and elucidate the multiple factors contributing to the variations in binding affinities. To the best of our knowledge, molecular descriptors such as ES_N/RB, DM_S/ES_S, and DM_S/KA have not been used to elucidate the H^+, Na^+, Ag^+/Cu^+

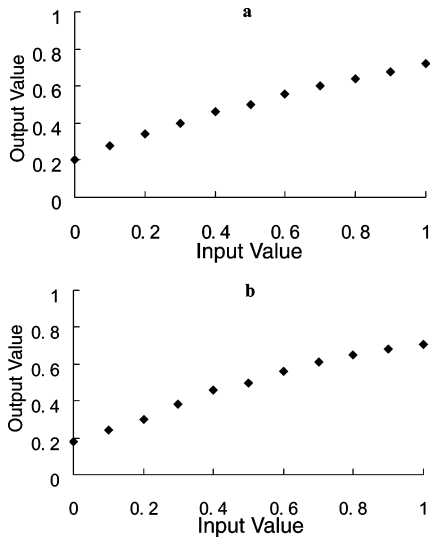

Figure 2. Dependency graphs of the ANN model on H^+ affinity: (a) ES_N; (b) RB.

TABLE 2: Correlation Coefficients (r^2) between Predicted and Experimental Affinities of the MLR, PLS, and ANN Models

model	MLR	PLS	ANN
H^+ affinity	0.85	0.85	0.86
Na^+ affinity	0.85	0.85	0.91
Cu^+ affinity	0.85	0.85	0.96
Ag^+ affinity	0.88	0.88	0.93
average	0.86	0.86	0.92

affinities of AA, respectively. We envision that these molecular descriptors may also be important in elucidating other binding properties of AA, as well as those of peptides/proteins.

3.2. Artificial Neural Network. The MLR and PLS methods have been found useful for the establishment of linear relationships. Recently, there is growing interest in the application of artificial neural networks (ANN) in the field of QSAR. The special interest in ANN arises from its ability to perform nonlinear mapping and its higher predictive power.[24−27] The relationships (linear or nonlinear) between molecular descriptors and the predicted affinities were shown in ANN dependence graphs.

The predictive power could be judged from a plot of predicted versus experimental affinities.[24] Higher correlation coefficients were given by the trained ANN architectures (average $r^2 = 0.92$, Table 2), in comparison with that of the MLR and PLS models (average $r^2 = 0.86$, Table 2).

In the ANN dependence graphs of the H^+ affinity model (Figure 2), PA increases approximately linearly with ES_N and RB. This is consistent with the positive coefficients found in the final MLR and PLS models.

In the ANN dependency graphs of the Na^+ affinity model, Na^+ affinity increases with the values of the molecular descriptors (DM, DM_S, and ES_S). The increasing trend is consistent with our MLR and PLS models (the coefficients are positive for all the descriptors). Nonlinear relationships were observed for DM_S and ES_S. The nonlinear relationships were explained with classical electrostatic theory. Ion−dipole interaction is chosen as an example. The ion−dipole interaction strength is proportional to $\mu \cos(\Phi)/r_\mu^2$, where μ is the permanent molecular dipole moment, Φ is the angle of deviation between the ion and the dipole vector (in deg), and r_μ is the distance between the ion and the center of the dipole moment vector (in angstrom).[48] As Φ and r_μ vary among the AA, nonlinear relationships were obtained.

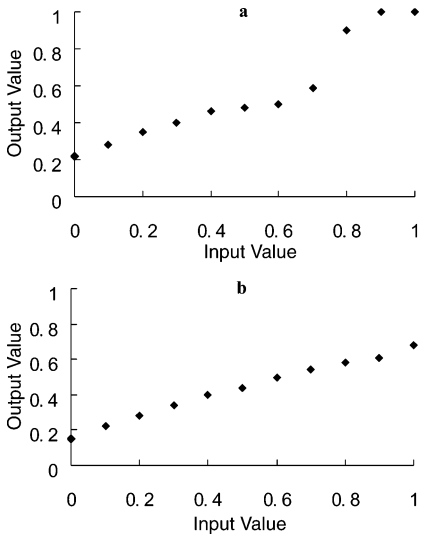

Figure 3. KA dependency graphs of the ANN models on (a) Cu^+ and (b) Ag^+ affinities.

For the ANN models on Cu^+ and Ag^+ affinities, all dependency graphs show increasing trends, except IP. This is consistent with the sign of the coefficients in the MLR and PLS models (negative for IP, positive for DM_S and KA). Cu^+ and Ag^+ affinities increase approximately linearly with all descriptors, except KA in the Cu^+ affinity model (Figure 3). It is of interest to note that the dependency graphs of IP and DM_S are similar for the Cu^+ and Ag^+ models but different in the case of KA (Figure 3). For Cu^+ affinity, the KA dependency graphs show a steady increase followed by a dramatic increase. However, no dramatic increase was observed for the case of the Ag^+ affinity model. This implied that Ag^+ gains less from the binding with AA, whose side chain is linear and flexible. Arg got the largest KA values; thus, we would like to get a better understanding about this discrepancy by investigating Arg in detail. It was reported that Arg surrounds the alkali metal cations and solvates the charge through multiple interactions with the N-terminus, the carbonyl oxygen, and terminal nitrogen of the side chain, except Cs^+; it cannot fit inside this pocket formed by Arg. Since Cs^+ sites on the side of the Arg, solvation is not nearly as effective as it is for Li^+.[49] As Ag^+ is larger than Cu^+ in size, one may expect Ag^+ may not fit as well as Cu^+ inside the pocket formed by Arg. We propose that this could be one of the explanations for the discrepancy. In conclusion, with the use of ANN models, higher predictive power could be obtained by considering the nonlinear relationships.

3.3. Zwitterionic versus Neutral. All AA are known to adopt the ZW form with deprotonated carboxyl and protonated amine or basic side chain in aqueous phases, as water preferentially stabilizes the ZW conformers over the CS one. In contrast, the inherently unstable ZW conformers are less stable than their CS forms in the gas phase. This has been unequivocally shown for Gly, Phe, as well as for the most basic Arg.[50] It was reported that the ZW conformers of AA could gain stability by complexation with metal cations. To study the binding forms of AA upon cationization, i.e., CS versus ZW, theoretical approaches were used.[30] However high-level ab initio calculations are always computationally demanding for large systems. Although it is relatively easy to measure binding affinities experimentally, it could be difficult to obtain direct or indirect evident on binding forms (CS vs ZW) experimentally. Several attempts were made to probe the binding forms experimentally using kinetic methods[51,52] or blackbody infrared radiative

dissociation approaches.[49,53] Recently, by coupling a free electron laser with a Fourier transform ion cyclotron resonance mass spectrometer,[54] vibrational fingerprints were used to provide direct evidence on the ZW structure of cationized proline. Here, we propose an alternative approach to provide a rough estimation on the binding form using QSAR models, with the steps mentioned above (Figure 1). In section 3.1, the descriptor values were first derived from the CS and ZW forms of AA. A stepwise MLR procedure was used to choose the best descriptors to develop MLR and PLS models. In general, we found descriptor values derived from the CS form of AA give better PA and Cu^+ affinity models with higher r^2(CV). For Na^+ and Ag^+ affinities models, descriptor values derived from the ZW form of Pro and the CS form for the rest of AA were used. However, nonlinear relationships have not been considered in the MLR and PLS models; thus, the results still need to be further confirmed with the ANN model.

In this section, the descriptor values derived from the CS and ZW forms of AA were input to the final trained ANN model to predict the affinities for the CS and ZW forms of AA. They were named as predicted CS affinity and predicted ZW affinity, respectively, in eq 12. These predicted affinities were compared with the actual experimental affinities (i.e., the actual affinity in eq 12). Then a CSvsZW index, defined as follows, is calculated for each AA:

$$\text{CSvsZW} = |\text{predicted CS affinity} - \text{actual affinity}| - |\text{predicted ZW affinity} - \text{actual affinity}| \quad (12)$$

If the CSvsZW index is positive, one may conclude that the experimentally determined affinity is better represented by the CS form of AA and vice versa. In other words, making use of experimental affinity, one can estimate the binding form of AA by QSAR models.

Comprehensive computational studies have been carried out on the possible structures of protonated/cationized glycine (Gly). Glycine is known to be in the CS form in all the protonated and cationized complexes.[13,16,29] Consistently, the CSvsZW indices for Gly are positive for all the cases studied here.

Proline is an interesting case. As reported previously, one would expect that Pro assumes the ZW form upon Na^+ and Ag^+ binding,[16,36,55,56] while H^+ and Cu^+ are stabilized by the CS form of Pro.[29,55,56] With the use of the final ANN model, consistent results were found, i.e., negative CSvsZW indices were obtained for Ag^+ and Na^+−Pro, but positive indices were found for H^+ and Cu^+−Pro complexes.

We have also extended our analysis to Na^+−Arg, whose binding form upon sodiation (CS or ZW) is still controversial. Ion mobility studies by Wyttenbach et al.[57] predicted that Arg adopts the ZW form. Blackbody infrared radiative dissociation and DFT calculations reported by Jockusch et al.[49] suggested that Arg is in the CS and ZW forms, respectively. Using a kinetic method, Wesdemiotis and co-workers predicted that Arg is in the CS form instead.[51,52] In this work, the CSvsZW index for Na^+−Arg was found to be nearly zero (−0.3). This may suggest that the CS and ZW forms of Arg have the same tendency to stabilize Na^+. This may explain why different results were reported in the literature.

No negative CSvsZW index was found for other protonated and cationized AA. Although no detailed studies on the CS versus ZW binding forms of these AA have been reported, it is generally believed that these AA have a lower tendency to form the ZW structure upon protonation and cationization compared with that of Pro (with a secondary amine) and Arg (with the most basic side chain).

We would like to emphasize that the QSAR approach discussed herein cannot replace the use of theoretical or experimental techniques cited previously. The absolute CSvsZW values, albeit inexact and nonrigorous, do appear to provide a fairly satisfactory estimation on the binding forms of CS versus ZW. We envision that QSAR could be used as a computationally less demanding method to estimate the binding forms of CS versus ZW forms for other complex systems in the future, once the binding affinities are determined.

3.4. Predicted Na^+ Affinities of Dihydroxyphenylalanine (DOPA). We believe the QSAR models developed herein may be beneficial for predicting affinities of other similar systems and bring new insight. In this section, we would like to vindicate this approach using sodiated dihyroxyphenylalanine (DOPA) as an example.

DOPA is a neurotransmitter that acts predominantly in the central nervous system and is associated with neurological diseases such as parkinsonism and schizophrenia.[58−60] The biosynthesis of DOPA begins with phenylalanine (Phe)/tyrosine (Tyr). Phenylalanine hydroxylase (PAH) converts Phe to Tyr. Tyrosine hydroxylase (TH) catalyzes the conversion of Tyr to DOPA, the rate-limiting steps in the biosynthesis of neurotransmitters. These aromatic amino acid hydroxylases are functionally and structurally closely related. The activity of these aromatic amino acid hydroxylases (PAH, TH) is subject to feedback inhibition by DOPA.[61] It was reported that the feedback inhibition requires the involvement of the amino acids residues Arg37 and Arg38 in TH.[62,63] Thus, it was hypothesized that the electrical field composted of the positive charge intrinsic to Arg is an essential regulator of the inhibition for TH.[62] It is generally believed that noncovalent interactions play a key role in the feedback inhibition. It would be interesting to systematically study the tendency for DOPA, Phe, and Tyr to interact with a positively charged cation. It was previously shown that sodium metal cations, mimicking positively charged sites, can be used to probe the intrinsic noncovalent interactions.[64] We believe the intrinsic noncovalent interactions between Na^+ and aromatic amino acids (Phe, Tyr) and DOPA may pave the way for more complete models describing these noncovalent interactions in chemistry and biology. Although Na^+−Phe and Tyr have been extensively studied, no data on Na^+−DOPA has been reported, rendering a systematic comparison difficult.

In virtue of the final Na^+ affinity ANN model reported herein, the Na^+ affinities of Phe, Tyr, and DOPA were systematically predicted to be 202.1, 202.4, and 202.5 kJ mol^{-1}, respectively. To our knowledge, this is the first reported Na^+−DOPA value. To check the quality of these predicted values, DFT affinities (Supporting Information) calculated for comparison give an absolute mean derivation of 2 kJ mol^{-1}. Thus, one may conclude that our predicted values are reasonable. Based on these predicted affinities, one may conceive that the tendency for Phe, Tyr, and DOPA to form electrostatic interactions with a positively charged site are comparable intrinsically.

4. Conclusions

To the best of our knowledge, this is the first QSAR study on proton, sodium, copper, and silver cation affinities of AA in the literature. We noted that the highest electrophilic superdelocalizability on nitrogen (major) and number of rotatable bonds (minor) are important molecular descriptors to explain the variation in PA. The variation in Na^+ affinity is attributed to the electrostatic interaction, especially ion−dipole interactions. The HSAB principles can be used to explain the Cu^+ and Ag^+ affinities. The variations in Cu^+ and Ag^+ affinities

would be explained by the softness, dipole moment, and degree of linearity properties of AA.

In comparison to MLR and PLS models, better predictive powers are obtained by using ANN models. The nonlinear relationships between molecular descriptors and affinities were discussed. We also demonstrated the use of QSAR models to study the (neutral vs zwitterionic) binding forms of AA upon protonation and cationization. Our results estimated that upon Na^+/Ag^+ bindings, Pro adopts the ZW form. On the other hand, other AA adopt the CS form upon protonation and cationization. We also illustrated the use of the Na^+ ANN model to predict the novel affinity for DOPA.

The use of QSAR models in predicting/elucidating the binding affinity/nature/form of protonated/cationized AA is illustrated in this work. Although gas-phase affinities of AA appear to be reasonably well established, studies of the affinities of peptides are much less definitive. It is believed that the intrinsic binding properties between proton/metal cation and peptides may have biological implications and values on their potential applications in mass spectrometric research fields. Our work on AA, as the building blocks of peptides and proteins, may pave the way for the application of QSAR models in large systems in the future.

Acknowledgment. This work is supported by the Areas of Excellence Scheme established under the University of Grants Committee of the Hong Kong Special Administrative Region, China (AoE/P-10/01). The small project funding approved by the Committee on Research and Conference Grants (CRCR) is gratefully acknowledged.

Supporting Information Available: Detailed statistics of the QSAR models and the DFT calculation of Na^+−DOPA/Phe/Tyr. This material is available free of charge via the Internet at http://pubs.acs.org.

References and Notes

(1) Stryer, L. *Biochemistry*, 3rd ed.; W. H. Freeman: New York; 1988.
(2) Lippard, S. J.; Berg, J. M. *Principles of Bioinorganic Chemistry*; University Science Books: Mill Valley, CA, 1994.
(3) Grier, N. *Disinfection, Sterilization and Preservation*, 3rd ed.; Lea & Febiger: Philadelphia, PA, 1983.
(4) Reiter, A.; Adams, J.; Zhao, H. *J. Am. Chem. Soc.* **1994**, *116*, 7827.
(5) Somogyi, Á.; Wysocki, V. H.; Mayer, I. *J. Am. Soc. Mass Spectrom.* **1994**, *5*, 704.
(6) Lee, S. W.; Kim, H. S.; Beauchamp, J. L. *J. Am. Chem. Soc.* **1998**, *120*, 3188.
(7) Bluhm, B. K.; Shields, S. J.; Bayse, C. A.; Hall, M. B.; Russell, D. H. *Int. J. Mass Spectrom.* **2001**, *204*, 31.
(8) Li, H.; Siu, K. W. M.; Guevremont, R.; Le Blanc, J. C. Y. *J. Am. Soc. Mass Spectrom.* **1997**, *8*, 781.
(9) Dongré, A. R.; Jones, J. L.; Somogyi, A.; Wysocki, V. H. *J. Am. Chem. Soc.* **1996**, *118*, 8365.
(10) Harrison, A. G. *Mass Spectrom. Rev.* **1997**, *16*, 201.
(11) Kish, M. M.; Ohanessian, G.; Wesdeomiotis, C. *Int. J. Mass Spectrom.* **2003**, *227*, 509.
(12) Cerda, B. A.; Wesdemiotis, C. *J. Am. Chem. Soc.* **1995**, *117*, 9734.
(13) Hoyau, S.; Ohanessian, G. *J. Am. Chem. Soc.* **1997**, *119*, 2016.
(14) Cerda, B. A.; Wesdemiotis, C. *Int. J. Mass Spectrom.* **1999**, *185/186/187*, 107.
(15) Lee, V. W. M.; Li, H.; Lau, T. C.; Guevremont, R.; Siu, K. W. M. *J. Am. Soc. Mass Spectrom.* **1998**, *9*, 760.
(16) Shoeib, T.; Siu, K. W. M.; Hopkinson, A. C. *J. Phys. Chem. A* **2002**, *106*, 6121.
(17) Bello-Ramírez, A. M.; Carreón-Garabito, B. Y.; Nava-Ocampo, A. A. *Epilepsia* **2002**, *43*, 475.
(18) Bello-Ramírez, A. M.; Buendía-Orozco, J.; Nava-Ocampo, A. A. *Fundam. Clin. Pharmacol.* **2003**, *17*, 575.
(19) Allinger, N. L.; Yuh, Y. H.; Lii, J.-H. *J. Am. Chem. Soc.* **1989**, *111*, 8551.
(20) Todeschini, R.; Consonni, V. *Handbook of Molecular Descriptors: Methods and Principles in Medicinal Chemistry*; Wiley-VCH: Weinheim, Germany, 2000; Vol. 11.
(21) (a) Goll, E. S.; Jurs, P. C. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 974. (b) Hemmateenejad, B.; Safarpour, M. A.; Miri, R.; Taghavi, F. *J. Comput. Chem.* **2004**, *25*, 1495.
(22) Karelson, M.; Lobanov, V. S.; Katritzky, A. R. *Chem. Rev.* **1996**, *96*, 1027.
(23) Ståhle, L.; Wold, S. *Prog. Med. Chem.* **1998**, *25*, 291.
(24) Douali, L.; Villemin, D.; Cherqaoui, D. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 1200.
(25) Douali, L.; Villemin, D.; Cherqaoui, D. *Curr. Pharm. Des.* **2003**, *9*, 1817.
(26) Su, Q.; Zhou, L. *J. Mol. Model.* **2006**, *12*, 869.
(27) Aoyama, T.; Suzuki, Y.; Ichikawa, H. *J. Med. Chem.* **1990**, *33*, 2583.
(28) Gancia, E.; Montana, J. G.; Manallack, D. T. *J. Mol. Graphics Modell.* **2001**, *19*, 349.
(29) Maksić, Z. B.; Kovačević, B. *Chem. Phys. Lett.* **1999**, *307*, 497.
(30) Siu, F. M.; Ma, N. L.; Tsang, C. W. *Chem. Eur. J.* **2004**, *10*, 1966.
(31) McMahon, T. B.; Ohanessian, G. *Chem. Eur. J.* **2000**, *6*, 2931.
(32) Maw, H. H.; Hall, L. H. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 1248.
(33) Cash, G. G.; Anderson, B.; Mayo, K.; Bogaczyk, S.; Tunkel, *J. Mutat. Res.* **2005**, *585*, 170.
(34) Hall, L. H.; Kier, L. B. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 1039.
(35) Gough, J. D.; Hall, L. H. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 356.
(36) Gapeev, A.; Dunbar, R. C. *Int. J. Mass Spectrom.* **2003**, *228*, 825.
(37) (a) Dunbar, R. C. *J. Phys. Chem. A* **2000**, *104*, 8067. (b) Ryzhov, V.; Dunbar, R. C.; Cerda, B.; Wesdemiotis, C. *J. Am. Soc. Mass Spectrom.* **2000**, *11*, 1037.
(38) Ruan, C.; Rodgers, M. T. *J. Am. Chem. Soc.* **2004**, *126*, 14600.
(39) Siu, F. M.; Ma, N. L.; Tsang, C. W. *J. Am. Chem. Soc.* **2001**, *123*, 3397.
(40) Shoeib, T.; Gorelsky, S. I.; Lever, A. B. P.; Siu, K. W. M.; Hopkinson, A. C. *Inorg. Chim. Acta* **2001**, *315*, 236.
(41) Deng, H.; Kebarle, P. *J. Am. Chem. Soc.* **1998**, *120*, 2925.
(42) Hall, L. H.; Kier, L. B. *Rev. Comput. Chem.* **1992**, *2*, 367.
(43) Hu, Q. N.; Liang, Y. Z.; Yin, H.; Peng, X. L.; Fang, K. T. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1193.
(44) De Mello Castanho Amboni, R. D.; Junkes, B. da S.; Yunes, R. A.; Heinzen, V. E. F. *J. Agric. Food Chem.* **2000**, *48*, 3517.
(45) Kier, L. B.; Hall, L. H. *Topological Indices and Related Descriptors in QSAR and QSPR*; Devillers, J., Balaban, A. T., Eds.; Gordon and Breach Science Publishers: Amsterdam, Netherlands, 1999.
(46) Hoyau, S.; Norrman, K.; McMahon, T. B.; Ohanessian, G. *J. Am. Chem. Soc.* **1999**, *121*, 8864.
(47) Pearson, R. G. *Inorg. Chim. Acta* **1995**, *240*, 93.
(48) Israelachvili, J. N. *Intermolecular & Surface Forces*, 2nd ed.; Academic Press: London, 1992.
(49) Jockusch, R. A.; Price, W. D.; Williams, E. R. *J. Phys. Chem. A* **1999**, *103*, 9266.
(50) Rak, J.; Skurski, P.; Simons, J.; Gutowski, M. *J. Am. Chem. Soc.* **2001**, *123*, 11695.
(51) Talley, J. M.; Cerda, B. A.; Ohanessian, G.; Wesdemiotis, C. *Chem. Eur. J.* **2002**, *8*, 1377.
(52) Cerda, B. A.; Wesdemotis, C. *Analyst* **2000**, *125*, 657.
(53) (a) Lemoff, A. S.; Bush, M. F.; Williams, E. R. *J. Phys. Chem. A* **2005**, *109*, 1903. (b) Dunbar, R. C. *Mass Spectrom. Rev.* **2004**, *23*, 127.
(54) Kapota, C.; Lemaire, J.; Maître, P.; Ohanessian, G. *J. Am. Chem. Soc.* **2004**, *126*, 1836.
(55) Marino, T.; Russo, N.; Toscano, M. *J. Phys. Chem. B* **2003**, *107*, 2588.
(56) (a) Hoyau, S.; Ohanessian, G. C. R. *Acad. Sci., C. R., Ser. IIc* **1998**, *1*, 795. (b) Shoeib, T.; Rodriquez, C. F.; Siu, K. W. M.; Hopkinson, A. C. *Phys. Chem. Chem. Phys.* **2001**, *3*, 853.
(57) Wyttenbach, T.; Witt, M.; Bowers, M. T. *Int. J. Mass Spectrom.* **1999**, *182/183*, 243.
(58) Cooper, J. R.; Bloom, F. E.; Roth, R. H. *The Biochemical Basis of Neuropharmacology*; Oxford University Press: New York, 1986.
(59) Birkmayer, W.; Riederer, P. *Understanding the Neurotransmitters: Key to the Workings of the Brain*; Springer-Verlag: New York, 1989.
(60) Mathews, C. K.; van Holde, K. E. *Biochemistry*; Benjamin/Cummings Publishing Co., Inc.: Menlo Park, CA, 1996.
(61) Maass, A.; Scholz, J.; Moser, A. *Eur. J. Biochem.* **2003**, *270*, 1065.
(62) Nakashima, A.; Mori, K.; Suzuki, T.; Kurita, H.; Otani, M.; Nagatsu, T.; Ota, A. *J. Neurochem.* **1999**, *72*, 2145.
(63) Nakashima, A.; Kaneko, Y. S.; Mori, K.; Fujiwara, K.; Tsugu, T.; Suzuki, T.; Nagatsu, T.; Ota, A. *J. Neurochem.* **2002**, *82*, 202.
(64) Kearney, P. C.; Mizoue, L. S.; Kumpf, R. A.; Forman, J. E.; McCurdy, A.; Dougherty, D. A. *J. Am. Chem. Soc.* **1993**, *115*, 9907.