PKa Calculation of the Positively Charged Amino acids in a Protein Chain, an Assessment of Theoretical Procedure

Ardabil University of Medical Sciences

M. Izadyar*, N. Zavvar, M. R. Housaindokht

Department of Chemistry, Faculty of sciences, Ferdowsi University of Mashhad, Mashhad, Iran

Izadyar@um.ac.ir

Ferdowsi University Of Mashhad

Abstract: In this study, the aqueous pK_a for cationic amino acid of lysine in a peptide chain which composed of eight glycine molecule have been calculated. HF and B3LYP methods using the 6-31G(d) combined with solvation energies that were computed by the SCRF-(CPCM/UFF) continuum models as implemented in Gaussian 09 computational package. pK_as were further calculated using two thermodynamic schemes, namely the direct method and the proton exchange method with the inclusion of solvent water molecule. The results of this research verify that the direct method is not suitable for computing pK_a of the amino acids in a peptide chain, while the other scheme in the presence of water molecule significantly improved the pk_a in comparison to the experimental data. The combination of the proton exchange scheme and CPCM-UFF model performed. Because of the convergence problems, the inclusion of large numbers of water molecule, the computation procedure in the solvated model produces failed. Comparison between the pK_a alpha-helix (α), beta-sheet (β) and random (R) structures of the studied proteins reveals that the strongest acidic character belongs to α, β and R, respectively at the HF and B3LYP levels of the theory which can be described according to natural bond orbital (NBO) and atom in molecule (AIM) analysis.

Keywords: pKa; SCRF; Lysine; Amino acid

Introduction

The field of computational chemistry is reaching the point where calculations at the level of chemical accuracy, within 1kcalmol^{-1}, are now possible. Numerous attempts to accurately calculate p*K*a values have been made, but none has achieved chemical accuracy [1-4]. p*K*a is obtained according to equation 1.

$$pK_a = -\log K_a \quad (1)$$

As pKa is related to Gibbs free energy (equation 2), the final version of equation 1 is equation (3).

$$\Delta G^\circ = -2.303RT\log K_a \quad (2)$$

$$pK_a = \Delta G^\circ/2.303RT \quad (3)$$

Because of the logarithmic relation between K_a and ΔG°, an error of1.36 kcalmol^{-1}in ΔG° gives an error of 1 pK_a unit. There are at least three sources of error in p*K*a calculations. The first error is the model which is used for calculation of pK_a, which generally involves a thermodynamic cycle such as scheme 1.The accuracy of the calculations for ΔG_{gas} and ΔG_s are the second and third major errors. In these thermodynamic cycles ΔG_{gas} is calculated with high level ab initio or density functional methods, and the ΔG_s values are calculated using a solvation method, typically a continuum dielectric approach. The p*K*a values can be determined from equation (3), where $\Delta G^\circ = \Delta G^\circ_{aq}$.

Accordingly, the accuracy in the calculation of relative p*K*a values depends strongly on the second and third main errors. Since we are going to investigate the second structure effects of proteins on their physical chemistry properties, in this paper we report absolute pK_a for random (R), beta-sheet (β) and alpha-helix (α) structures of positively charged amino acid of lysine including a glycine tailed using the thermodynamic cycles 1 and 2. Natural bond orbital (NBO) analysis and AIM are of importance to study the electronic density changes on these structures.

In this study for pK_as calculation, we examine the two approaches highlighted above, namely the direct method, the proton exchange method. The test set of molecules, positively charged amino acid of lysine in an amino acidic frame with eight molecule of glycine, have been designed to have a comprehensive understanding of second structural effects on pK_as. These molecules are further categorized as cationic as shown in Figure 1.

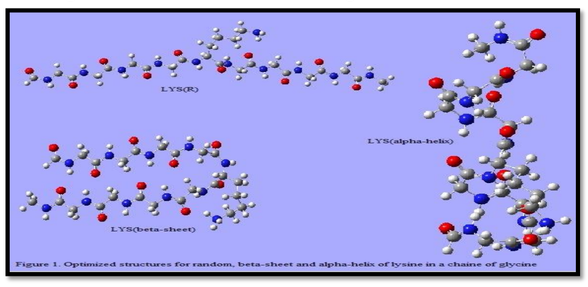
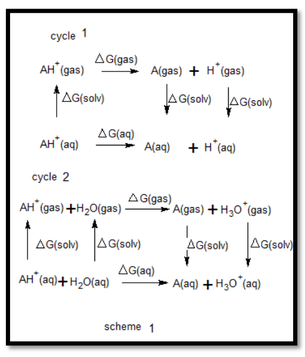


Figure 1. Optimized structures for random, beta-sheet and alpha-helix of lysine in a chains of glycine



scheme 1

Material and Methods

p*K*a calculations involve the representation of the acid dissociation process as a sum of several intermediate steps such as in the thermodynamic cycle shown above in Scheme 1.the free energy of acid dissociation in solution, ΔG°_{aq}, according to Hess's law can be calculated by equation(4).

$$\Delta G^\circ_{aq} = \Delta G^\circ_{gas} + \Delta\Delta G^\circ_{solv} \quad (4)$$

Where $\Delta\Delta G^\circ solv = \Sigma\Delta G_{solv,products} - \Sigma\Delta G_{solv,reactants}$. The symbol is used for a standard state of 1 molL^{-1} in any phase. K_a and pK_a can be obtained through the thermodynamic equation (3).

Equation 1 allows us to decompose the errors in the acidity constant into a gaseous component, ΔG_{gas}, and a solvation component, $\Delta\Delta G_{solv}$. Gas phase Gibbs free energies and solvation energies have been determined from the Hartree-Fock (HF) and post-HF methods such as density functional theory at the B3LYP level of the theory [5,6].

The situation mostly due to the difficulty of treating the solvent-solute inter actions is much less satisfactory in solution. Dissociation of positively charged acids in water involves the formation of charged and neutral molecules. Therefore short range intermolecular interactions such as ion-dipole and hydrogen bonding are important in solvation of charged species. For this solvation energies, ΔG°_{solv}, have been calculated using the conductor like polarizable continuum model(CPCM) [7]. By default, the CPCM method builds up the cavity using a united atom (UA) model, i.e. by putting a sphere around each solute heavy atom. We have used UFF radii (United force field model). Natural bond orbital charges have been used for all calculations of solvation energies. In CPCM model, the solvation energies partitioned into two components: electrostatic energies(ΔG_{elec}) and non-electrostatic energies (ΔG_{nelec}). The cavity, dispersion and repulsion energies form non-electrostatic interactions between solute and solvent.6-31G(d) basis set has been chosen since its geometries and energies is fairly good and accordance to couple cluster theory [8] by considering the size and type of studied molecules. All structures were fully optimized without any symmetry constrains using the HF and B3LYP method with6-31G(d) basis set as implemented in Gaussian 09 package [9].

Results and Discussion

We have chosen eight amino acids of glycine in the peptide's structures in which one molecule of lysine emerged to detect the second structure effects of proteins on the thermodynamics of cationic amino acids and also investigate the role of internal hydrogen bonding on pKa values. Experimental value for side chain pKa lysine is10.5.Figure 1shows the optimized structures of the studied proteins. Table 1 and 2 shows the calculated gas phase free energy and solvation free energy of each amino acid and its corresponding anion together with the standard Gibbs free energy of reaction (1), ΔG°. Comparison of the calculated solvation energies with the available experimental values shows that the CPCM results are consistent with the experimental values, though the yare slightly overestimated [7]. The standard Gibbs free energy of reaction (1), ΔG°, presented in table 1 and 2, have been obtained by two methods(Two thermodynamic cycles) described earlier. By using these energies, one can calculate the pKa values according to equation (2).

The theoretical values of pKa which are obtained by both methods have been shown in table 3. A comparison between the theoretical and experimental values reveals that there is small discrepancy between theory and experiment. Some sources for this discrepancy are as followed:

1.Theoretical method and basis set selection are two important factors which determine the accuracy extent of the computed pKas. Increasing the number of atoms in a structure lowers the level of computation. The corresponding error comes from the gas-phase energies. Different level of theories employed by ab initio calculations result in different gas-phase energies. The highest level of theory and the largest basis set are not practical for very large acids and for a numerous studied molecules from the computational point of view. For solving the convergence problems for 89-90 atoms in these proteins we limited all computations only two methods of HF and DFT/B3LYP level of the theory.

2. Although these two methods of thermodynamics for pKa evaluation in scheme 1 present similar results, there are some advantages and disadvantages for each method. The calculated pKa based on cycle 1can be reliable since many inherent systematic errors in calculation of free energies will be cancelled in determination of the total change of Gibbs free energy of proton-transfer reaction. But, these results are reference-dependent. Naturally, the non-substituted glycine can be chosen as reference, however, one can change it in order to improve the results. The results of cycle 2 in scheme 1 are reference-independent since the water molecule is involved. Although this model is more compatible with experiment, the large uncertainty accompanied with Gibbs free energy of solvation of$H3O^+$ can seriously affect the results [10,11].

3. The third error appears when we use different models of solvation in order to calculate the Gibbs free energy of solvation of species. Several versions of PCM and SMD models have been frequently used for these calculations. This part of uncertainty is usually larger than for the gas-phase energies.

4. The real chemistry of dissociation of acids is an important factor which must be taken into account. For example, there is a probability of finding the amino acids in dimmer form in the gas phase and in solution phase, therefore the dimerism forms of the acid alters the pka of corresponding acid and for producing more real data it should be also considered.

5. Finally considering the lysine as a single molecule and comparing its thermodynamic parameters with a nearly long chain of added glycine molecules, induces that the intermolecular interaction contribution on these parameters is not negligible and it is the source of some degrees of discrepancies between the theoretical data and experiments. Therefore, we focus on the hydrogen bonds which are important in the formation of the second structures of the proteins.

Table1. Computed Gibbs free energies at the HF/ 6-31G (d) in Hartree·

	G(gas)	ΔG (solv)	ΔG (aq)
Random+	-2280.194012	-0.132523	-2280.326535
Random-H	-2279.813691	-0.062899	-2279.87659
Alpha helix+	-2280.144979	-0.157128	-2280.302107
Alpha helix -H	-2279.778516	-0.074867	-2279.853383
Beta sheet +	-2280.161449	-0.118771	-2279.897287
Beta sheet-H	-2279.775308	-0.055345	-2279.830653
ΔG(aq) [H+]= -0.428678	ΔG (aq)[H3O+]= -76.867528	ΔG(aq) [H2O]= -76.441968	

Table2. Computed Gibbs free energies at the B3LYP/6-31G (d) in Hartree.

	G(gas)	ΔG (solv)	ΔG (aq)
Random+	-2293.689966	-0.119265	-2293.809231
Random-H	-2293.307773	-0.052493	-2293.360266
Alpha helix+	-2293.644567	-0.142977	-2293.787544
Alpha helix -H	-2293.274527	-0.059728	-2293.334255
Beta sheet	-2293.671043	-0.105830	-2293.776873
Beta sheet-H	-2293.280099	-0.042740	-2293.322839
ΔG(aq)[H+]= -0.428678	ΔG(aq)[H3O+]= -76.867528	ΔG (aq)[H2O]= -76.441963	

Table 3. Computed pKa for lysine in the HF/B3LYP

	Pka1 (H^+)	Pka2 ($H3O^+$)
Random	9.81/9.36	9.51/9.06
Alpha	9.24/11.07	8.9/11.05
beta	9.64/11.69	9.34/11.40

Computed data show that the acidic character of alpha-helix is more than beta-sheet and random structure. As in pka evaluation H transfer plays an important role on acidic property of an acid, acidic H ability for hydrogen bond formation alters the pKa. NBO analysis and AIM investigation on these structures confirmed the hydrogen bonding in alpha is stronger and more than in beta and random. Therefore we can conclude that not only this discrepancy between the theory and experiments is not error, but also it is a driving force for better analysis of H bonding in second structures of the proteins.

Conclusion

In this study, the aqueous pK_a for cationic amino acid of lysine in a peptide chain which composed of eight glycine molecule have been calculated. pK_as were further calculated using two thermodynamic schemes, namely the direct method and the proton exchange method with the inclusion of solvent water molecule. The results of this research verify that the direct method is not suitable for computing pK_a of the amino acids in a peptide chain, while the other scheme in the presence of water molecule significantly improved the pk_a in comparison to the experimental data. Comparison between the pK_a alpha-helix (α), beta-sheet (β) and random (R) structures of the studied proteins reveals that the strongest acidic character belongs to α, β and R, respectively at the HF and B3LYP levels of the theory which can be described according to natural bond orbital (NBO) and atom in molecule (AIM) analysis.

References

[1] M.D. Liptak, G.C. Shields, J. Am. Chem. Soc. 123 (2001) 7314.

[2] C. Scharnagl, R.A. Raupp-Kossmann, J. Phys. Chem. B 108 (2004)477.

[3] D.M. Chipman, J. Phys. Chem. A 106 (2002) 7413.

[4] M.D. Liptak, K.C. Gross, P.G. Seybold, S. Feldgus, G.C. Shields, J.

[5] A.D. Becke, Phys. Rev. A, 38 (1988) 3098.

[6] C. Lee, W. Yang, R.G. Parr, Phys. Rev. B, 371 (1988) 785.

[7] M. Cossi, V. Barone, R. Cammi, J. Tomasi, Phys. Lett. 255 (1996) 327.

[8] L.J.,Bartolotti, K., Fluchick, Reviews in Computational Chemistry,Vol. **7**, VCH: New York(1996) 187.

[9] Gaussian, Inc., Wallingford CT, Gaussian 09, Revision A.02, (2009).

[10] M.W. Palascak, G.C. Shields, J. Phys. Chem. A 108 (2004) 3692.

[11] J.R. Pliego, J.M. Riveros, Chem. Phys. Lett. 332 (2000) 597.