THEORETICA CHIMICA ACTA

Theoret. Chim. Acta (Berl.) 51, 323—329 (1979)

Non-Empirical SCF MO Studies on the Protonation of Biopolymer Constituents

I. Protonation of Amino Acids*

Paul G. Mezey**

Department of Chemistry and Chemical Engineering,
University of Saskatchewan, Saskatoon, Canada S7N 0W0

János J. Ladik and Sándor Suhai

Chair of Theoretical Chemistry at the Friedrich-Alexander-University Erlangen-Nürnberg,
D-8520 Erlangen, Federal Republic of Germany; and
Laboratory of the National Foundation for Cancer Research
at the Chair of Theoretical Chemistry, University of Erlangen-Nürnberg,
D-8520 Erlangen, Federal Republic of Germany

Side-chain proton affinities for a series of peptide-forming amino acids have been calculated using *ab initio* Hartree–Fock–Roothaan SCF method; the order of proton affinities is Arg > His > Ser ≈ Tyr ≫ Lys. Protonation of some side-chains in a protein may introduce new energy levels in the band gaps thus fundamentally altering the conduction properties of the proteins.

Key words: Biopolymer constituents, protonation of ~

1. Introduction

Most biochemical processes involving amino acids take place in aqueous media, and protonated forms of amino acids are certainly of some importance as intermediates. While in most free amino acids protonation is likely to occur at the amino or carboxyl moieties, for amino acids built in a peptide chain, protonation at electronegative centres in the side-chains becomes particularly important.

Side-chain protonation of amino acid units may change the conformational-structural properties of a polypeptide substantially, simply by the presence of the additional hydrogen atom and by all its interactions with the surroundings, and also by altering the overall electronic structure of the peptide.

All these effects are dependent on the actual site of protonation. In biologically important peptides there is a great variety of amino acid side-chains and the relative ease of protonation varies from side-chain to side-chain. Since these variations are reflected in the proton affinities, it is of some interest to study the

* This paper is dedicated to Professor H. Hartmann for his 65th birthday.

** Guest at the Laboratory of the National Foundation for Cancer Research at the Chair of Theoretical Chemistry, University of Erlangen-Nürnberg.

0040-5744/79/0051/0323/$01.60

side-chain proton affinities of amino acids particularly prone to protonation, i.e. those containing electronegative centres besides the NH_2 and COOH groups.

The importance of semiconductive properties of proteins in explaining their biochemical behaviour has been recognized for some time, the earliest suggestions are dated back to the pioneering works of Szent-Györgyi and Laki [1]. Following the experimental demonstration that the conduction in proteins is of the electronic type [2], there has been a revived theoretical interest in the energy band structure of proteins [3–8]. *Ab initio* crystal orbital calculations [7] have indicated that simple periodic protein models have forbidden band gaps of ~ 10 eV width, and the presence of (donor or acceptor) impurity centres in real proteins has been suggested to account for the experimentally observed semi-conducting properties [9]. The simplest chemical reactions that may generate impurity centres in the peptide chain are proton-transfer and tautomerization in the amino acid side-chain. Only a few natural amino acids, while built in a peptide chain, are capable of undergoing genuine tautomerization processes (a notable example is histidine), however, in many amino acid side-chains there are electronegative centres that are possible targets for protonation reactions. Since proton transfer may cause a significant shift in the energy levels of the amino acid unit, such an "irregular" side-chain may generate few discrete energy levels in the forbidden band gap, thus fundamentally altering the conduction properties of the polypeptide.

In addition to altering the conduction properties of proteins, a tautomerization or protonation reaction of some of the side-chains may substantially modify several properties of the polypeptide, ranging from conformational stabilities of chain-segments on the microscopic level to the stability of colloid solutions on the macroscopic level. Consequently, it is of some interest to study the effects of side-chain tautomerization and protonation reactions on the total energy and MO energy levels of biologically important amino acids. In the present *ab initio* study the protonation and tautomerization reactions of five amino acids, serine, lysine, arginine, histidine and tyrosine, have been investigated in some detail.

2. Method

Throughout this study the Hartree–Fock–Roothaan LCAO SCF MO method [10] was applied, using an adaptation of the Gaussian 70 program system [11] and a set of contracted 6^s3^p (minimum) Gaussian basis functions [12]. Due to the large size of the amino acids and the corresponding side-chain protonated ions, only the serine structures were considered in full. For all the other structures the actual calculations were carried out on simplified models, where only the side chains were considered and the HOOC—CH—NH_2 moieties were replaced by a H atom. No attempt was made to optimize the geometries of the above amino acid models, and standard geometries were used throughout [13]. The geometrical information on these amino acid models is summarized in Figs. 1 and 2.

The serine, arginine and histidine structures marked with A are those most commonly observed experimentally. Serine and arginine structures marked with B have

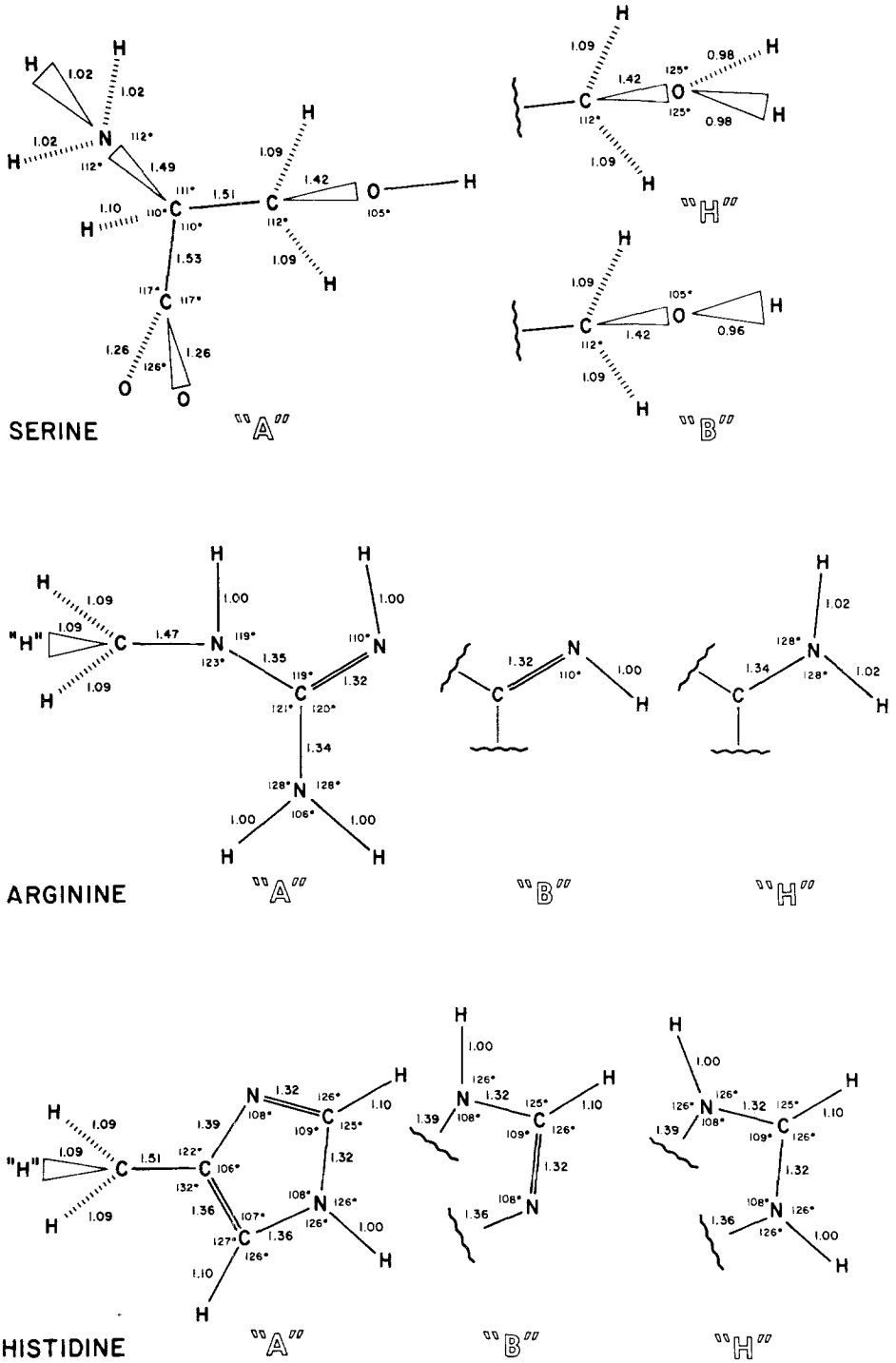

Fig. 1. Geometries of various serine, arginine and histidine model compounds and protonated forms

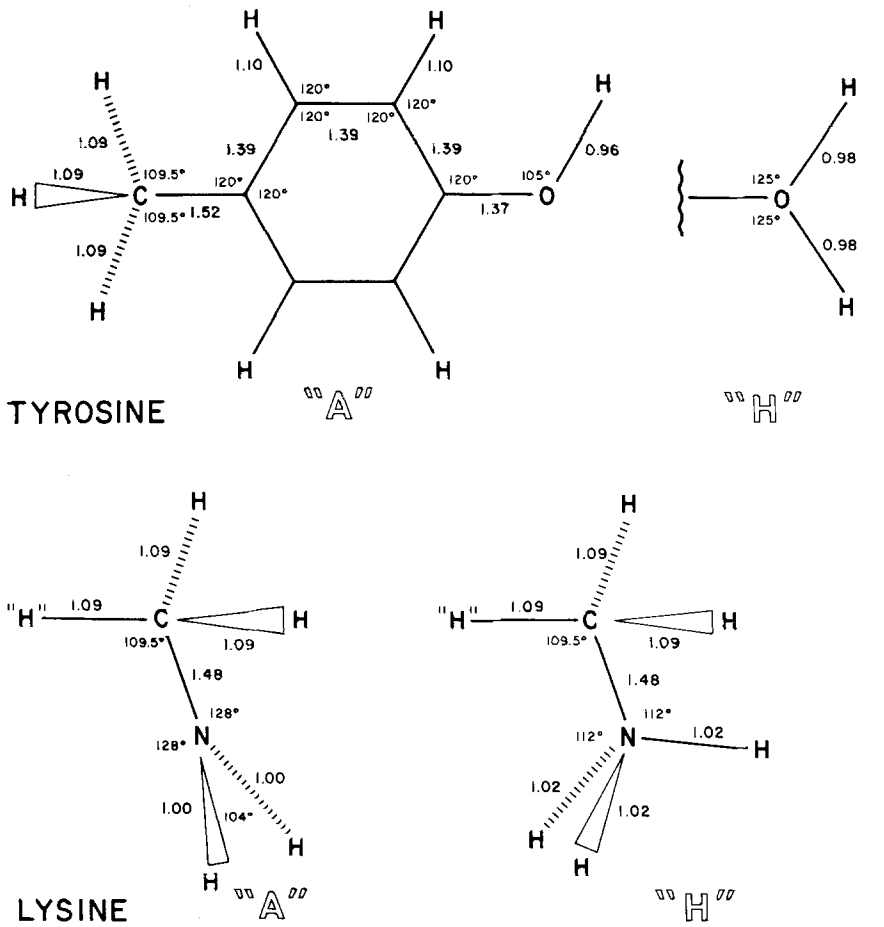

Fig. 2. Geometries of lysine and tyrosine model compounds and protonated forms

been derived by a 180° rotation of the side-chain C—O and C=N bonds, respectively, thus structures marked with B correspond to conformational and geometric isomers, respectively. Histidine B is a genuine tautomer of histidine A, with a H in the N3 position rather than one in the N1 position (see Fig. 1).

The calculations were carried out on the Cyber 72 computer of the University of Erlangen-Nürnberg, and in part on the IBM 370/158 computer at the University of Saskatchewan, Saskatoon.

3. Results and Discussion

The calculated total energies and proton affinities of various species are listed in Table 1. The calculated energy differences between conformers or tautomers A and B are relatively small in all three cases: 1.39 kcal/mole, 1.24 kcal/mole and 1.46 kcal/mole for serine, arginine and histidine, respectively. Calculated energy

Table 1. Calculated total energies and proton affinities of various amino acid models

Amino acid model compound	Total energy (a.u.)	Total energy of protonated form (a.u.)	Proton affinity (kcal/mole)
Serine A	−391.35638		236.42
Serine B	−391.35416	−391.73314[a]	237.81
Lysine	−94.00651	−94.18215	110.22
Arginine A	−240.02020	−240.52352[a]	315.84
Arginine B	−240.02217		314.60
Histidine A	−260.55994	−261.02392[a]	291.15
Histidine B	−260.56226		289.70
Tyrosine	−340.31255	−340.68943	236.50

[a] Protonation of both forms A and B results in the same ion.

differences between neutral and protonated species indicate that the arginine side-chain is the most likely site for protonation, although large proton affinity values have been obtained for histidine, serine and tyrosine as well. Lysine is exceptional with a relatively low proton affinity value of 110 kcal/mole.

In Table 2 the orbital energies of the four highest occupied MOs are collected for all species, together with the MO energies for the two lowest unoccupied orbitals. In Table 3 the variations in the above MO energy levels, caused by protonation, are listed. One may expect the appearance of discrete energy levels in the forbidden gap regions due to both types of impurities, i.e. the presence of isomeric forms (or different conformers) and protonated forms.

Table 2. MO energy levels of amino acid model compounds and protonated forms **K**

	HOMO energies (a.u.)				LUMO energies (a.u.)	
Structure	4th	3rd	2nd	1st	1st	2nd
Serine A	−0.35836	−0.19064	−0.18021	−0.12358	0.39695	0.44208
Serine B	−0.33557	−0.16806	−0.15444	−0.09778	0.41284	0.46689
Serine, protonated	−0.51772	−0.35425	−0.33953	−0.27995	0.11916	0.22994
Lysine	−0.54771	−0.53144	−0.48250	−0.28118	0.56597	0.64707
Lysine, protonated	−0.93786	−0.85150	−0.78598	−0.74150	0.13061	0.25066
Arginine A	−0.49757	−0.33937	−0.29412	−0.22188	0.37475	0.55393
Arginine B	−0.49348	−0.33822	−0.29697	−0.22101	0.37527	0.55323
Arginine, protonated	−0.72930	−0.71215	−0.58164	−0.54792	0.06603	0.30503
Histidine A	−0.46643	−0.32109	−0.30671	−0.26723	0.32441	0.39215
Histidine B	−0.49758	−0.32268	−0.32168	−0.25926	0.31580	0.40088
Histidine, protonated	−0.70872	−0.68543	−0.63372	−0.51030	0.03267	0.12718
Tyrosine	−0.41211	−0.40437	−0.28594	−0.23929	0.26603	0.28088
Tyrosine, protonated	−0.59863	−0.56459	−0.45897	−0.44668	0.08300	0.09315

Table 3. Variations of MO energy levels of amino acid model compounds upon protonation

	ΔE_i of HOMOs (a.u.)				ΔE_i of LUMOs (a.u.)	
Structure	4th	3rd	2nd	1st	1st	2nd
Serine A	−0.15936	−0.16361	−0.15932	−0.15637	−0.27579	−0.21214
Serine B	−0.18215	−0.18619	−0.18509	−0.18217	−0.29369	−0.23695
Lysine	−0.39015	−0.32006	−0.30348	−0.46032	−0.43536	−0.39641
Arginine A	−0.23173	−0.37278	−0.28752	−0.32604	−0.30872	−0.24890
Arginine B	−0.23582	−0.37393	−0.28467	−0.32691	−0.30924	−0.24820
Histidine A	−0.24229	−0.36434	−0.32701	−0.24307	−0.29174	−0.26497
Histidine A	−0.21114	−0.36275	−0.31204	−0.25104	−0.28313	−0.27370
Tyrosine	−0.18652	−0.19022	−0.17303	−0.20739	−0.18303	−0.18773

While the differences in the MO energies between the A forms and B forms are approximately 0.02, 0.004 and 0.03 hartree for serine, arginine and histidine, respectively, protonation causes a shift in the MO energy levels approximately 10 times larger. It is surprising that lysine, with the smallest proton affinity value in the series, shows the largest changes in the MO energies if protonated (see Table 3). This suggests that there is a fortuitous partial cancellation in the variations of the actual MO contributions to the total molecular wavefunction. A similar trend is apparent for the other species as well if one compares the calculated MO and total energies.

It is remarkable that though an A $\rightleftharpoons$ B isomerization or tautomerization results in a much larger (approximately 10 times larger) change in the orbital energies than in the total energy, these MO energy shifts are still rather small. On the other hand, the protonation-induced changes in the MO energies are of the same order of magnitude and in some cases smaller than total energy changes due to protonation. These findings suggest that isomerization or conformational changes on the one hand cause insertion of new energy levels approximately only 0.03 hartree $\sim$1 eV from the nearest band edge (either in the gap, or inside the bands). If the gap regions are approximately 10 eV wide (as earlier crystal orbital studies [7] have indicated for simple protein models), such shifts cannot account for fundamental changes in the conducting properties of the polymer. On the other hand, as Table 3 shows, protonation may introduce approximately 0.2–0.3 hartree, i.e. $\sim$5–8 eV shifts towards smaller energies and thus the LUMO levels can be situated only by 2–3 eV above the upper edge of the valence band of a polypeptide chain. These new levels may become in this way very effective acceptor levels.

This comparison suggests that protonation of few side-chains in the protein may indeed convert a polypeptide into a potential semiconductor. In addition, an isomerization or conformational change may also cause minor changes in the energy band structure of the protein. In real proteins a combination of both effects is likely to be responsible for the observed properties, partly because the net effect of the protonation-deprotonation may be equivalent for a form A $\rightleftharpoons$ form B isomerization or conformational change.

4. Conclusions

Ab initio minimum basis set calculations on various neutral and protonated forms of amino acid models (serine, lysine, arginine, histidine and tyrosine) indicated that in proteins the arginine side chain is the most likely target for protonation. The order of calculated side-chain proton affinities is Arg > His > Ser $\approx$ Tyr $\gg$ Lys. MO energy levels are strongly affected by protonation, but only to a minor extent by isomerization or conformation change in the side-chains. It is likely that protonated side-chains may act as "impurity-centres" in a protein, leading to the appearance of new energy levels in the "forbidden" band gap regions, thus altering the conductance properties of the polypeptide.

Acknowledgement. The authors express their thanks to the National Foundation for Cancer Research, USA, and the Natural Sciences and Engineering Research Council of Canada, for continuing financial support.

References

1. Szent-Györgyi, A.: Nature **148**, 157 (1941); Laki, K.: Stud. Inst. Med. Chem. Unit. Szeged **2**, 43 (1942)
2. Eley, D. P., Spivey, D. I.: Trans. Faraday Soc. **56**, 1432 (1960)
3. Ladik, J.: Acta Phys. Hung. **11**, 239 (1960); Ladik, J.: Nature **202**, 1208 (1964)
4. Yomosa, S.: Biopolymers Symp. **1**, 1 (1964)
5. Fujita, H., Imamura, J.: J. Chem. Phys. **53**, 4555 (1970)
6. Beveridge, D. L., Jano, I., Ladik, J.: J. Chem. Phys. **56**, 4744 (1972)
7. Suhai, S., Ladik, J.: Theoret. Chim. Acta (Berl.) **34**, 147 (1974); Suhai, S., Collins, T. C., Ladik, J.: Biopolymers, in press
8. Suhai, S.: Biopolymers **13**, 1731 (1974)
9. Szent-Györgyi, A.: Intern. J. Quantum Chem. Quantum Biol. Symp. **3**, 45 (1976)
10. Roothaan, C. C. J.: Rev. Mod. Phys. **23**, 69 (1951)
11. Hehre, W. J., Lathan, W. A., Ditchfield, R., Newton, M. D., Pople, J. A.: QCPE program 236, Indiana University, Bloomington, Indiana
12. Hehre, W. J., Stewart, R. F., Pople, J. A.: J. Chem. Phys. **51**, 2657 (1969)
13. Gurskaya, G. V.: The molecular structure of amino acids. New York: Plenum Publishing Corporation 1968

Received January 10, 1979