Neutral *vs.* zwitterionic form of arginine—an *ab initio* study

Zvonimir B. Maksić *[a,b] **and Borislav Kovačević**[a]

[a] *Quantum Organic Chemistry Group, Ruđer Bošković Institute, PO Box 1016, 10000 Zagreb, Croatia. E-mail: zmaksic@spider.irb.hr*
[b] *Faculty of Science and Mathematics, The University of Zagreb, Marulićev trg 19, 10000 Zagreb, Croatia*

Received (in Cambridge, UK) 25th March 1999, Accepted 31st August 1999

The problem of the intramolecular proton transfer isomerism in arginine, leading to conventional neutral and zwitterionic forms of this compound, is addressed by high level theoretical models. It is shown that arginine has two neutral and two zwitterionic isomers implying that there exist two additional unconventional isomers, which have not been identified so far. It appears also that the most stable neutral isomer is energetically more favourable than both zwitterions, which implies that the former should be preferred in the gas phase. Examination of atomic charges obtained by the electron density partitioning techniques reveals that the charge distributions of neutral and zwitterionic isomers are not as widely different as expected. This finding is counterintuitive, since it contradicts the classical notion of chemical bonding and a customary picture of zwitterions involving two local complementary fragments possessing unit charges of opposite sign. The true distribution of the electron density is more uniform and quite similar to that of the neutral form. The proton affinity of arginine is estimated to be 249 kcal mol^{-1}. Hence, it follows that arginine is a very basic compound although it belongs to a family of 20 fundamental α-amino acids. A very high proton affinity is interpreted in terms of the resonance effect spurred by protonation in the guanidine moiety and by a strong hydrogen bonding taking place in the protonated form.

Introduction

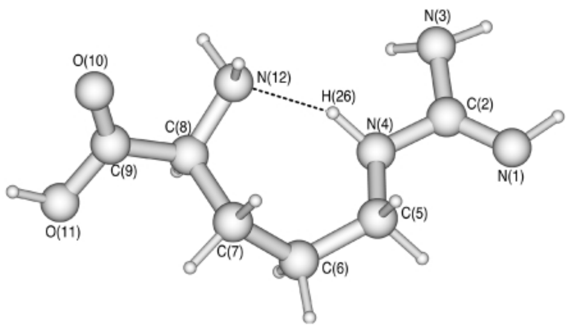
Twenty α-amino acids (AAs) provide the elementary building blocks of all proteins found in living species. They are characterized by a carboxylic acid fragment and amino group substituted at the α-carbon atom whereas the side chain varies from one AA to another leading to remarkable diversity and versatility. It is common knowledge that AAs exist in neutral forms in the gas phase[1–3] implying that the bifocal zwitterionic charge localization is unfavourable in the absence of additional stabilization effects provided either by crystal forces or by solvent molecules.[4] This point of view was challenged recently by Price *et al.*[5] who claimed, on the basis of Fourier transform-mass spectrometry (FTMS) measurements and BLYP/6-31G* calculations, that the most stable configuration of gaseous arginine (**R**) was its zwitterion. This evidence is not quite conclusive, however, because the applied theoretical approach was not accurate enough, whereas the experimental information was obtained from the proton-bound dimer complex **RH**$^+$**R**, where both guanidine groups were protonated and the carboxylic group of one **R** was deprotonated at the same time. In other words, one of the coupled arginine molecules is protonated at the imino nitrogen whereas the other appears in a form of the zwitterion. It is quite conceivable that in this particular experiment **R** occurs finally as the zwitterion, but this does not automatically imply that it is the energetically most favourable configuration in general. The work of Price *et al.*[5] has prompted Saykally *et al.*[6] to perform infrared cavity ringdown laser absorption spectroscopy investigations on **R**, which led to the conclusion that arginine exsisted in the neutral configuration.

Stimulated by this intriguing but controversial experimental evidence, we performed a series of relatively high level theoretical calculations on possible arginine conformations and examined its most favourable protonated form. A particularly interesting question of the existence of other neutral tautomers of arginine is addressed too. Anticipating forthcoming results we can state that the most stable conventional neutral configuration and the customary zwitterionic conformer of arginine are very close in total molecular energy, the former being slightly more stable. Another remarkable finding is the identification of two additional isomers possessing neutral and zwitterionic distribution of formal atomic charges. Finally, a word on terminology is appropriate here. Strictly speaking both “neutral” and zwitterionic forms are in fact neutral molecules. The only difference is that the intramolecular charge transfer is somewhat more pronounced in the latter form. We shall keep with this widely adopted terminology, although it is clear that a more adequate distinction between these two forms of isomers is obviously needed in the future.

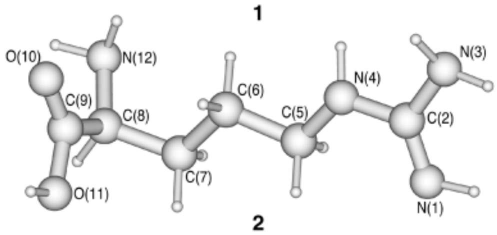
Methodology

The initial search of the Born–Oppenheimer energy hypersurface was performed at the economical Hartree–Fock level employing a modest 6-31G* basis set in order to get an idea about positions of true minima. Since the electron correlation apparently plays a significant role particularly in the zwitterionic and protonated forms, additional geometry optimizations have been carried out at the MP2(fc)/6-31G* degree of sophistication. This model gives certainly a better spatial description of the studied systems. A choice of the appropriate basis set is important too. It is well established by now that the 6-31G* basis set is not flexible enough to describe properly the loosely bound lone pair of nitrogen, particularly if the energetic properties are desired. For instance, a good reproduction of the experimental proton affinities of nitrogen atoms in various chemical environments required a more advanced MP2(fc)/6-311+G**//HF/6-31G* model.[7–9] Consequently, the total molecular energies of various conformers of **R** were estimated by the single point MP2(fc)/6-311+G**// MP2(fc)/6-31G* approach. Subsequently, the structural parameters and energies were computed in a fully consistent way by the

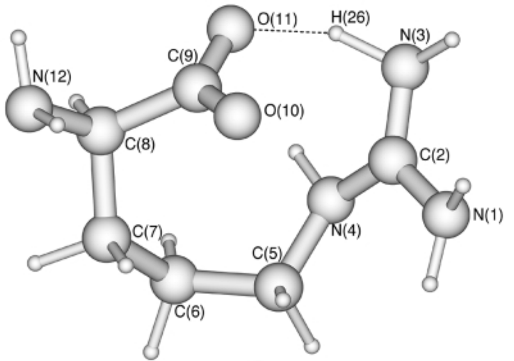
This journal is © The Royal Society of Chemistry 1999



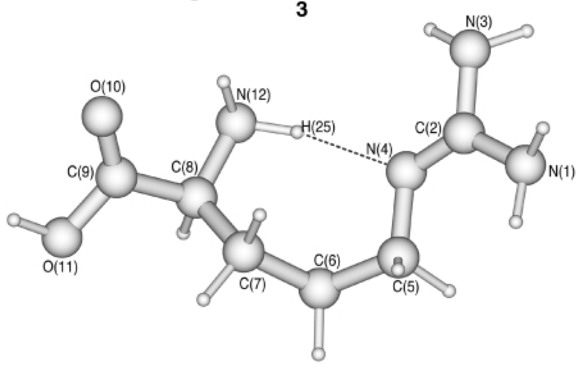
1



2



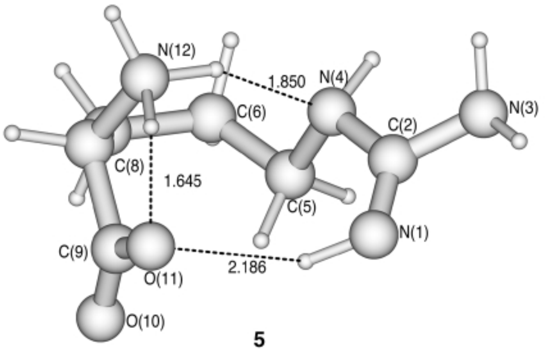
3



4

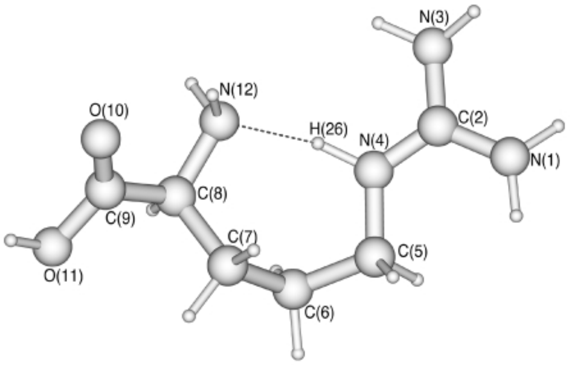
Fig. 1 Schematic representation and numbering of heavy atoms in the conventional neutral arginine **1**, its zig-zag conformation **2** the zwitterionic isomer **3** and the newer neutral tautomer **4**.

MP2(fc)/6-311+G** model in order to obtain the relative stability of neutral and zwitterionic configurations as reliably as possible at reasonable costs. Finally, sensitivity of the relative energies to the even more flexible 6-311+G(2df,p) basis set is examined within the MP2(fc)/6-311+G(2df,p)// MP2(fc)/6-31G* model. The HF/6-31G*, MP2(fc)/6-31G*, MP2(fc)/6-311+G**// MP2(fc)/6-31G*, MP2(fc)/6-311+G** and MP2-(fc)/6-311+G(2df,p)// MP2(fc)/6-31G* models are denoted from here onwards, as M(I), M(II), M(III), M(IV) and M(V), respectively. The zero point vibrational energies (ZPVEs) were estimated at the M(I) level by using a customary weighting factor of 0.89. It should be emphasized that all examined species **1**–**5**, **1p** and **2p** are true minima on the potential energy hypersurfaces as verified by the vibrational analyses. All calculations have been executed using the GAUSSIAN 94 suite of programs.[10]

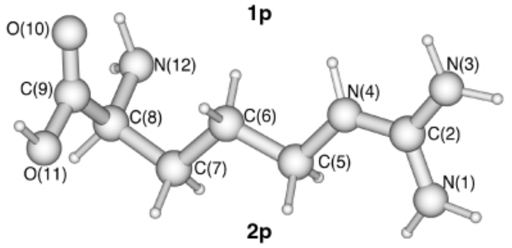


5

Fig. 2 Schematic representation and numbering of atoms in the unconventional zwitterion **5**.



1p



2p

Fig. 3 The most stable protonated form of arginine **1p** and its zig-zag conformation **2p**.

Results and discussion

Structural parameters

The neutral conformation of **R** denoted by **1**, its unfolded zig-zag form **2**, the conventional zwitterion of arginine **3**, the unconventional neutral and zwitterionic isomers **4** and **5**, respectively, and two protonated species **1p** and **2p** are schematically depicted in Figs. 1–3. the relevant structure data are given in Table 1. The zig-zag conformation serves the purpose of estimating the strength of the H-bond in **1**. It appears that the hydrogen bonding contributes about 1.4 kcal mol^{-1} to the overall stability of the parent arginine (*vide infra*). In the zwitterion configuration **3** the proton of the COOH fragment is transferred to the terminal imino group. One observes that **3** assumes a quasi-ring structure where the Coulomb attraction between the formal COO$^-$ anion and the protonated guanidine cation leads to an energetically favourable interaction. Furthermore, an additinoal H-bond occurs (Fig. 1) offering some additional stabilization. Since the M(II) model represents a good compromise between reliability and practicability in estimating the relative stability of isomers (see later), we have used it in a search of other possible neutral and zwitterionic forms. Much to our surprise we found an additional tautomer **4** and a novel zwitteionic isomer **5**. This is a remarkable finding because the existence of two neutral tautomers (**1** and **4**) of arginine has not been discussed in the literature so far to the best of our knowledge. The neutral tautomer **4** depicted in Fig. 1 has the imino nitrogen N(4) placed in the middle of the system thus forming a hydrogen bond with the H(25) atom. In contrast, the imino nitrogen in the slightly more stable tautomer **1** is peripherally placed sticking out of the compound. The zwitterionic form **5** is formed by the proton transfer from the carboxylic to the amino group attached to the α-carbon atom (Fig. 1). It is stabilized by three hydrogen bonds, the one providing N(12)–H–––O(11) bridge being particularly short and strong. In spite of that, isomer **5** is less stable than **3** by some 15 kcal mol^{-1} (*vide infra*). The protonated **R** species **1p** differs relatively little from the parent compound **1** at the AA core, the main change taking place at the guanidine end of the side chain. The same holds for the protonated zig-zag form **2p** where a new $N(1)H_2$ amino group is formed. A close scrutiny of the presented bond distances reveals that values offered by the Hartree–Fock M(I) model deviate sometimes considerably from the corresponding MP2 M(II) and M(IV) bond lengths. This is particularly pronounced in the description of some double bonds like N(1)=C(2) and C(9)=O(10) or in some single bonds such as, *e.g.* the C(9)–O(11) bond in **1** and related systems. On the other hand the differences in bond distances separating heavy atoms estimated by the M(II) and M(IV) models are practically negligible as shown by the average absolute deviation of only 0.003 Å. For that reason we did not carry out M(IV) model calculations for **2** and **2p**. Also, the rest of the discussion will be based on the M(II) model geometries. However, it should be pointed out in passing that there is one notable exception where the M(II) and M(IV) models do disagree to a larger extent. This is the case of the hydrogen bond contact O(11)– H(25) in the zwitterionic form **3**. In this context it should be mentioned that the M(I) model appreciably overshoots the H-bond distances. In contrast, all three models yield compatible and quite reasonable bond angles. This is in harmony with a known fact that it is easier to reproduce molecular shape than molecular size. An interesting structural feature of **3** is that two CO bonds of the O(10)–C(9)–O(11) fragment are not equivalent, in spite of an apparent tendency to be equal. In fact, C(9)–O(10) and C(9)–O(11) bonds assume distances of 1.256 Å and 1.293 Å, respectively, implying that the former is more "double bond"-like, whereas the latter is more elongated due to hydrogen bonding with the H(25) hydrogen atom of the $N(1)H_2$ group. It is also noteworthy that the O(10)–C(9)–O(11) angle in **1p** is 126.6° thus being enlarged by 3.6° relative to the corresponding angle in the parent arginine **1**.

The protonation at the N(1) atom leads to two NH_2 groups attached to C(2) atom in **1p**, which in turn exhibit very similar spatial and electronic structures. Tiny differences are caused by variation in weak interactions with non-nearest neighbours. A point of interest is their degree of pyramidalization. This can be conveniently estimated by a deviation of the three bond angles of NH_2 groups from the full angle normalized to 90°, shown in eqn. (1), where DP (in percent) stands here for the degree of

$$\mathrm{DP}\,(\%) = \left[360 - \sum_{i=1}^{3} \alpha_i\right]/0.9 \qquad (1)$$

pyramidalization and the summation $\sum_{i=1}^{3} \alpha_i$ is extended over bond angles a_i (in degrees) of a pyramidal nitrogen atom. If the summation is 360°, then the amino group is obviously planar. On the other hand, it is tacitly assumed here that the maximal pyramidalization is achieved if the sum of three angles is equal to 270°, corresponding to three mutually perpendicular $2p_N$ atomic orbitals. This supposition is plausible since it is impossible to form real hybrid AOs closing an angle less than 90°. It should be mentioned that it is in principle feasible to construct hybrid orbitals composed of 2s and 2p AOs possessing the interhybrid angle smaller than 90°, but they are then complex functions. The latter are, however, energetically unfavourable thus being unsuitable for formation of strong covalent bonds.[11] Employing eqn. (1) one finds that the pyramidalization DP values for N(1) and N(3) atoms in **1p** are very small, being 1.4% and 6.3%, respectively. It follows that the distribution of covalent bonds of the protonated N(1) atom is practically planar, whereas a slight nonplanarity of the $N(3)H_2$ group occurs presumably due to interaction with the H-bonded central region of the molecule. This is remarkable because the pyramidalization of the $N(3)H_2$ amino group in the parent compound **1** is as large as 33.8%. Apparently, the protonation causes both NH_2 groups to become planar, thus allowing for a strong delocalization effect. This is also obvious from the N(1)–C(2) and N(3)–C(2) bond distances, which are lengthened/shortened upon protonation by roughly 0.05/0.06 (in Å), respectively. Similarly, both NH_2 groups attached to the C(2) atom in the zig-zag protonated form **2p** (Fig. 2) are practically planar as evidenced by the pyramidalization values of 0.0% and 1.1% for N(1) and N(3) atoms respectively. Pyramidalization in the zwitterion **3** is of particular interest in view of the strong Coulomb interactions between the $N(3)H_2$ group and the carboxylic COO fragment. The degree of pyramidalization of this amino group should be larger than that of the $N(1)H_2$ one. This is indeed the case as evidenced by the DP values of 16.9% and 11.4%, respectively.

The electron density distribution

The charge distribution in arginine isomers **1** and **3** is of particular interest. We shall make use of Mulliken population analysis, which is known to be imperfect,[12] but it is employed here for qualitative purposes only. Perusal of the atomic charges given in Table 1 shows that the distributions of electron densities in **1** and **3** are not as widely different as intuitively expected. The conventional zwitterionic form has a somewhat more pronounced intramolecular charge transfer particularly in the molecular fragments which lose/gain the proton, but this feature does not provide a rationale for distinguishing isomers **1** and **3** as neutral and zwitterionic, respectively. Their charge distribution is similar in a broad sense. A conventional picture based on the negative unit charge localized on the COO moiety and the positive unit charge placed at the protonated imino nitrogen atom is obviously misleading. One of the reasons behind this remarkable finding is a well known fact that the protonated atom in cations recovers practically all of its previous electron density due to a strong charge relaxation effect.[13] This is also obvious from the atomic charges in the protonated species **1p** and **2p**. The protonated nitrogen atom gains electron density from all remaining atomic groupings and not exclusively from the carboxylic COO fragment. In fact, the protonated nitrogen N(1) becomes, in these particular cations, more negative than in the unprotonated parent arginine **1** as far as Mulliken charges are concerned. A more realistic picture is obtained by using Löwdin atomic charges,[11] but the main conclusions remain the same. The protonated atom regains the largest portion of its initial electron density. We note in passing that the intramolecular charge transfer is much less pronounced, if Löwdin partitioning of the mixed electron density is utilized (Table 1). This is in accordance with Pauling's electroneutrality principle and with the simple idea that atomic densities are perturbed as little as possible by formation of covalent bonds.

A one-electron property closely related to the charge distribution is the dipole moment. It measures a separation between the centers of gravity of positive and negative charges in the molecule. The dipole moments calculated by M(I), M(II) and M(IV) models are given in Table 1. It appears that dipole moments do not critically depend on the electron correlation in accordance with the general knowledge that one-electron properties are not highly sensitive to the correlation effects.

Table 1 Selected structural parameters of some arginine conformers (**1**–**3**) and the most stable protonated form (**1p**) as obtained by theoretical models M(I), M(II) and M(IV). Atomic charges offered by Mulliken and Löwdin analyses and dipole moments refer to the M(II) model[a]

Molecular species	Distance or angle	M(I)	M(II)	M(IV)	Atom	Charge	
						Mull.	Löw.
1	N(1)–C(2)	1.264	1.290	1.290	N(1)	−0.65	−0.49
	C(2)–N(3)	1.393	1.410	1.408	C(2)	0.56	0.13
	C(2)–N(4)	1.372	1.384	1.383	N(3)	−0.81	−0.54
	N(4)–C(5)	1.447	1.454	1.455	N(4)	−0.70	−0.36
	C(5)–C(6)	1.530	1.526	1.528	C(5)	−0.13	−0.20
	C(6)–C(7)	1.541	1.536	1.538	C(6)	−0.32	−0.29
	C(7)–C(8)	1.541	1.539	1.541	C(7)	−0.32	−0.29
	C(8)–C(9)	1.523	1.520	1.522	C(8)	−0.09	−0.10
	C(8)–N(12)	1.448	1.461	1.460	C(9)	0.55	0.18
	C(9)–O(10)	1.118	1.220	1.211	O(10)	−0.44	−0.26
	C(9)–O(11)	1.329	1.358	1.355	O(11)	−0.60	−0.38
	N(12) H(26)	2.100	1.960	1.953	N(12)	−0.78	−0.54
	N(1)–C(2)–N(3)	126.0	127.1	126.7	H(26)	0.41	0.30
	N(1)–C(2)–N(4)	122.1	121.5	121.6			
	C(2)–N(4)–H(26)	114.4	113.8	113.8			
	C(2)–N(4)–C(5)	118.8	117.0	117.3			
	N(4)–C(5)–C(6)	112.7	112.0	112.1			
	C(5)–C(6)–C(7)	118.4	117.8	117.6			
	C(6)–C(7)–C(8)	116.9	116.8	116.6			
	C(7)–C(8)–N(12)	112.1	117.7	111.5			
	C(7)–C(8)–C(9)	109.3	108.7	107.8			
	C(8)–C(9)–O(10)	124.9	125.0	125.0			
	C(8)–C(9)–O(11)	126.6	111.9	111.5			
	C(9)–O(11)–H	108.4	105.9	106.3			
	H–N(12)–H	105.8	104.9	105.4			
	H–N(3)–H	117.7	110.5	111.1			
	H–N(1)–C(2)	111.7	110.1	109.5			
	Dipole moment	3.7	3.7	3.6			
2	N(1)–C(2)	1.261	1.288	—	N(1)	−0.64	−0.47
	C(2)–N(3)	1.392	1.407	—	C(2)	0.56	0.12
	C(2)–N(4)	1.378	1.391	—	N(3)	−0.81	−0.54
	N(4)–C(5)	1.452	1.459	—	N(4)	−0.67	−0.34
	C(5)–C(6)	1.525	1.523	—	C(5)	−0.12	−0.20
	C(6)–C(7)	1.531	1.527	—	C(6)	−0.35	−0.31
	C(7)–C(8)	1.539	1.536	—	C(7)	−0.30	−0.29
	C(8)–C(9)	1.522	1.520	—	C(8)	−0.11	−0.11
	C(8)–N(12)	1.446	1.459	—	C(9)	0.55	0.18
	C(9)–O(10)	1.189	1.221	—	O(10)	−0.45	−0.26
	C(9)–O(11)	1.330	1.359	—	O(11)	−0.60	−0.38
	N(1)–C(2)–N(3)	126.7	126.7	—	N(12)	−0.75	−0.56
	N(1)–C(2)–N(4)	121.6	121.6	—			
	C(2)–N(4)–C(5)	119.1	119.2	—			
	N(4)–C(5)–C(6)	110.1	110.0	—			
	C(5)–C(6)–C(7)	111.5	111.6	—			
	C(6)–C(7)–C(8)	114.9	114.6	—			
	C(7)–C(8)–N(12)	111.5	111.4	—			
	C(7)–C(8)–C(9)	109.5	110.8	—			
	C(8)–C(9)–O(10)	114.1	125.2	—			
	C(8)–C(9)–O(11)	123.9	112.4	—			
	C(9)–O(11)–H	107.8	108.3	—			
	H–N(12)–H	106.5	105.8	—			
	H–N(3)–H	111.6	111.7	—			
	H–N(1)–C(2)	111.9	111.9	—			
	Dipole moment	2.1	2.2	—			
3	N(1)–C(2)	1.336	1.357	1.358	N(1)	−0.81	−0.47
	C(2)–N(3)	1.328	1.337	1.329	C(2)	0.77	0.22
	C(2)–N(4)	1.317	1.333	1.340	N(3)	−0.79	−0.46
	N(4)–C(5)	1.461	1.464	1.467	N(4)	−0.64	−0.26
	C(5)–C(6)	1.535	1.533	1.537	C(5)	−0.16	−0.21
	C(6)–C(7)	1.544	1.539	1.539	C(6)	−0.35	−0.30
	C(7)–C(8)	1.541	1.534	1.531	C(7)	−0.30	−0.30
	C(8)–C(9)	1.535	1.532	1.535	C(8)	−0.08	−0.11
	C(8)–N(12)	1.459	1.473	1.471	C(9)	0.52	0.12
	C(9)–O(10)	1.228	1.256	1.252	O(10)	−0.58	−0.41
	C(9)–O(11)	1.257	1.293	1.286	O(11)	−0.65	−0.48
	O(11)–H(26)	1.874	1.725	1.503	N(12)	−0.77	−0.57
	N(1)–C(2)–N(3)	120.1	120.7	121.9	H(25)	0.43	0.34
	N(1)–C(2)–N(4)	118.4	118.0	118.1			
	C(2)–N(4)–C(5)	126.5	124.9	124.0			
	N(4)–C(5)–C(6)	111.7	111.7	111.9			
	C(5)–C(6)–C(7)	117.2	116.8	116.6			
	C(6)–C(7)–C(8)	118.4	117.7	117.4			
	C(7)–C(8)–N(12)	107.0	106.7	107.2			
	C(7)–C(8)–C(9)	110.8	110.2	111.0			
	C(8)–C(9)–O(10)	116.7	116.6	117.4			
	C(8)–C(9)–O(11)	116.9	116.8	116.3			
	H–N(12)–H	105.7	104.6	105.8			
	H–N(1)–H	116.7	114.7	114.3			
	H–N(3)–H	115.7	116.9	118.3			
	Dipole moment	6.4	7.1	8.5			
4	N(1)–C(2)	1.379	1.392	—	N(1)	−0.82	−0.53
	C(2)–N(3)	1.385	1.397	—	C(2)	0.59	0.13
	C(2)–C(4)	1.261	1.288	—	N(3)	−0.81	−0.52
	N(4)–C(5)	1.446	1.457	—	N(4)	−0.48	−0.28
	C(5)–C(6)	1.527	1.525	—	C(5)	−0.19	−0.22
	C(6)–C(7)	1.546	1.542	—	C(6)	−0.32	−0.29
	C(7)–C(8)	1.546	1.544	—	C(7)	−0.32	−0.31
	C(8)–C(9)	1.515	1.511	—	C(8)	−0.07	−0.10
	C(8)–C(12)	1.448	1.458	—	C(9)	0.57	0.20
	C(9)–C(10)	1.189	1.220	—	O(10)	−0.45	−0.26
	C(9)–C(11)	1.330	1.360	—	O(11)	−0.61	−0.26
	N(12)–H(25)	1.002	1.023	—	N(12)	−0.81	−0.59
	N(1)–C(2)–N(3)	112.4	112.3	—			
	N(1)–C(2)–N(4)	127.1	127.4	—			
	C(2)–N(4)–C(5)	119.8	118.0	—			
	N(4)–C(5)–C(6)	110.9	109.9	—			
	C(5)–C(6)–C(7)	115.0	113.8	—			
	C(6)–C(7)–C(8)	113.7	112.8	—			
	C(7)–C(8)–N(12)	115.5	115.3	—			
	C(7)–C(8)–C(9)	108.2	107.8	—			
	C(8)–C(9)–O(10)	125.3	125.5	—			
	C(8)–C(9)–O(11)	112.5	107.9	—			
	H–N(12)–H	106.8	105.8	—			
	H–N(1)–H	112.8	112.1	—			
	H–N(3)–H	111.9	111.1	—			
	Dipole moment	4.91	5.0	—			

Molecular species	Distance or angle	M(I)	M(II)	M(IV)	Atom	Charge		Molecular species	Distance or angle	M(I)	M(II)	M(IV)	Atom	Charge	
						Mull.	Löw.							Mull.	Löw.
5	N(1)–C(2)	—	1.275	—	N(1)	−0.61	−0.43	**2p**	N(1)–C(2)	1.325	1.339	—	N(1)	−0.82	−0.44
	C(2)–N(3)	—	1.405	—	C(2)	0.53	0.11		C(2)–N(3)	1.328	1.349	—	C(2)	0.86	0.22
	C(2)–N(4)	—	1.438	—	N(3)	−0.79	−0.53		C(2)–N(4)	1.316	1.329	—	N(3)	−0.83	−0.45
	N(4)–C(5)	—	1.487	—	N(4)	−0.73	−0.37		N(4)–C(5)	1.472	1.473	—	N(4)	−0.67	−0.24
	C(5)–C(6)	—	1.533	—	C(5)	−0.20	−0.22		C(5)–C(6)	1.525	1.521	—	C(5)	−0.17	−0.20
	C(6)–C(7)	—	1.541	—	C(6)	−0.34	−0.31		C(6)–C(7)	1.531	1.526	—	C(6)	−0.37	−0.31
	C(7)–C(8)	—	1.520	—	C(7)	−0.32	−0.31		C(7)–C(8)	1.542	1.537	—	C(7)	−0.31	−0.29
	C(8)–C(9)	—	1.578	—	C(8)	−0.16	−0.14		C(8)–C(9)	1.522	1.518	—	C(8)	−0.10	−0.10
	C(8)–N(12)	—	1.499	—	C(9)	0.57	0.11		C(8)–N(12)	1.448	1.459	—	C(9)	0.57	0.19
	C(9)–O(10)	—	1.231	—	O(10)	−0.53	−0.34		C(9)–O(10)	1.189	1.218	—	O(10)	−0.42	−0.24
	C(9)–O(11)	—	1.281	—	O(11)	−0.61	−0.44		C(9)–O(11)	1.324	1.357	—	O(11)	−0.61	−0.38
	N(1)–C(2)–N(3)	—	120.3	—	N(12)	−0.84	−0.43		N(1)–C(2)–N(3)	119.1	119.0	—	N(12)	−0.79	−0.58
	N(1)–C(2)–N(4)	—	127.8	—					N(1)–C(2)–N(4)	120.2	120.5	—			
	C(2)–N(4)–C(5)	—	116.2	—					C(2)–N(4)–C(5)	125.8	125.4	—			
	N(4)–C(5)–C(6)	—	109.0	—					N(4)–C(5)–C(6)	109.3	109.2	—			
	C(5)–C(6)–C(7)	—	116.9	—					C(5)–C(6)–C(7)	111.5	111.9	—			
	C(6)–C(7)–C(8)	—	118.0	—					C(6)–C(7)–C(8)	111.5	111.9	—			
	C(7)–C(8)–N(12)	—	116.3	—					C(7)–C(8)–N(12)	110.8	109.8	—			
	C(7)–C(8)–C(9)	—	113.8	—					C(7)–C(8)–C(9)	109.6	111.0	—			
	C(8)–C(9)–O(10)	—	116.6	—					C(8)–C(9)–O(10)	124.2	125.2	—			
	C(8)–C(9)–O(11)	—	112.1	—					C(8)–C(9)–O(11)	112.5	111.4	—			
	H–N(3)–H	—	110.1	—					H–O(11)–C(9)	109.1	106.1	—			
	Dipole moment	—	11.3	—					H–N(12)–H	106.8	107.6	—			
1p	N(1)–C(2)	1.328	1.344	1.346	N(1)	−0.83	−0.46		H–N(1)–H	117.0	117.6	—			
	C(2)–N(3)	1.331	1.352	1.354	C(2)	0.83	0.21		H–N(3)–H	116.8	117.3	—			
	C(2)–N(4)	1.309	1.319	1.317	N(3)	−0.83	−0.45								
	N(4)–C(5)	1.463	1.463	1.463	N(4)	−0.69	−0.26								
	C(5)–C(6)	1.528	1.525	1.527	C(5)	−0.19	−0.21								
	C(6)–C(7)	1.545	1.541	1.543	C(6)	−0.35	−0.29								
	C(7)–C(8)	1.540	1.537	1.539	C(7)	−0.35	−0.30								
	C(8)–C(9)	1.524	1.523	1.526	C(8)	−0.09	−0.10								
	C(8)–N(12)	1.461	1.475	1.475	C(9)	0.57	0.18								
	C(9)–O(10)	1.187	1.220	1.211	O(10)	−0.42	−0.24								
	C(9)–O(11)	1.320	1.347	1.342	O(11)	−0.59	−0.36								
	N(12)–H(26)	1.911	1.754	1.725	N(12)	−0.85	−0.56								
	N(1)–C(2)–N(3)	118.7	117.8	117.7	H(23)	0.47	0.33								
	N(1)–C(2)–N(4)	119.7	120.1	120.3											
	C(2)–N(4)–C(5)	125.5	124.4	124.1											
	N(4)–C(5)–C(6)	110.0	110.2	109.8											
	C(5)–C(6)–C(7)	116.3	115.4	114.8											
	C(6)–C(7)–C(8)	114.7	114.6	114.3											
	C(7)–C(8)–N(12)	112.2	112.0	111.9											
	C(7)–C(8)–C(9)	109.4	109.0	108.6											
	C(8)–C(9)–O(10)	123.6	123.6	123.4											
	C(8)–C(9)–O(11)	112.3	111.5	111.4											
	C(5)–N(4)–H(23)	115.8	115.9	116.2											
	H–N(12)–H	105.0	104.5	104.7											
	H–N(1)–H	117.0	117.2	116.6											
	H–N(3)–C(2)	117.4	116.1	115.8											

[a] Interatomic distances and bond angles are given in Å and degrees, respectively. Dipole moments are in Debye.

Table 2 Total molecular electronic energies (in au), zero point vibration energies and proton affinities (in kcal mol^{-1})[a]

Molecular species	M(I)	M(II)	M(III)	M(IV)	M(V)	ZPVE
1	−602.89238	−604.66878	−605.06690	−605.06738	−605.36820	134.8
2	−602.89212	−604.66507	−605.06421	—	—	134.5
3	−602.88251	−604.66852	−605.06189	−605.06327	−605.36757	135.1
4	−602.88824	−604.66633	−605.06446	—	—	134.5
5	—	−604.64461	−605.03706	—	—	—
1p	−603.31664	−605.08581	−605.47581	−605.47632	−605.77661	142.1
2p	−603.30327	−605.06596	−605.45895	—	—	141.5
PA(**1p**)	258.8	254.4	249.3	249.3	249.0	—
PA(**2p**)	251.1	242.5	239.3	—	—	—
Δ_{13}	−6.2	−0.2	−3.1	−2.6	−0.4	−0.3
Δ_{13} (t)	−6.5	−0.5	−3.4	−2.9	−0.7	—

[a] The difference in the total electronic (or ZPVE) energies between **1** and **3** is denoted by Δ_{13}. The corresponding entity involving a sum of Δ_{13}(el) + Δ_{13}(ZPVE) is given by Δ_{13} (t).

Hence, the Hartree–Fock values are close to dipole moments obtained by the M(II) model for systems **1** and **2**. The dipole moment is not large in **1**, which is dramatically changed in the zwitterionic configuration **2** exhibiting a dipole moment of 7.1 D. It is therefore plausible to assume that the zwitterion will be more stabilized by polar solvent molecules than the neutral tautomer **1** and that this form will predominantly occur in solutions. It is noteworthy that in less conventional systems like zwitterion **3** both the electron correlation and flexibility of the employed basis set are important in estimating the dipole moment. This is illustrated by dipole moment estimates obtained by models M(I), M(II) and M(IV) which are 6.4, 7.1 and 8.5 (in D), respectively.

Energetic properties

The total electronic and ZPV energies are presented in Table 2. We shall first focus on the H-bond interaction in **1**. It appears that the hydrogen bonding contributes about 2.0 or 1.4 kcal mol^{-1} to the stability of the neutral arginine depending on the applied models M(II) or M(III), respectively. The difference in electronic energies between neutral **1** and zwitterionic **3** forms is given by Δ_{13}(el). It appears that all five theoretical models predict that the neutral configuration is more stable. However, the Hartree–Fock model exaggerates the difference in stability of the tautomer **1**, compared to all other MP2 models. Apparently, the electron correlation plays a very important role in determining energetics and relative stability of these systems. This conclusion is in accordance with our earlier finding that the MP2 level of theory is a *conditio sine qua non* for a satisfactory description of the proton affinity. The HF values are useful only if they are properly scaled.[9] It should be mentioned that the ZPV energy gives only 0.3 kcal mol^{-1} to the enhanced stability of **1**. Taking into account both electronic and ZPVE contributions Δ_{13}(t) = Δ_{13}(el) + Δ_{13}(ZPVE) one obtains the difference in the total energy contents of **1** and **3**. They are −6.5, −0.5, −3.4, −2.9 and −0.7 (in kcal mol^{-1}) for models M(I)–M(V), respectively. Comparison of numbers offered by models M(II), M(III) and M(IV) gives an insight into the dependence of the relative stability on the quality of the basis sets, which in turn increase in flexibility ranging from 6-31G* to 6-311+G(2df,p) *via* the intermediate 6-311+G** basis functions. It is surprising that the 6-31G* and the 6-311+G(2df,p) basis sets yield practically the same result, which is probably fortuitous because the proton affinities (PA) for **1p** obtained by these functions differ by 5 kcal mol^{-1} (*vide infra*). Additionally, comparison of Δ_{13}(t) values for M(III) and M(IV) models shows that differences in geometries estimated by MP2(fc)/6-31G* and MP2(fc)/6-311+G** models have little influence on the relative stability of arginine tautomers. It follows as a corollary that the neutral distribution of the charges in **R** is energetically slightly more favourable by approximately 1–3 kcal mol^{-1}. It is important to stress that the tautomer **4** is by only ≃2 kcal mol^{-1} less stable than **1** (Table 2) and that its ZPVE is just 0.3 kcal mol^{-1} lower that the corresponding value for **1**. On the other hand, the zwitterionic form **5** is, by 15 kcal mol^{-1}, higher in energy than **3** thus being probably less important in the gas phase chemistry. The present results show also why Price *et al.*[5] come to the wrong conclusion that the zwitterion is a more stable configuration in the gas phase: their theoretical (BLYP) search of the potential energy surface identified a false absolute minimum, which corresponds to a particular conformation of the less stable neutral tautomer **4**! Finally it should be strongly pointed out that difference Δ_{13}(t) is surprisingly small in view of the considerable variation in the spatial and electronic structure of these two isomeric forms (Fig. 1).

A property of considerable interest is the absolute proton affinity of **R**. The intrinsic or gas phase proton affinity gives useful information about a number of properties, such as the acid–base behaviour of organic molecules, the intermolecular hydrogen bond ability and the enzymatic activity[14–16] *etc.* We found by extensive calculations that the MP2 model employing the 6-311+G** basis set offers reliable proton affinities for nitrogens in molecules involving atoms of the first row elements.[7–9] It appears that the PA values of arginine calculated by models M(III) and M(IV) are virtually the same, being 249.3 kcal mol^{-1} for the protonation of the neutral tautomer **1**. The PA obtained by the M(V) model is only 0.3 kcal mol^{-1} lower, indicating that the basis set saturation is already achieved at the 6-311+G** level of flexibility. The proton affinity of the zwitterionic form **3** is slightly higher assuming values of 252.4 and 251.9 kcal mol^{-1} for M(III) and M(IV) models, respectively. It would be useful to compare these results with some available experimental data. There are two types of measurements which were performed on **R**: one involving the thermodynamic equilibrium[17] approach and the other based on the kinetic data related to the proton bound collision dimers.[18] Unfortunately, both techniques give only the lower limit of the proton affinity of **R** indicating that the PA value should be higher than 243 kcal mol^{-1}. All theoretical results obtained here show that the PA should indeed be larger than that threshold, the most reliable value being 249 kcal mol^{-1}. It is of some interest to answer the question of why **R** protonates at the imino nitrogen belonging to the α-AA chain and to provide at the same time some rationalization of the high basicity of this important compound. Our extensive studies of the proton affinity of nitrogen containing compounds unambiguously show that the imino group is the most basic one.[7–9,19] Its susceptibility toward protonation is particularly enhanced, if it is flanked by two NH_2 groups like in the guanidine moiety, because in this case the resonance effect is very strong in the resulting conjugate acid. Additional alkyl substituent(s) increase the proton affinity in view of the amplified charge reorganization effect spurred by the protonation. It is known that the proton affinity of guanidine is 234 kcal mol^{-1} [8] implying that the backbone of arginine starting with the C(5) atom contributes 15 kcal mol^{-1} to the total value of 249 kcal mol^{-1}. It is interesting to note that a large portion of this increase in the PA originates from the intramolecular H-bonding in **1p**. Comparing the PA values of **1p** and the zig-zag protonated form **2p** one finds that the former puckered form is more stable by 10 kcal mol^{-1} (Table 2). Since the hydrogen bond strength in the parent arginine **1** is 1.4 kcal mol^{-1}, it follows that protonation amplifies the H-bonding by 8.6 kcal mol^{-1}. Consequently, the remaining 6.6 kcal mol^{-1} should be ascribed to the enhanced electron density relaxation effect of **1p** relative to that in the protonated guanidine. This is a reasonable estimate as evidenced by the alkyl effect in other related systems.[8,9]

As a final comment we note that the protonated form of tautomer **4** results in the same final structure **1p** implying that its PA value should be, by 2 kcal mol^{-1}, higher than that of tautomer **1**.

Concluding remarks

The present *ab initio* analysis shows that **R** should exist in the gas phase predominantly as a neutral conformer since it is energetically slightly more stable. The difference in the total energy between the energetically most favourable neutral and zwitterionic forms is rather small, lying within the range of 1–3 kcal mol^{-1} depending on the applied theoretical model. This is surprising in view of the dramatic differences in the structure of these two isomeric forms. The small difference in stability and much higher dipole moment of **3** shows that this zwitterion can be easily stabilized by just a few polar molecules of the solvent. This conjecture is confirmed by a computational study carried out on glycine (**G**) indicating that only two water molecules can stabilize the zwitterion of **G** as shown by Jensen and Gordon.[4] Our results indicate that two experimental studies, claiming that either the zwitterion or the neutral arginine species have been observed under specific conditions defined by measurements, might both be correct. Price *et al.*[5] started with the proton bound arginine dimer complex, which can subsequently easily dissociate into $\mathbf{RH}^+$ and **R** (zwitterion) by utilizing two reaction channels. The latter is probably capable of existing for some time, since the neutral form is only slightly more stable. Saykally *et al.*[6] started their spectroscopic investigations with jet-cooled arginine, which was in the neutral configuration from the outset. Consequently, arginine apparently stayed in its more stable neutral form in a supersonic molecular beam as was unambiguously established by analysis of the carbonyl stretch bands.[6] The very small difference in stability between the neutral and zwitterionic configurations possessing lowest energies indicates that arginine might be another case of a molecular chameleon. This conjecture is strengthened by a remarkable finding that there is another neutral tautomer **4**, which is less stable than **1** by roughly 2 kcal mol^{-1} meaning that its total energy is equal to that of the zwitterion **3**. Similarly, there is an additional zwitterionic isomer **5**, which, on the other hand, is substantially less stable than **3** by some 15 kcal mol^{-1}.

The estimated proton affinity of **1** is 249 kcal mol^{-1}, which shows that arginine is a very strong base despite the fact that it is nominally an amino acid. Its high basicity can be resolved in three ways: (1) the PA of the guanidine moiety has a value of 234 kcal mol^{-1}. (2) The intramolecular hydrogen bonding contributes some 9 kcal mol^{-1} and (3) the charge relaxation effect triggered by protonation provides ~6 kcal mol^{-1}. It is worth noting that the PA value of **4** should be close to 250.5 kcal mol^{-1} *i.e.* 2 kcal mol^{-1} higher than that of **1**.

Finally, it should be kept in mind that the classical picture of a molecular zwitterion as a bipolar system possessing highly localized ±1 charges placed on separated atoms or atomic groupings should be taken *cum grano salis*. In reality the electron density is more evenly distributed over the molecular framework.

Note added in proof: We found that the recently published experiment result for the proton affinity of arginine[20] of 251.2 kcal mol^{-1} was in good accordance with our PA values for both neutral isomers **1** and **4**. In fact, it is very close to the theoretical estimate for the latter isomer.

Acknowledgements

We thank Professor L. Klasinc for useful discussions and the Center for Advanced Computations of the University of Zagreb for a generous donation of computing time.

References

1 R. D. Suenram and F. J. Lovas, *J. Mol. Spectrosc.*, 1978, **72**, 372.
2 L. Schäfer, H. L. Selers, F. J. Lovas and R. D. Suenram, *J. Am. Chem. Soc.*, 1980, **102**, 6556.
3 M. J. Locke and R. T. McIver, Jr., *J. Am. Chem. Soc.*, 1983, **105**, 4226.
4 J. H. Jensen and M. S. Gordon, *J. Am. Chem. Soc.*, 1995, **117**, 8159.
5 W. D. Price, R. A. Jockusch and E. R. Williams, *J. Am. Chem. Soc.*, 1997, **119**, 11988.
6 C. J. Chapo, J. P. Paul, R. A. Provencal, K. Roth and R. J. Saykally, *J. Am. Chem. Soc.*, 1998, **120**, 12956.
7 C. Hillebrand, M. Klessinger, M. Eckert-Maksić and Z. B. Maksić, *J. Phys. Chem.*, 1996, **100**, 9698.
8 B. Kovačević and Z. B. Maksić, *Chem. Phys. Lett.*, 1998, **288**, 289; B. Kovačević, Z. B. Maksić and P. Rademacher, *Chem. Phys. Lett.*, 1998, **293**, 245.
9 Z. B. Maksić and B. Kovačević, *J. Phys. Chem. A*, 1998, **102**, 7324.
10 M. J. Frisch, G. W. Trucks, H. B. Schlegel, P. M. W. Gill, B. G. Johnson, M. A. Robb, J. R. Cheeseman, T. Keith, G. A. Petersson, J. A. Montgomery, K. Raghvachari, M. A. Al-Laham, V. G. Zakrzewski, J. V. Ortiz, J. B. Foresman, J. Cioslowski, B. B. Stefanov, A. Nanayakara, M. Challacombe, C. Y. Peng, P. Y. Ayala, W. Chen, M. W. Wong, J. L. Andres, E. S. Replogle, R. Gomperts, R. L. Martin, D. J. Fox, J. S. Binkley, D. J. Defrees, J. Baker, J. P. Stewart, M. Head-Gordon, C. Gonzales, J. A. Pople, GAUSSIAN 94, Revision D.1, Gaussian Inc. Pittsburg, PA, 1995.
11 Z. B. Maksić, in *Theoretical Models of Chemical Bonding*, Vol. 2, Z. B. Maksić, Ed., Springer Verlag, Berlin, Heidelberg, 1990, p. 137 and references cited therein.
12 For a critical survey of the existing electron density partitioning techniques the reader should consult *e.g.*: K. Jug and Z. B. Maksić, in *Theoretical Models of Chemical Bonding*, Vol. 3, Z. B. Maksić, Ed., Springer Verlag, Berlin, Heidelberg, 1991, p. 235.
13 Z. B. Maksić and M. Eckert-Maksić, in *Theoretical Organic Chemistry*, C. Párkányi, Ed., Elsevier, Amsterdam, 1998, p. 203.
14 A. J. Russel, P. G. Thomas and A. R. Fersht, *J. Mol. Biol.*, 1987, **100**, 803.
15 M. Meot-Ner, *J. Am. Chem. Soc.*, 1984, **106**, 278.
16 M. J. S. Dewar and D. M. Storch, *Proc. Natl. Acad. Sci. USA*, 1985, **82**, 2225.
17 G. S. Gorman, J. P. Speir, C. A. Turner and I. J. Amster, *J. Am. Chem. Soc.*, 1992, **114**, 3986.
18 G. Bojesen and T. Breindahl, *J. Chem. Soc., Perkin Trans. 2*, 1994, 1029.
19 Z. B. Maksić and B. Kovačević, *J. Phys. Chem. A*, 1999, **103**, 6678.
20 E. P. L. Hunter and S. G. Lias, *J. Phys. Chem. Ref. Data*, 1998, **27**, 413.

Paper 9/02404B