Electronic Structure and Bonding of the Amino Acids Containing First Row Atoms*

(Received for publication, March 25, 1976)

DAVID A. DIXON+ AND WILLIAM N. LIPSCOMB
From the Gibbs Chemical Laboratory, Harvard University, Cambridge, Massachusetts 02138

The electronic structures of the amino acids containing first row atoms have been determined for the zwitterionic form using an approximate self-consistent field method, partial retention of diatomic differential overlap. Various energetic quantities including certain proton affinities are presented as are eigenvalues for the highest occupied and lowest unoccupied molecular orbitals. It is found that our method, in common with all methods employing minimum basis sets, yields eigenvalues for the highest occupied molecular orbital that are too high. The method does predict the location of this orbital correctly when compared to calculations employing larger basis sets. It is predicted that electron loss due to ionizing radiation should occur from the carboxylate group for the nonaromatic amino acids, while for tyrosine and tryptophan, electron loss should occur from the ring system. No choice between these two sites can be made for phenylalanine. Charge distributions have been obtained which show that only partial zwitterionic character is found in the backbone and that little delocalization of charge from the backbone to the side chain occurs. Localized molecular orbitals have been obtained using the Boys criteria and the bonding in the amino acids is discussed in terms of these orbitals. Hybridization of various bonds and bond polarities are discussed as is the phenomenon of fractional bonding to carbon.

The amino acids, even though they are of great biological importance, have not yet been studied by ab initio or near ab initio methods. With the advent of the method of partial retention of diatomic differential overlap (PRDDO) (1, 2), wave functions of ab initio minimum basis set quality can be obtained for large asymmetric molecules in reasonable computer times (3, 4). Recent advances in the accurate determination of the structural parameters of the amino acids have been made using the technique of neutron diffraction primarily by the group at Brookhaven (5–7). This high quality crystal structure work provides accurate geometrical parameters which we have used in an investigation of the electronic structure and properties of the amino acids containing first row atoms as listed in Table I. We discuss energetics, eigenvalues for the highest occupied molecular orbital and lowest unoccupied molecular orbital, charge distributions, and localized molecular orbitals obtained with the Boys criteria (8–11).

Geometries and Calculations—Geometries for the amino acids were obtained from the crystal structures determined primarily by neutron diffraction (12–27) and in two cases by x-ray diffraction (28–31). These amino acids form zwitterions in the crystal and the wave functions have been calculated for the molecules in the zwitterion form. For the molecules, the structures of which have been determined by neutron diffraction methods, the crystal geometries were used directly, including the experimental hydrogen positions, as the technique of neutron diffraction does not shorten A–H bond lengths in contrast to the technique of x-ray diffraction which does (32). The geometries of leucine and isoleucine were taken from the x-ray diffraction studies (28–31) and hydrogens were positioned assuming tetrahedral coordination and appropriate C–H (1.08 Å) and N–H (1.05 Å) bond lengths. The geometries of proline and phenylalanine (33) were generated by removal of a hydroxyl group from 4-hydroxyproline and tyrosine, respectively, and replacing this group with a hydrogen atom at the appropriate C–H distance along the original C–O vector. The coordinates for aspartic acid were generated from the coordinates of asparagine by substituting a hydroxyl group for an amine (NH₃⁺) group. The oxygen atom was substituted at the nitrogen atom position and the hydrogen atom was placed on an N–H vector at the correct O–H (1.01 Å) bond distance. The geometries of the negative ions of glutamic acid and aspartic acid were generated from the geometries of the neutral molecules by removal of the proton from the carboxylic acid group and then averaging the two C=O distances to give two equal CO bonds. The geometry of the positive ion of arginine was generated by replacing the amine (NH₃⁺) group on the α carbon with a tetrahedral protonated amine (NH₄⁺) group. The two possible geometries for neutral lysine were generated by replacing the respective protonated

* This research was supported by National Institutes of Health Grant GM08920.
† Junior Fellow, Society of Fellows, Harvard University.
+ The crystal coordinates for tryptophan were kindly provided by T. Koetzle from unpublished neutron diffraction work.

After our study was completed, the structure of phenylalanine has been determined by neutron diffraction as given in Ref. 33.
nneglect of differential overlap (34) and are comparable to those used for carbon, nitrogen, and oxygen (1s,2s,2p basis), and the moments, and eigenvalues. Standard Slater exponents were obtained with the minimum basis set STO-3G (35) for propper-

msions to the 2-electron matrix are parameterized to reproduce experimental parameters are employed but certain contribu-
tions to the molecular orbital eigenvalues are on the indole ring.  h These negative ions have three positive eigenvalues and thus the ionization potential can not be calculated.

Computing times for an IBM 360/91 for convergence in the density matrix less than 0.0001.

° ND = neutron diffraction.

amine groups in the lysine-positive ion structure by amine (NH₃) groups. The geometrical results are summarized in Table I.

The method of partial retention of diatomic differential overlap uses a Slater minimum basis set, and all electrons are explicitly included in the self-consistent field calculation. No experimental parameters are employed but certain contributions to the 2-electron matrix are parameterized to reproduce ab initio matrix elements for a variety of small molecules; thus, our method is a nonempirical self-consistent field method. The results from the method of partial retention of differential diatomic overlap are greatly superior to those obtained from the methods of complete (CNO) or intermediate (INDO) neglect of differential overlap (34) and are comparable to those obtained with the minimum basis set STO-3G (35) for properties such as energy differences, charge distributions, dipole moments, and eigenvalues. Standard Slater exponents were used for carbon, nitrogen, and oxygen (1s,2s,2p basis), and the 1s exponent for hydrogen was set at 1.2. Computing times for our method are given in Table I for an IBM 360/91 computer for convergence in the density matrix less than 0.0001. These numbers are given in order to facilitate comparison of the method of partial retention of diatomic differential overlap with other methods.

Energies and Eigenvalues—Total energies and virial ratios are given in Table I. The proton affinity of lysine in the zwitterion form is 260 kcal/mol where the zwitterion with the NH₄ terminus protonated is more stable by 64 kcal/mol than the zwitterion with the side chain protonated. The proton affinity for the zwitterion of arginine is 314 kcal/mol when the side chain is protonated. The proton affinities of the negative ions of glutamic acid and aspartic acid were calculated and are 481 kcal/mol. The calculations also show that isoleucine is 9 kcal/mol more stable than leucine. We note that these results are for gas phase zwitterions and that the geometries are unoptimized; the trends, however, should be qualitatively correct. Thus, the proton affinities for the negative ions of glutamic acid and aspartic acid should be similar,

### Table I

<table>
<thead>
<tr>
<th>No.</th>
<th>Molecule</th>
<th>Structure</th>
<th>Energy</th>
<th>Viral ratio</th>
<th>HOMO</th>
<th>LUMO</th>
<th>LUMO location</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Alanine</td>
<td>ND (17)</td>
<td>-1230.726</td>
<td>0.939</td>
<td>-0.110</td>
<td>0.379</td>
<td>-NH₄⁺</td>
<td>16</td>
</tr>
<tr>
<td>2</td>
<td>Arginine</td>
<td>ND (10)</td>
<td>-601.041</td>
<td>0.955</td>
<td>-0.017</td>
<td>0.135</td>
<td>N=C=O</td>
<td>85</td>
</tr>
<tr>
<td>3</td>
<td>Arginine positive ion</td>
<td>From 2'</td>
<td>-601.541</td>
<td>0.977</td>
<td>-0.196</td>
<td>0.046</td>
<td>N=C=O</td>
<td>85</td>
</tr>
<tr>
<td>4</td>
<td>Asparagine</td>
<td>ND (11)</td>
<td>-487.053</td>
<td>0.992</td>
<td>-0.107</td>
<td>0.299</td>
<td>&gt;C=O</td>
<td>44</td>
</tr>
<tr>
<td>5</td>
<td>Aspartic acid</td>
<td>From 4</td>
<td>-307.644</td>
<td>0.992</td>
<td>-0.113</td>
<td>0.269</td>
<td>&gt;C=O</td>
<td>30</td>
</tr>
<tr>
<td>6</td>
<td>Aspartic acid negative ion</td>
<td>From 5</td>
<td>-408.304</td>
<td>0.987</td>
<td>+0.0655</td>
<td>0.531</td>
<td>-NH₄⁺</td>
<td>30</td>
</tr>
<tr>
<td>7</td>
<td>Glutamic acid</td>
<td>ND (13)</td>
<td>-546.624</td>
<td>0.992</td>
<td>-0.107</td>
<td>0.268</td>
<td>&gt;C=O</td>
<td>44</td>
</tr>
<tr>
<td>8</td>
<td>Glutamic acid negative ion</td>
<td>From 7</td>
<td>-545.889</td>
<td>0.988</td>
<td>+0.059</td>
<td>0.499</td>
<td>NH₄⁺</td>
<td>44</td>
</tr>
</tbody>
</table>

* Virial ratio is \(-E/T\).*

* Conversion to eV is 1 au - 27.2 eV; negative values correspond to Koopmans' theorem ionization potentials. HOMO is the highest occupied molecular orbital and LUMO is the lowest unoccupied molecular orbital.

* Computing times for an IBM 360/91 for convergence in the density matrix less than 0.0001.

* ND = neutron diffraction.

* Numbers in parentheses refer to reference for the structure.

* From number; structure taken from molecule of that number as described in the text.

* These negative ions have three positive eigenvalues and thus the ionization potential cannot be calculated.

* The Ex⁺ phenyl eigenvalues split slightly. The two lowest unoccupied molecular orbital eigenvalues are on the indole ring.

### Table I

Energies and Eigenvalues—Total energies and virial ratios are given in Table I. The proton affinity of lysine in the zwitterion form is 260 kcal/mol where the zwitterion with the NH₄ terminus protonated is more stable by 64 kcal/mol than the zwitterion with the side chain protonated. The proton affinity for the zwitterion of arginine is 314 kcal/mol when the side chain is protonated. The proton affinities of the negative ions of glutamic acid and aspartic acid were calculated and are 481 kcal/mol. The calculations also show that isoleucine is 9 kcal/mol more stable than leucine. We note that these results are for gas phase zwitterions and that the geometries are unoptimized; the trends, however, should be qualitatively correct. Thus, the proton affinities for the negative ions of glutamic acid and aspartic acid should be similar,
but the actual magnitude is too large due to deficiencies of the minimum basis set in dealing with negative ions (36). Even though these are molecules of reasonable size, the minimum basis set is not quite adequate for these negative ions as evidenced by the three positive eigenvalues. However, the proton affinities for the neutral molecules should be more accurate in magnitude as evidenced by other studies (36). For instance, the experimental proton affinity of ammonia is 207 kcal/mol, while our calculated value is 235 kcal/mol.

The highest occupied molecular orbital is found to reside on the two oxygens of a carboxylate ion group for all 24 of the molecules studied. In the negative ions of glutamic acid and aspartic acid, oxygens from both carboxylate ion groups participate in the highest occupied molecular orbital. The average eigenvalue for this orbital for 18 neutral amino acids is -0.111 au, with a spread of ±0.01 au. This eigenvalue corresponds to an ionization potential of roughly 3 eV.

The only neutral amino acids that differ from this are arginine and the form of lysine with the protonated side chain where the carboxylate ion group for each molecule is destabilized because of the distance from the positive center and thus have eigenvalues for the highest occupied molecular orbital near zero. The 2 positive ions have more stabilized carboxylate ion groups due to the excess positive charge and thus have a more negative eigenvalue. As discussed previously the negative ions of glutamic acid and aspartic acid each have three occupied orbitals with positive eigenvalues. The eigenvalues for the occupied π orbitals in histidine are located on C^1-N^1-C^2 (-0.350 au) and on N^1 and N^2 (-0.406 au).

The eigenvalue (Table I) and location for the lowest unoccupied molecular orbital being lower in the acid form than in the amine form. In histidine the lowest unoccupied molecular orbital (0.212 au) is in the C^1-N^1-C^2 group, while the next virtual orbital (0.274 au) is located on N^3 and C^2.

Since a large negative charge is localized on the carboxylate ion group, the minimum basis set may make the magnitude of the eigenvalue too small and, consequently, the ionization potential would be too low. In order to ascertain whether our results for the eigenvalues of the highest occupied and lowest unoccupied molecular orbitals are due to deficiencies in the minimum basis set description of zwitterions, the wave function for glycine was calculated at the STO-4-31G level (36). This level corresponds to an ab initio self-consistent field calculation where the valence shell is treated at the double zeta level and a better description of ionic molecules is given. However, it is very time consuming to perform these STO-4-31G calculations as seen by comparing the times for the calculation in the method of partial retention of diatomic differential overlap (9 s) and for the STO-4-31G calculation (15 min and 10 s) on an IBM 360/91 computer. The eigenvalues for the highest occupied and lowest unoccupied molecular orbitals are -0.326 au and 0.106 au, respectively, at the STO-4-31G level, while the corresponding eigenvalues for our calculations using the method of Halgren and Lippecomb (1, 2) are 0.106 au and 0.374 au. Although the difference between eigenvalues of the highest occupied and lowest unoccupied molecular orbitals is similar for both calculations, the minimum basis set (our work) seriously underestimates the size of the ionization potential, which is 8.87 eV at the STO-4-31G level and only 2.88 eV at the minimum basis set level. We do note that the method of partial retention of diatomic differential overlap does correctly predict the location of the highest occupied molecular orbital on the oxygens of the carboxylate ion group and the location of the lowest unoccupied molecular orbital on the protonated amine group. Thus, ionization of the nonaromatic amino acids should occur primarily by electron loss from the carboxylate ion group to form the radical ion NH^+_2CHRCOO with an ionization potential of roughly 9 eV.

We now discuss the results for the amino acids phenylalanine, tyrosine, and tryptophan which have aromatic side chains. The eigenvalues for the occupied π orbitals which correspond to ionization potentials from Koopman's theorem (37) are given in Table II along with the results for benzene (38, 39). We present these results separately due to the recent experimental studies involving the laser flash photolysis (40, 41) and fluorescence (42, 43) of these molecules. The results found for phenylalanine are remarkably similar to those found for benzene. The ionization potentials are nearly identical and only a slight splitting of the E_σ orbitals is observed toward a higher ionization potential. Thus, the backbone portion of the
The E, orbitals found in p-cresol is retained in tyrosine. The potential found in p-cresol by 0.23 eV, but the large splitting of the E, orbitals found in p-cresol is retained in tyrosine. The mixing of lone pair orbitals from the oxygen with the ring orbitals. The second ring ionization potential is similar to that found for benzene in both p-cresol and tyrosine. The first π ionization potential for tryptophan is larger than that found in indole (geometry obtained from tryptophan), just as the first π ionization potential of tyrosine was greater than the ionization potential found for p-cresol. The remaining π ionization potentials in tryptophan are all greater by ~0.3 eV than the comparable ionization potentials found for indole. The STO-3G results for glycine show that ionization from the carboxylate ion group occurs near 9 eV. Thus, in phenylalanine, it is difficult to know whether ionization occurs by electron loss from the carboxylate ion group or the aromatic ring. However, for tyrosine and tryptophan, ionization should occur by electron loss from the aromatic rings since these ionization potentials are significantly lower than 9 eV.

The lowest unoccupied molecular orbitals for these 3 molecules are localized on the aromatic rings in the π* states rather than at the center of positive charge. The results for phenylalanine and tyrosine are quite similar, showing a slight splitting of the E, orbitals at roughly 0.20 au. In tryptophan the lowest unoccupied molecular orbital falls to a lower value (0.166 au) and involves only carbon atoms, while the second virtual orbital occurs at 0.216 au and with some “density” on N'.

Charge Distributions—In Fig. 1 (44), we show the group charges and the qualitative localized molecular orbitals obtained by the Boys criteria (8-11). The group charges are calculated by adding the atomic population of the hydrogens connected to a heavy atom to the population of that atom. This procedure, used with success in the boron hydrides (3, 4), alleviates some of the deficiencies of the Mulliken (45) division of charge. A result is a simpler comparison of various portions of the molecules.

In the carboxylate ion groups, the average charge on carbon is +0.19 e; only one variation is greater than 0.02 e. The oxygen atoms in the carboxylate ion groups exhibit an average charge of -0.42 e, but both oxygen atoms rarely have the same charge; the usual difference between the two charges is 0.03 to 0.04 e. The variance in the charges of the oxygen atoms about the average is at most 0.04 e. Thus, the two oxygens in the carboxylate ion group carry a total of approximately -0.84 e in reasonable agreement with the simple zwitterionic model which places unit negative charge on these 2 atoms. Inclusion of the charge on the carbon atom gives an effective charge of -0.72 e on the carboxylate ion group.

The average group charge for the protonated amine group is 0.61 e with a maximum variation of 0.02 e except for the negative ion of aspartic acid where the variation is 0.04 e. The positive charge of 0.61 e on the protonated amine group is somewhat less than that expected from the simple zwitterion model for this region of the molecule. In proline and 4-hydroxy-proline, the protonated amine group has a lower charge of 0.40 e as expected, since one hydrogen is missing. Comparison of the charge for the protonated amine group with the carboxylate group charge shows that the two charges almost balance; the carboxylate group has 0.11 e greater absolute charge. Most of the positive charge needed for electroneutrality is concentrated on C' which has an average charge of 0.08 e as long as it is adjacent to both the protonated amine and carboxylate ion groups. On the average, very little of the zwitterion charge (-0.03 e) delocalizes to the side chains. Moreover, the backbone portion of the amino acids shows a regular charge distribution. Thus, in modeling the interactions of amino acids in polypeptides using, in part, charge distributions, we see that separability of the backbone and side chain charge distributions is not unreasonable.

We now discuss the charge distributions for the side chains, noting that our values are the most accurate ones yet presented in the literature. The important side chain charges are summarized in Table III. As found above for the backbone regions, the negative charges tend to be more strongly localized than do the positive charges. Those molecules which can become positively charged tend to show an increase in the side chain charge in going from the neutral to the positive ion with the charge on the NH4+ group in lysine changing from 0.66 e to 0.68 e and the charge on the C5N4 ion group changing from 0.72 e to 0.78 e.

The hydroxyl groups are found to be slightly negative, ranging from -0.11 e in threonine to -0.03 e in tyrosine. The oxygen atoms in carbonyl groups are more negative, having charges between -0.17 e and -0.24 e. In general the positive charge on the carbon atom in the carboxyl group tends to balance the negative oxygen atom charge. The amine groups found in asparagine and glutamine are slightly negative (-0.01 e and -0.02 e), while the amine group in lysine is more negative (-0.07 e). Aliphatic carbons not adjacent to heteroatoms have small charges ranging from -0.05 e to methyl groups to 0.05 e for more highly substituted carbons. As expected, the atom C4 adjacent to the NH4+ group in proline and 4-hydroxyproline is quite positive with more charge than found on C'.

A comparison of the ring charges between tyrosine and phenylalanine shows that the OH group strongly polarizes the ring charge. Some effect of the backbone in the ring is indicated by the observations that the ring carbons on the same side as the protonated amine group are somewhat negative, and that the carbons on the same side as the carboxylate group are slightly positive. In histidine, N6 is quite negative at -0.24 e, while the NH group is slightly positive (0.05 e); the C8H group between N8 and N4 is quite positive (0.13 e). The benzene ring in tryptophan is slightly negative as is the N3-N5 group. However, C9 and C10, which are adjacent to N3, are positive with charges of 0.14 e and 0.10 e, respectively. The charge on C10 is lower showing the moderating effect of the benzene ring.

Localized Molecular Orbitals—In Fig. 1, we show the localized molecular orbitals (LMO's) obtained from the Boys criteria (8-11), while data for certain types of bonds are given in Table IV. The localized molecular orbitals for boranes and boron hydrides have been discussed extensively (46, 47) but few localized orbitals derived from wave functions have been published for compounds containing predominantly carbon atoms. Kleier et al. (48) have discussed the localized molecular orbitals from wave functions generated from the method of partial retention of diatomic differential overlap for...
Fig. 1. Group charge distributions and localized molecular orbital structures for the amino acids studied. The atomic labeling follows the rules given in Ref. 44.
formamide, aniline, nitrobenzene, guanine, and cytosine, while Kleier et al. (38) have discussed the localized molecular orbitals for the monocyclic aromatics in detail.

The carboxylate ion groups contain a carbon which participates in five bond orbitals: a C—C bond to C\textsuperscript{+} and two polarized and equivalent bent σ bonds to each of the oxygen atoms which we represent as

\begin{equation}
\text{C—C} \quad \text{O—O}
\end{equation}
TABLE III

<table>
<thead>
<tr>
<th>Ionic Species</th>
<th>Charge</th>
<th>Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lysine positive ion</td>
<td>0.68</td>
<td>-NH₄⁺</td>
</tr>
<tr>
<td>Arginine positive ion</td>
<td>0.78</td>
<td>C⁺</td>
</tr>
<tr>
<td>Glutamic acid negative ion</td>
<td>-0.92</td>
<td>2 O atoms in CO₃⁻</td>
</tr>
<tr>
<td>Aspartic acid negative ion</td>
<td>-0.90</td>
<td>2 O atoms in CO₂⁻</td>
</tr>
</tbody>
</table>

Neutral Molecules

<table>
<thead>
<tr>
<th>Molecule</th>
<th>OH group charge</th>
</tr>
</thead>
<tbody>
<tr>
<td>Threonine</td>
<td>-0.11</td>
</tr>
<tr>
<td>4-hydroxyproline</td>
<td>-0.08</td>
</tr>
<tr>
<td>Serine</td>
<td>-0.07</td>
</tr>
<tr>
<td>Glutamic acid</td>
<td>-0.06</td>
</tr>
<tr>
<td>Aspartic acid</td>
<td>-0.05</td>
</tr>
<tr>
<td>Tyrosine</td>
<td>0.03</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Molecule</th>
<th>C=O group charge</th>
<th>NH₂ group charge</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glutamic acid</td>
<td>-0.24</td>
<td></td>
</tr>
<tr>
<td>Aspartic acid</td>
<td>-0.17</td>
<td></td>
</tr>
<tr>
<td>Asparagine</td>
<td>-0.21</td>
<td>-0.01</td>
</tr>
<tr>
<td>Glutamine</td>
<td>-0.23</td>
<td>-0.02</td>
</tr>
<tr>
<td>Lysine</td>
<td>-0.07</td>
<td></td>
</tr>
</tbody>
</table>

This description shows that the electron density is polarized toward oxygen away from carbon, and, also, that the bonds are bent and are not co-axial with C-O internuclear axes. The bonding of the oxygen atoms to carbon can be thought of as the average of the two valence bond structures

\[
\begin{array}{c}
\text{\text{C=O}} \\
\begin{array}{c}
\hat{\alpha} \\
\hat{\gamma}
\end{array}
\end{array}
\]

with fractional bonding at the carbon atom. This formulation is analogous to the results (46, 47) found in the boron hydrides where fractional bonding (more than four bonds to boron) is observed

\[
\begin{array}{c}
\text{\text{B–B}} \\
\begin{array}{c}
\hat{\text{B}} \\
\hat{\text{B}}
\end{array}
\end{array}
\]

where \( \text{B}_1 \) has five bonds to it as is found in \( \text{B}_1\text{H}_{14} \). We note that the carbons still have approximately 4 valence electrons and the borons approximately 3 valence electrons even though these electrons participate in more than the four bond orbitals required by the octet rule. The localization corresponds to an average of two sets of resonance bond orbitals in order to obey the constraints of molecular geometry. The localized molecular orbitals yield fractional bonding because of this averaging of bonds and also not to violate the formal octet rule. The populations (see Table IV) show that there is more density in the C=O bonds on oxygen than on carbon: the hybridization at oxygen is sp\(^1\) (almost pure p), while at carbon the hybridization is between sp\(^2\) and sp\(^3\). As a consequence, we find that the two lone pairs on each oxygen atom have high s character being sp and sp\(^4\) hybrids. The large s character suggests that the lone pairs are tightly bound. The unequal nature of the two lone pairs may be an artifact of the Boys criteria but in work on CO\(_2\) and CO\(_2\) the lone pairs are found to be identical. The C-C\(^*\) bond is strongly polarized toward C\(^*\) with roughly 0.70 e on carbon and 1.30 e on C\(^\bullet\) giving an average bond dipole moment of 1.0 Debye. This result is in part a consequence of the fractional bonding at the C carbon. The C\(^*\)-N bond is strongly polarized toward nitrogen with a bond dipole moment averaging to 1.6 Debye. The C\(^\bullet\)-N and C\(^*\)-N bonds are essentially bonds in which the centroid of charge lies close to the internuclear axis; here the hybridizations are between sp\(^1\) and sp\(^3\). The delocalization index (5) for the C\(^\bullet\)-N bond is comparable to that found for most C-C single bonds averaging 14.5%. However, the delocalization index for the C-C\(^*\) bond is remarkably high, averaging 18.5% with most of the delocalization onto the carboxylate oxygens. It is the highest delocalization index found in these molecules except for the \( \tau \) bonds in aromatic rings which have a delocalization index of approximately 20%.

Alcoholic oxygens have two equivalent lone pairs with slightly more s character than found in sp\(^3\) hybrids with an exception occurring for glutamic acid. In glutamic acid, one lone pair has roughly sp character, while the other lone pair is an sp\(^3\) hybrid with the latter lone pair having some delocalization (0.130) to C\(^\bullet\). The C–OH bond has a high s character at carbon (sp\(^3\) and high p character at oxygen (sp\(^2\)~) with the bond polarized toward oxygen (1.15 e on oxygen and 0.85 e on carbon). As expected for a polarized bond with such high p character on 1 atom, the centroid of charge is displaced from the C–O internuclear axis at the angles given in Table IV. The OH bonds also have low s character on oxygen; again they are sp\(^3\) hybrids.

The bonding in the carbonyl groups is similar to that found for the C=O bond in the carboxylate ion group, but there is less polarization of the bond (0.90 e on carbon and 1.10 on oxygen). Again the hybridization at oxygen (sp\(^2\)~) shows almost pure p character on oxygen while the hybridization on carbon shows higher p character with sp\(^3\) hybridization.

Aliphatic C-C bonds tend to show little deviation from simple predictions. The centroid of charge lies close to the internuclear axes; the hybridizations are between sp\(^2\) and sp\(^3\) and the populations differ by at most 0.10 e.

The two equivalent sets of localized molecular orbitals for benzene (38) are shown below

\[
\begin{array}{c}
\text{\text{C–C}} \\
\begin{array}{c}
\hat{\alpha} \\
\hat{\beta}
\end{array}
\end{array}
\]

and consist of alternating \( \alpha \) bonds and equivalent bent \( \tau \) bonds around the ring. These sets correspond to the simple valence-bond resonance structures for benzene (Kekulé structures). The \( \tau \) bonds are somewhat delocalized, have roughly sp\(^4\) hybridization at each carbon, and the centroid of charge lies

Stabilization of the bonds in the benzene rings of phenylalanine and tryptophan show only minor polarization effects, and the charges on carbon atoms are almost equal. In tyrosine, definite polarization effects in the σ bonds are observed due to the presence of the OH group as shown in Table IV.

Except for the bonding in the carboxylate ion group, which shows a carbon participating in five bond orbitals, the bonding patterns given by the localized molecular orbitals have conformed qualitatively to simple chemical ideas. We now discuss the bonding that does not conform to simple chemical stereotypes as found in the molecules asparagine and glutamine, histidine, tryptophan, and arginine. These molecules all contain nitrogen atoms. The bonding patterns as found in these molecules have previously been discussed by Kleier et al. (48).

The major feature is that the lone pair on nitrogen tends to interact with the adjacent carbon (electron sink) with the carbon thus participating in more than four bond orbitals. In tryptophan, NC' has a highly delocalized lone pair with populations on NC1 of 1.60 e, on C*' of 0.20 e, and CY of 0.17 e with the lone pair having almost pure p character. The C—N σ bonds consequently bend away from the axis with the angles given in Table IV. The interaction of the σ bonds and the “lone pair” is also observed in the second derivative test (8) where the highest eigenvector corresponds to mixing these three localized molecular orbitals. This same delocalized lone pair was also observed by Kleier et al. (48) in guanine. The lone pairs in asparagine and glutamine actually bond to the adjacent carbon as found for formamide (48) but in different fashions. For asparagine, two highly polarized but equivalent C—N bonds are observed, but in glutamine the pair of bonds...
split and unequal populations are found (see Table IV) which when averaged give the same results obtained for asparagine. As found for the carboxylate ion group, the carbon atom participates in five bond orbitals. In arginine, the carbon C\textsuperscript{\alpha} participates in six bond orbitals as the 3 nitrogen atoms adjacent to the central C\textsuperscript{\alpha} are each connected by two \( \pi \) bonds to this carbon. The atom N\textsuperscript{\textordervigiliae} is connected by two equivalent \( \tau \) bonds to C\textsuperscript{\alpha}, while the other two nitrogens have unequal \( \tau \) bonds with all parameters given in Table IV. For histidine we find a normal set of C\equiv C \( \pi \) bonds and a normal set of C\equiv C \( \sigma \) bonds to this carbon. The atom N\textsuperscript{\textordervigiliae} is connected by two

Acknowledgment—The authors thank Professor D. A. Kleier for helpful discussions. We also thank Jean Evans for the drawings.

REFERENCES

37. Koopmans, T. (1934) Physica 1, 104-113
46. Lipscomb, W. N. (1973) Accds. Chem. Res. 6, 257-262
Electronic structure and bonding of the amino acids containing first row atoms.
D A Dixon and W N Lipscomb


Access the most updated version of this article at http://www.jbc.org/content/251/19/5992

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 0 references, 0 of which can be accessed free at http://www.jbc.org/content/251/19/5992.full.html#ref-list-1