Iowa State University

From the SelectedWorks of Mark S. Gordon

October, 1992

Uncatalyzed Peptide Bond Formation In the Gas Phase

Jan H. Jensen
Kim K. Balgridge
Mark S. Gordon

Available at: https://works.bepress.com/mark_gordon/118/
Uncatalyzed Peptide Bond Formation In the Gas Phase

Jan H. Jensen,* Kim K. Baldrige,1 and Mark S. Gordon*1

Department of Chemistry, North Dakota State University, Fargo, North Dakota 58105-5316, and San Diego Supercomputer Center, San Diego, California 92186-9784 (Received: May 26, 1992)

Several levels of electronic structure theory are used to analyze the formation of a peptide bond between two glycine molecules. Both a stepwise and concerted mechanism were considered. The energetic requirements for the stepwise and concerted mechanisms are essentially the same within the expected accuracy of the methods used. A simpler model system comprised of formic acid and ammonia is found to provide a good representation of the essential features of dipeptide formation. Total electron densities and localized molecular orbitals are used to interpret the mechanisms.

I. Introduction

The peptide bond is of central importance to protein chemistry in particular and biological chemistry in general. It provides the link between amino acid subunits of proteins and imposes an important conformational restriction on the main chain. While Nature has crafted a very complicated machinery for the making and breaking of peptide bonds, chemists have succeeded in this regard as well, starting with Fisher's first peptide bond synthesis in 1903. Controlled hydrolysis of the peptide bond is central to the field of protein sequencing, initiated by Sanger's determination of the amino acid sequences of insulin. Therefore, analysis of the details of the mechanism leading to the formation of a dipeptide warrants investigation. Important questions to be answered in this regard are as follows: (1) Does this bond formation occur in a concerted or stepwise manner? (2) What is the molecular and electronic structure of each transition state? (3) What are the associated barrier heights? (4) What are the effects of entropy on the details of the mechanism? (5) What is the nature of solvent effects on the apparent mechanism? (6) How do enzymes and other catalysts aid peptide bond formation/breaking? The first four of these questions will be addressed in this paper.

The more general case of amide bond formation has been extensively studied. The mechanism of amide bond formation was studied experimentally by Jencks and co-workers1 for the acid/base catalyzed aminolysis of alkyl esters in aqueous solutions. Evidence was presented to support a preference for a stepwise mechanism, based on pH-dependence studies using various esters. The mechanism shown in (1) was proposed.5c The rate-determining step is the proton transfer (step 2), rather than the amine attack (step 1), mediated by an acid (step 2a) or base (step 2b) catalyst (through either $T_1$ or $T_2$ respectively) or the solvent. The third step, i.e. the breakdown of the tetrahedral intermediate and the formation of the amide bond, could not be probed due to the lower barrier. One way to address this problem is to consider a related reaction, namely the hydrolysis of the amide bond, which mechanistically is related to amide bond formation since it is the reverse of (1) for $R' = H$. Thus, step 3 in reaction 1 corresponds to the initial hydration of the amide bond, to be followed by complete hydrolysis through additional steps.6

The hydration of amides is a special case of the general nucleophilic attack on carboxyl centers for which three competing mechanisms exist (2).7 Depending on the nucleophile and pH, Q and S may either represent intermediates or mechanistically extreme representations of the transition state for the concerted path. In acidic solution the nucleophile is weak (Nuc: = H$_2$O) and the mechanism is "$S$-like." For amides, this is facilitated through the delocalization of positive charge on the nitrogen$^8$

Additional steps lead to complete hydrolysis$^8$

In basic solution the nucleophile is strong (Nuc: = OH$^-$) and Q is regarded as an intermediate that is broken down by the following steps$^9$

No experimental evidence exists for a concerted ester aminolysis/amide hydrolysis mechanism.

There have been several previous theoretical studies on prototypical reactions which are intended to mimic peptide bond
Uncatalyzed Peptide Bond Formation in the Gas Phase

II. Computational Approach

An important aspect of this study is to determine levels of theoretical treatments of large molecules which are both efficient and reliable. Therefore, several levels of theory will be discussed. The molecular structures of all stationary points have been determined with both the semiempirical AM1 method and the minimal STO-3G basis set, at the self-consistent field (SCF) level of theory. For the model system, stationary points were identified with the 6-31G(d,p) basis set as well, at both the SCF and MP2 levels of theory. Geometry optimizations were performed with the aid of analytically determined gradients and the search algorithms contained in MOPAC (version 5.0) (AM1; the "NOMM" option was used where applicable), GAMES, and GAUSSIAN86,19 and GAUSSIAN98 (ab initio). The nature of each SCF stationary point was established by calculating (analytically for ab initio wavefunctions, numerically for semiempirical methods) and diagonalizing the matrix of energy second derivatives (hessian) to determine the number of imaginary frequencies (zero for a local minimum, one for a transition state). The MP2 and SCF geometries are sufficiently similar so that the considerable computational expense of MP2 hessians was considered unnecessary.

The two mechanisms (stepwise and concerted) for the model system were initially explored with the semiempirical methods. The three transition states (TS1, TS2, and TS3; see Schemes I and II) were located and identified by following the gradient downhill in both directions using the gradient following routine implemented in MOPAC. The resulting structures were optimized and verified as minima by calculating the hessian. These geometries were used as initial guesses for subsequent ab initio calculations.

Several computational problems arise in the study of systems as complex as those in reaction 7. One is how to select the lowest-energy conformation for each structure from the many conformational isomers that exist in a system this large. Another is how to efficiently optimize structures this complex. The approach taken in this study is the following. First we define that part of the structure that glyglyglycine and the model system have in common as the "model system part" of each structure. Those atoms which are directly involved in a transition state are collectively referred to as the "TS part." The AM1 geometry for the model system part of the three transition states was taken from the model system calculations. Then (N−)H and (C−)H were replaced with CH₂COOH and CH₂NH₂, respectively, each arranged so as to most closely resemble the global minimum conformation of gas-phase glycine.21 These two parts were subsequently energy minimized while the geometry of the TS part was kept frozen. Then all geometrical parameters were relaxed and the TS was optimized using the NLLSQ22 option in MOPAC. Again, the gradient was followed downhill in both directions starting at each TS, the resulting structures optimized and verified as minima. These structures were then used as initial guesses for the subsequent STO-3G optimizations. It was found that this "freeze-unfreeze" technique was necessary to obtain the STO-3G transition states as well. Clearly, many more conformational isomers of the stationary points found in this study exist. However, the structures presented here are thought to be representative of both the structures and energies involved in gas-phase peptide bond formation.

To obtain improved predictions of energetics, higher level calculations were performed at the SCF-optimized geometries. These single point calculations were performed with the 6-31G(d,p) and 6-311G(d,p) basis sets and frozen-core many-body perturbation theory through second (MP2) or full fourth (MP4) order. The notation used to describe such single point calculations is A/B, signifying a calculation at theoretical level A performed at a geometry obtained at theoretical level B. The MP4/6-311G(d,p) results were obtained by extrapolation from MP4/6-31G(d,p) and MP2/6-311G(d,p), assuming that improvements in basis set and level of correlation are additive.24

The technique used to obtain localized molecular orbitals (LMOs) is that described by Boys.25 The electron density of each
LMO is calculated over a 61 x 61 grid, squared, multiplied by the orbital occupation number (two for RHF) and summed up to give the total electron density. In the analysis of the total electron density the presence of a bond critical point\(^2\)\(^6\) (that is, a saddle point in the electron density) connecting two atoms is considered to be indicative of bonding. Internuclear distances and LMOs then provide qualitative information about the relative strength of each bonding interaction. The information thus obtained is used to elucidate key features of the mechanisms by translating the MO-based results into a valence-bond-like description (section III.B).

The zero point and Gibbs free energies are calculated using standard statistical-mechanical formulas\(^2\)\(^7\) (with the harmonic oscillator–rigid rotor approximation) as implemented in GAMESS.

### III. Results and Discussion

The concerted and stepwise mechanisms for the dipeptide bond formation are illustrated in Schemes I and II, respectively. In these schemes, and in the related tables and discussions, R = R′ = H for the model system ([5]), while for the actual dipeptide bond formation ([7]), R = CH\(_2\)NH\(_2\) and R′ = CH\(_2\)COOH. For the concerted mechanism, there is a single four-center transition state, denoted TS3 in Scheme II, with an associated classical barrier height \(\Delta E_b\). For the stepwise process, an intermediate INT2 is separated from reactants by a barrier \(\Delta E_b\); at transition state TS1 and from products by a barrier \(\Delta E_f\) at transition state TS2. A key question is therefore whether the barrier height \(\Delta E_b\) is greater or smaller than the larger of the two barriers \(\Delta E_f\) and \(\Delta E_b\). The barrier separating INT2 and INT1 in Scheme I corresponds to an internal rotation in the intermediate.

This section is organized as follows. First the SCF-optimized molecular structures for each basis set and reaction are compared. Second, those parts of the TS wavefunctions (localized molecular orbitals and total density) common to both (5) and (7) are analyzed to give a valence-bond-like description of the mechanisms. Third, the energetics of the two mechanisms are discussed and the effect of entropy is addressed. Lastly, the effect of electron correlation (MP2) on the molecular structure is examined.

#### A. Molecular Structures

The structures of the model system (5) are investigated using both the 6-31G(d) and STO-3G basis sets and the semiempirical AMI Hamiltonian so that the accuracy of the latter two may be gauged by comparison to the former. These are then used to investigate the glycylglycine system (7).

Two key features of the predicted geometries are of interest: (1) How well do the various levels of theory agree with regard to the prediction of key geometric parameters? (2) How similar are the key common geometric features in the model compound and the full dipeptide? The two questions are addressed first for the products and reactants of both systems and subsequently for the transition states.

The predicted geometries for glycine and the dipeptide are displayed in Figures 1 and 2, respectively. In each of these figures, both the STO-3G and the AM1 (in parentheses) parameters are given and in Figure 1 experimental parameters\(^2\)\(^{12}\) are included in brackets. The glycine conformation shown in Figure 1 is the global minimum on the potential energy surface of gas-phase glycine\(^2\)\(^1\) and is used as reactant for both mechanisms. The structure agrees well with structures obtained using a higher level of theory\(^2\)\(^{12}\) as well as with neutron diffraction\(^2\)\(^{12}\). Bond lengths calculated with STO-3G are too long by 0.01–0.03 Å relative to experiment (due to the small basis set). Parts a and b of Figure 2 depict the dipeptide product of the stepwise and concerted mechanisms, respectively. Bond lengths predicted by AM1 and STO-3G generally agree to within 0.05 Å and bond angles to within 7°. Judging from the O=CC–N dihedral angles, the two glycine subunits in Figure 2 retain the conformation of the monomer, the largest deviation being 31.9° calculated with AM1 for 2a. The peptide bond linkage deviates from planarity by up to 30°. Both methods predict a 0.02–0.03 Å increase in the C–N bond length relative to formamide. Bond lengths and angles of the two glycylglycine conformers differ little. The largest disagreements are for the AM1 C\(_2\)–N and C\(_1\)–C bond lengths.

The predicted geometries for formic acid and formamide are presented in Table I, while those for the three transition states (TS1, TS2, TS3) and the intermediate INT2 are given in Tables II–V, respectively. Only the most interesting geometric parameters are given in the latter four tables. Complete geometries are available as supplementary material. The molecular structures of the three transition states and one intermediate are depicted schematically in Figures 3–6, respectively.

The structures of formic acid and formamide are well predicted by all methods. For the transition states, there are significant deviations among the various theoretical methods. The SCF/6-
Table II: Structure of Transition State TSI for Diglycine (Model Compound)

<table>
<thead>
<tr>
<th>Bond</th>
<th>MP2/6-31G(d)</th>
<th>6-31G(d)</th>
<th>STO-3G</th>
<th>AM1</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\text{R(C}^-\text{O}))</td>
<td>1.350</td>
<td>1.323</td>
<td>1.386</td>
<td>1.357</td>
</tr>
<tr>
<td>(\text{R(C}^-\text{O}))</td>
<td>1.212</td>
<td>1.181</td>
<td>1.214</td>
<td>1.230</td>
</tr>
<tr>
<td>(\text{R(C}^-\text{H}))</td>
<td>1.096</td>
<td>1.084</td>
<td>1.104</td>
<td>1.103</td>
</tr>
<tr>
<td>(\text{R(O}^-\text{H}))</td>
<td>0.980</td>
<td>0.953</td>
<td>0.990</td>
<td>0.971</td>
</tr>
<tr>
<td>(\text{A(O}^-\text{C}^-\text{O}))</td>
<td>125.1</td>
<td>124.9</td>
<td>123.6</td>
<td>117.6</td>
</tr>
<tr>
<td>(\text{A(O}^-\text{C}^-\text{O}))</td>
<td>125.4</td>
<td>124.7</td>
<td>126.0</td>
<td>130.1</td>
</tr>
<tr>
<td>(\text{A(O}^-\text{C}^-\text{O}))</td>
<td>106.1</td>
<td>108.7</td>
<td>104.8</td>
<td>110.6</td>
</tr>
<tr>
<td>(\text{D(H}^-\text{N}^-\text{C}^-\text{O}))</td>
<td>1.362</td>
<td>1.348</td>
<td>1.403</td>
<td>1.367</td>
</tr>
<tr>
<td>(\text{R(C}^-\text{O}))</td>
<td>1.224</td>
<td>1.193</td>
<td>1.218</td>
<td>1.243</td>
</tr>
<tr>
<td>(\text{R(C}^-\text{H}))</td>
<td>1.105</td>
<td>1.091</td>
<td>1.105</td>
<td>1.114</td>
</tr>
<tr>
<td>(\text{R(N}^-\text{H}))</td>
<td>1.011, 1.009</td>
<td>0.996, 0.993</td>
<td>1.014, 1.013</td>
<td>0.990, 0.986</td>
</tr>
<tr>
<td>(\text{A(O}^-\text{C}^-\text{N}))</td>
<td>124.7</td>
<td>124.9</td>
<td>124.4</td>
<td>122.0</td>
</tr>
<tr>
<td>(\text{A(H}^-\text{C}^-\text{N}))</td>
<td>112.4</td>
<td>112.7</td>
<td>111.4</td>
<td>112.6</td>
</tr>
<tr>
<td>(\text{A(C}^-\text{N}^-\text{N}))</td>
<td>118.4, 121.4</td>
<td>119.2, 121.9</td>
<td>120.5, 121.3</td>
<td>120.6, 121.2</td>
</tr>
</tbody>
</table>

*Bond lengths in angstroms, angles in degrees.

Table III: Structure of Transition State TS2 for Diglycine (Model Compound)

<table>
<thead>
<tr>
<th>Bond</th>
<th>MP2/6-31G(d)</th>
<th>6-31G(d)</th>
<th>STO-3G</th>
<th>AM1</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\text{R(C}^-\text{N}))</td>
<td>1.552</td>
<td>1.548</td>
<td>1.787 (1.756)</td>
<td>1.552 (1.531)</td>
</tr>
<tr>
<td>(\text{R(C}^-\text{O}))</td>
<td>1.336</td>
<td>1.322</td>
<td>1.317 (1.315)</td>
<td>1.346 (1.346)</td>
</tr>
<tr>
<td>(\text{R(N}^-\text{H}))</td>
<td>1.224</td>
<td>1.211</td>
<td>1.139 (1.128)</td>
<td>1.281 (1.284)</td>
</tr>
<tr>
<td>(\text{R(O}^-\text{H}))</td>
<td>1.380</td>
<td>1.334</td>
<td>1.378 (1.404)</td>
<td>1.423 (1.419)</td>
</tr>
<tr>
<td>(\text{A(H}^-\text{N}^-\text{C}))</td>
<td>72.9</td>
<td>72.4</td>
<td>69.3 (70.7)</td>
<td>80.6 (80.7)</td>
</tr>
<tr>
<td>(\text{A(N}^-\text{C}^-\text{O}))</td>
<td>96.6</td>
<td>95.9</td>
<td>88.7 (89.9)</td>
<td>93.3 (94.0)</td>
</tr>
<tr>
<td>(\text{A(C}^-\text{O}^-\text{H}))</td>
<td>112.7</td>
<td>77.0</td>
<td>80.8 (79.8)</td>
<td>83.4 (83.0)</td>
</tr>
<tr>
<td>(\text{D(H}^-\text{H}^-\text{N}^-\text{C}))</td>
<td>-9.0</td>
<td>-5.3</td>
<td>-1.2 (-0.4)</td>
<td>0.6 (-0.2)</td>
</tr>
</tbody>
</table>

*Bond lengths in angstroms, angles in degrees.

Table IV: Structure of Transition State TS3 for Diglycine (Model Compound)

<table>
<thead>
<tr>
<th>Bond</th>
<th>MP2/6-31G(d)</th>
<th>6-31G(d)</th>
<th>STO-3G</th>
<th>AM1</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\text{R(C}^-\text{N}))</td>
<td>1.577</td>
<td>1.524</td>
<td>1.695 (1.676)</td>
<td>1.546 (1.522)</td>
</tr>
<tr>
<td>(\text{R(C}^-\text{O}))</td>
<td>1.812</td>
<td>1.877</td>
<td>1.615 (1.603)</td>
<td>1.575 (1.583)</td>
</tr>
<tr>
<td>(\text{R(N}^-\text{H}))</td>
<td>1.200</td>
<td>1.135</td>
<td>1.241 (1.220)</td>
<td>1.332 (1.340)</td>
</tr>
<tr>
<td>(\text{R(O}^-\text{H}))</td>
<td>1.329</td>
<td>1.378</td>
<td>1.167 (1.176)</td>
<td>1.303 (1.294)</td>
</tr>
<tr>
<td>(\text{A(H}^-\text{N}^-\text{C}))</td>
<td>81.2</td>
<td>85.5</td>
<td>73.1 (73.6)</td>
<td>81.9 (82.5)</td>
</tr>
<tr>
<td>(\text{A(N}^-\text{C}^-\text{O}))</td>
<td>82.9</td>
<td>81.5</td>
<td>81.4 (82.0)</td>
<td>87.0 (87.1)</td>
</tr>
<tr>
<td>(\text{A(C}^-\text{O}^-\text{H}))</td>
<td>69.3</td>
<td>66.3</td>
<td>78.0 (77.8)</td>
<td>81.7 (81.6)</td>
</tr>
<tr>
<td>(\text{D(H}^-\text{H}^-\text{N}^-\text{C}))</td>
<td>-6.5</td>
<td>-6.6</td>
<td>0.0 (-0.8)</td>
<td>-2.8 (-2.0)</td>
</tr>
</tbody>
</table>

*Bond lengths in angstroms, angles in degrees.

Figure 3. RHF/STO-3G optimized structure of TS1 in the stepwise mechanism of (7).

Figure 4. RHF/STO-3G optimized structure of TS2 in the stepwise mechanism of (7).

Figure 5. RHF/STO-3G optimized structure of TS3 in the concerted mechanism of (7).

Figure 6. RHF/STO-3G optimized structure of INT2 in the stepwise mechanism of (7).

31G(d) geometry is used as reference for each structure. The TS best represented by the simpler methods is TS1. The only serious deviation (0.208 Å) is the overestimation of the C-N bond length by 0.208 Å using STO-3G. Large deviations in the key geometric
TABLE V: Structure of Intermediate INT2 for Diglycine (Model Compound)

<table>
<thead>
<tr>
<th>MP2/6-31G(d)</th>
<th>6-31G(d)</th>
<th>STO-3G</th>
<th>AM1</th>
</tr>
</thead>
<tbody>
<tr>
<td>R(C-O-N)</td>
<td>(1.443)</td>
<td>(1.433)</td>
<td>1.504 (1.489)</td>
</tr>
<tr>
<td>R(C-O)</td>
<td>(1.409)</td>
<td>(1.386)</td>
<td>1.416 (1.426)</td>
</tr>
<tr>
<td>R(C=O)</td>
<td>(1.404)</td>
<td>(1.383)</td>
<td>1.432 (1.426)</td>
</tr>
<tr>
<td>A(N-C-O)</td>
<td>(108.2)</td>
<td>(108.6)</td>
<td>113.5 (109.1)</td>
</tr>
<tr>
<td>A(O-C-O)</td>
<td>(116.0)</td>
<td>(115.2)</td>
<td>108.1 (114.0)</td>
</tr>
<tr>
<td>A(O-H)</td>
<td>(106.5)</td>
<td>(107.0)</td>
<td>108.1 (107.1)</td>
</tr>
</tbody>
</table>

*Bond lengths in angstroms, angles in degrees.

parameters are more common for TS2 and TS3. In the case of STO-3G only the C-O (N-H) distance is within 0.1 Å of the 6-31G(d) values for TS2 (TS3). The largest deviation is the 0.274 Å underestimation of the C-O distance in TS3. Bond angles are generally within 10° of those predicted by 6-31G(d). The AM1 method does somewhat better in that it brings the O-H, H-O, and O-H, distance deviations down to 0.075 and 0.084 Å, respectively. The largest AM1 deviation (0.294 Å) is the underestimation of the C-O distance in TS3. The deviations in bond angles predicted by AM1 are generally within 15°. An exception is that the O-H-O angle in TS2 is underestimated by 24°. It appears the leaving water molecules in the TSs are more intact while also more strongly attached to the carbon for AM1 and especially STO-3G relative to 6-31G(d).

For INT2 (Table V; Figure 6) bond lengths vary between basis sets, but the deviations for both AM1 and STO-3G are on the order of 0.05 Å. Bond angle deviations are on the order of 3-5°. Both STO-3G and AM1 have been used to analyze the reactions of the model system and the actual dipeptide bond forming reaction. For the STO-3G basis set, the key geometric parameters predicted for the model system are in excellent agreement with those for the dipeptide, even though there are significant differences relative to 6-31G(d). All bond lengths agree to within 0.03 Å and the angles generally to within ±1.5° (2-3° for some INT2 parameters). The agreement obtained using AM1 is comparatively good. So, the reaction between formic acid and ammonia appears to be a realistic structural model for the glycine + glycine reaction. Thus, the 6-31G(d) results obtained for the model system are very likely to be an accurate representation of (7) and can be used to interpret its mechanism.

In Figure 6 a hydrogen bond is drawn between an oxygen at the C-terminal and one of the hydroxyl hydrogens, as the O-H distance is 2.652 Å and the O-H-O angle is 159.5°, both of which are optimum for hydrogen bonding. In addition, one of the O(---H-O) LMO lone pairs is delocalized along the C-O-H bond axis.

B. Electronic Wavefunction. Figures 7-9 display four key 6-31G(d) localized molecular orbitals (LMOs) and the total density in the bond breaking/making region for each transition state. Key features of each molecular structure are included for comparison (and explained in the figure legend). In this manner, key features of the mechanism can be elucidated.

First consider TS1 (Figure 7). The C-N bond is essentially formed at the TS. Here, the C-N bond distance in a fully optimized structure of zwitterionic glycine, rather than the conformer shown in Figure 1, serves as a reference since the C-N(H4) bond length is considerably shorter (0.07 Å) than that in C-N(H4). While the C-N distance is 0.044 Å longer than the bond length in the glycine zwitterion, there is a clear bonding orbital and the total electron density exhibits a saddle point connecting C and N. The C-O bond appears to be a single bond judging from the bond length, identical to the single bond in formic acid. From the density it is apparent that although there is some H2-O bonding, there is more electron density between H4 and N. This is consistent with the fact that the N-H bond distance is closer to an equilibrium value than is the O-H distance: the N-H bond is elongated by 0.204 Å while the O-H distance is stretched by almost twice that (0.380 Å). This suggests that there is an N-H bond and only weak bonding between O and H. Thus, the bonding picture of TS1 (2) is very much like the 7' intermediate in reaction 1 in that C-N bond formation precedes hydrogen transfer.

Next consider TS2 (Figure 8). No bond appears to exist connecting O2 and the C atom. This is especially apparent from the lack of a C-O bond critical point in the total density and the very long 1.925 Å internuclear distance. It is also apparent that the corresponding LMO (Figure 8d) is essentially an O2 lone pair. Weak bonding exists between O2 and H4, but a much stronger bond is found between H4 and O. Evidence for these assertions is found in the total density (bond critical points are evident between both H4-O2 and H2-O, but there is a larger build-up of charge density between the latter than the former) as well as in the bond lengths. The deviation from equilibrium bond distance is more than twice as large for H4-O2 than for H2-O (0.369 Å versus 0.166 Å). The LMO connecting O2 and H4 (Figure 8) shows significant delocalization onto the C-O2 bond which adds to a rather diffuse density build-up between the latter two atoms. This, in conjunction with C-O2 and C-N bond lengths that are intermediate between double and single bond lengths (the C-N distance is 1.330 Å), suggests a partial C-O2 double bond. Hence, TS2 might be written as

\[
\text{O} = \text{C} = \text{N} \\
\text{H} \quad \text{H} \\
\text{O} \quad \text{H} \quad \text{C} \quad \text{N} \quad \text{H}_2
\]

Since the C-O2 distance is longer and the C-N distance shorter than those in formamide, one expects more [O=C=O] character in formamide. Note the strong similarity between 3 and 1 in the acid catalyzed hydrolysis mechanism, indicating a S-like TS in the gas phase as well as in acidic aqueous solution.

Finally, an intermediate picture emerges for TS3 (Figure 9). The N-H4 bond is more completely formed than in TS1, since the N-H distance is closer to its equilibrium value. The O-H interaction is weaker than in TS1 and TS2 although a bond critical point is present. While the C-N bond distance is 0.02 Å shorter than in TS1, the LMO and total electron density remain virtually unchanged. Thus, the C-N bond is almost completely formed. The C-O2 bond length is 0.05 Å shorter than in TS2 but there is still no significant density build-up between C and O (and no bond critical point), and the LMO remains primarily an O orbital. These observations lead one to conclude that there is no C-O3 bond. A possible bonding picture of TS3 is

\[
\text{O} = \text{C} = \text{N} \\
\text{H} \quad \text{H} \\
\text{O} \quad \text{H} \quad \text{C} \quad \text{N} \quad \text{H}_2
\]
Thus, in the concerted aminolysis-mechanism the proton transfer is preceded by N-C bond formation as it is in the stepwise mechanism. Conversely, amide hydrolysis via a concerted mechanism involves a nucleophilic attack on a protonated substrate, much like the first step of the stepwise mechanism in the gas phase and acidic aqueous solution.

C. Energetics. The total energies, zero point vibrational energies (ZPE), and Gibbs free energies at 298 K (\(G_{\text{ZPE}}\)) for all species of interest at several different levels of theory are included in the supplementary material. The corresponding energy differences (\(\Delta E\)), 0 K enthalpy differences [\(\Delta H_0 = \Delta E + \Delta(\text{ZPE})\)], and Gibbs free energy differences are listed in Tables VI and VII. Referring to Schemes I and II, the values of greatest interest are the heights of the barriers at TS1 and TS2 (the stepwise transition states) vs the barrier height at TS3 (the transition state for the concerted process). These barriers are referred to as \(\Delta E_1\), \(\Delta E_3\), and \(\Delta E_6\), respectively (Table VI). Also of interest is the relative stability of the two intermediates (\(\Delta E_2\) and \(\Delta E_4\)) and the overall endothermicity of the reaction (\(\Delta E_1\) for the model system and the stepwise process of (7), \(\Delta E_3\) for the concerted mechanism of glycyglycine formation). These values are listed in Table VII.

Model System. First, consider the relative values of \(\Delta E_2\) and \(\Delta E_6\), the two steps in the stepwise mechanism. At the SCF level of theory for the model system, the minimal basis sets (AM1, STO-3G) predict the second step in this process (\(\Delta E_6\), conversion of the intermediate to products) to be considerably more ener-
Figure 7. Four localized RHF/6-31G(d) MOs involved in bond making/breaking in TS1 of (5) (a-d) and the total density in the N-H-O plane (f). Plots a (N-H bond LMO) and b (predominantly O lone pair) use the N-H-O plane while plots c (N-C bond LMO) and d (C-O bond LMO) use the N-C-O plane. Figure 7e schematically represents the orientation of the molecule, the RHF/6-31G* bond lengths in angstroms and related parameters from (g) formic acid and (h) glycine zwitterion are in parentheses. All plots in this and the subsequent two figures have maximum contour lines of 1 bohr^{-3/2} and increments of 0.05 bohr^{-3/2}.

The relative values of \( \Delta E_1 - \Delta E_3 \) range from 10 kcal/mol for AM1 to 28 kcal/mol for STO-3G. The use of the larger basis sets 6-31G(d) and 6-31G(d,p) at the minimal basis set geometries decreases these differences to 2-4 kcal/mol. This is a dramatic basis set effect. Only when the larger 6-31G(d) basis set (denoted “C” in Tables VI and VII) is used to predict the geometry do the SCF calculations predict the first step in the stepwise mechanism to be the higher energy of the two. Using the largest basis set employed here, 6-311G(d,p) at the 6-31G(d) geometries, the SCF calculations predict \( \Delta E_1 \) to be 7.7 kcal/mol greater than \( \Delta E_3 \). The addition of correlation corrections reduces both \( \Delta E_1 \) and \( \Delta E_3 \) by about 10 kcal/mol, but has a much smaller effect on the relative values of these two barrier heights. At the highest level of theory, MP4/6-31G(d,p)/RHF/6-31G(d), \( \Delta E_1 \) is greater than \( \Delta E_3 \) by 1.7 kcal/mol. Note, however, that the difference between MP2 and MP4 barriers is generally less than 2.0 kcal/mol.

At the SCF level of theory, the barrier height, \( \Delta E_6 \), for the concerted mechanism is predicted with all basis sets except STO-3G to be higher than either of the barriers in the stepwise mechanism. Recall that STO-3G predicts a value for \( \Delta E_6 \) that is much too large (see Table VI). Improvement of the basis set to 6-31G(d), 6-31G(d,p), or 6-311G(d,p) raises \( \Delta E_6 \) by about 10 kcal/mol, relative to the minimal basis sets. So, at the SCF level of theory, the stepwise process is predicted to be more viable than
the concerted mechanism by about 10 kcal/mol, using any of the three largest basis sets. The spread in barrier heights for $\Delta E_4$ predicted by these three basis sets at the 6-31G(d) geometry is only 2.6 kcal/mol. Addition of correlation corrections (MP2) reduces the predicted value of $\Delta E_4$ by as much as 20 kcal/mol, with the effect being the greatest for the larger basis sets. Even so, $\Delta E_4$ is still found to be higher than either $\Delta E_1$ or $\Delta E_3$ at all correlated levels except MP2/STO-3G. The barriers predicted with MP4 are only slightly different from those predicted with MP2 for a given basis set. At the 6-31G(d) geometries, all correlated calculations find the concerted mechanism to require about 4 kcal/mol more energy than the stepwise mechanism. Use of a minimal basis (STO-3G or AM1) geometry raises this difference to 6-8 kcal/mol. It is gratifying that most levels of theory investigated here are in reasonable agreement on this point [MP4/6-311G(d,p)//C and MP2/6-31G(d)//A favor the stepwise mechanism by 3.3 and 7.5 kcal/mol, respectively].

Of particular interest with regard to the next subsection is the following comparison: The values of $\Delta E_1$, $\Delta E_3$, and $\Delta E_6$ predicted by MP2/6-31G(d)//AM1 are 44.6, 46.2, and 52.7 kcal/mol, respectively, while MP2/6-31G(d)//STO-3G predicts 41.3, 40.6, and 48.8 for the same three barrier heights. These may be compared with the predictions of 41.4, 39.7, and 44.7 at the highest level of theory performed here, MP4/6-311G(d,p)//6-31G(d). The agreement between each of the two less time-consuming levels of theory with the latter is quite good, particularly when STO-3G geometries are used. All three levels of theory predict that the stepwise mechanism has a smaller overall energy requirement than the concerted mechanism, but only by a few kcal/mol.

Glycylglycine. The barriers and relative energies for the gly-
Figure 9. Plots similar to those in Figure 7 for TS3, except that plot d represents an LMO that is predominantly O₃ lone pair.

cylglycine system are listed in Tables VI and VII, respectively. The overall trends found for the dipeptide are similar to those discussed above for the model system: There are significant changes as a result of both basis set improvement and the addition of correlation corrections. On the basis of the detailed comparisons presented for the model system, one expects that barrier heights obtained at the highest levels of theory used for the dipeptide will be at least qualitatively similar to those predicted by MP4/6-311G(d,p)//RHF/6-31G(d) for the model system. The values of ΔE₁, ΔE₃, and ΔE₆ predicted by MP2/6-31G(d)//AM1 are 43.4, 52.7, and 52.7 kcal/mol, respectively, while MP2/6-31G-(d)//STO-3G predicts 40.5, 44.7, and 47.7 for the same barrier heights. The barrier heights predicted for glycylglycine are very similar to those found for the model system. Notice in particular that MP2/6-31G(d)//A favors the stepwise process (ΔE₁ > ΔE₃) by 3.0 kcal/mol. This is essentially what was found for the model system with the highest level of theory. On the basis of these results, it may be concluded that, as found for the model system, the two alternative mechanisms have similar overall energy requirements. This suggests that the essence of the energetics for dipeptide bond formation is largely unaffected by the ancillary groups not directly involved in the bond-making and bond-breaking. This observation provides some justification for the previous calculations on model systems, as well as some impetus for performing similar time-saving calculations for more elaborate amino acids.

Thermodynamics. Table VII lists the energies of the products and intermediates relative to the reactants for both the model system and the glycylglycine system. The latter has two (conformationally) distinct products, and two energy differences must be considered. The results for the model system and reaction 7 are very similar and both can be represented schematically by Figure 10. First the model system is considered and the RHF/6-31G(d) results are taken as the reference. It can be seen
still being favored by a few kcal/mol. The change is mainly due to a 2-4 kcal/mol decrease for $\Delta H_f$ and very little change in $\Delta G_f$. Also, $\Delta H_f$ increases slightly so that at the highest level of theory the stepwise mechanism is still preferred, by 2.4 kcal/mol. MP2/6-31G(d)//A continues to overestimate the gap by about 4 kcal/mol. The same trends are evident for the dipeptide, such that the gap between the energy requirements for the stepwise and concerted mechanisms increases by about 3 kcal/mol. While $\Delta H_f$ is slightly smaller than $\Delta H_f$ for the dipeptide at the MP2/6-31G(d)//STO-3G level, these two may well reverse at higher levels of theory, based on the model system.

Given the molecular structures and vibrational frequencies, it is straightforward to calculate the Gibbs free energy at any temperature, using the harmonic oscillator/rigid rotor approximation. This is useful, since it allows one to estimate both temperature and entropy effects on the reaction energetics. The free energy differences at 298 K are listed in Tables VI and VII. The inclusion of entropy increases the (Gibbs free) energies of all structures relative to the reactants. The intermediates increase by roughly 5 kcal/mol more than the TSs, which increase by about 10-12 kcal/mol relative to the electronic energy, and the first step thus becomes 15-17 kcal/mol higher in energy than the second. All the TSs increase by the same amount for all levels and the stepwise mechanism is still favored by no more than 3 kcal/mol at the highest level. MP2/6-31G(d)//A now overestimates this difference by only 2.1 kcal/mol since $\Delta G_f$ increases 2 kcal/mol more than $\Delta G_f$. AM1 continues to overestimate the barrier for the concerted mechanism. For the dipeptide system the gap widens a bit (by 3 kcal/mol for MP2/6-31G(d)///A) and if one considers the 2.1 kcal/mol overestimation observed in the model system, the difference drops to around 6 kcal/mol.

D. Effect of Correlation on the Molecular Structure. It is evident from the previous discussion that correlation has a significant effect on the calculated barriers. The possibility that the calculated geometries may be equally affected has been investigated. The principal concern is whether the bonding description of the three transition states outlined in section B changes qualitatively. All previously discussed stationary points in reaction 5 were optimized at the MP2/6-31G(d) level of theory, and the resulting key geometric parameters are listed in Tables I-V. In addition, the glycine zwitterion was also optimized to provide reference C-N and N-H equilibrium bond distances (1.507 and 1.024 Å, respectively).

First, consider TS1. One finds that all the key bond distances and their associated equilibrium bond distances considered in Figure 7, increase relative to the SCF values. The distance deviations from equilibrium, calculated using MP2, give a bonding description similar to 2. Again, the O-H distance shows a deviation from its equilibrium value that is twice as large as that for N-H (0.400 Å versus 0.200 Å). This seems to indicate that $H_2$ is primarily bonded to the nitrogen. The N-C distance of 1.562 Å (equilibrium value = 1.507 Å) seems to indicate that the N-C bond is essentially formed, while the 1.336 Å C-O distance is indicative of a single bond when compared to the formamide C-O distances of 1.350 Å (single) and 1.212 Å (double). Thus, the proton transfer seems to proceed after the N-C bond is essentially formed, and the bonding picture of TS1 (2) remains similar to the labile $T^+$ intermediate proposed by Jencks and co-workers. Now consider TS2. The C-O distance of 1.836 Å, compared to the equilibrium value of 1.350 Å for a single bond, does not suggest any significant bonding between those two atoms. Judging from the C-O and N-C bond (equilibrium) distances of 1.318 Å (1.350 Å, single; 1.224 Å, double) and 1.358 Å (1.362 Å for formamide) there is some delocalization of charge from the C-O bond into the N-C bond. However, the C-O distance has increased and the O-H distance decreases relative to the SCF geometry, so that the proton transfer from the incoming water is less complete. However, the respective distances of 1.211 and 1.272 Å, in conjunction with the C-O distances, suggest that proton transfer precedes nucleophilic attack in the first step of formamide hydrolysis, in accordance with SCF and experimental results.

Figure 10. Schematic representation of the RHF/6-31G(d) (bold), RHF/STO-3G (solid), and AM1 (dashed) PES for (a) the stepwise and (b) concerted mechanism.
Next, consider TS3. As in TS1, the N–C bond is largely formed, although the deviation from equilibrium is larger for MP2 (0.070 Å) than for SCF (0.020 Å). H₂ appears primarily bonded to N considering N–H = 0.72 Å (equilibrium) distances of 1.200 Å (1.204 Å) and 1.329 Å (0.980 Å), respectively. Although the gap between bond and equilibrium bond distance for O=O has decreased relative to SCF (0.462 Å versus 0.354 Å), 1.812 Å seems unreasonably large for any significant bonded interaction between O₃ and C. Thus the conclusions regarding TS3, and indeed for TS1 and TS2, in section B remain consistent with results obtained with correlated wave functions.

The energetics of both the stepwise and concerted mechanisms are essentially unaffected by the changes in geometry. The MP2/6-31G(d) values are 1.6 kcal/mol for ΔEₚ. The latter remains the highest of the three barriers considered in Table VI.

IV. Conclusions

Two mechanisms of peptide bond formation are considered in this study, a stepwise mechanism proceeding through a tetrahedral intermediate (Scheme I) and a concerted mechanism (Scheme II). These mechanisms are explored in two systems, a model system (5) leading to the formation of formamide and an extended system (7) leading to the formation of the dipeptide glycylglycine. The model system is studied extensively using high levels of theory to establish how well the model system represents the actual extended system. The main conclusions that may be drawn from this work are as follows:

1. In the first step of the stepwise mechanism, C–N bond formation precedes proton transfer, in accordance with the experimental results for acidic aqueous solutions.

2. In the second step of the stepwise mechanism, a proton is transferred after the leaving group has departed in the dehydration reaction. This is equivalent to nucleophilic attack preceded by proton transfer; i.e., an S-like transition state, in the reverse reaction, hydration.

3. The mechanism for the concerted reaction is essentially a composite of points 1 and 2.

4. The overall reaction is predicted to be slightly endothermic.

5. While the first step in the stepwise process seems to require slightly more energy than the second, the difference is too small to distinguish the two steps with certainty.

6. The stepwise process requires slightly less energy than the concerted mechanism, but the two are too close to call.

7. The inclusion of entropy effects slightly widens the difference in energy requirements for concerted versus stepwise.

8. The model system appears to be a good representative of the actual dippeptide system. In particular, the geometric parameters common to (5) and (7) are very similar, and the barriers and relative energetics for (5) and (7) are very close. Thus one can use the results obtained with the highest level of theory for (5) to analyze its mechanism and be fairly confident that they apply to (7) as well.

9. The semiempirical AM1 method predicts transition state geometries that are in better agreement with 6-31G(d) structures than STO-3G, because STO-3G greatly overestimates the stability of the intermediates relative to the reactants (Figure 10).

10. The MP2/6-31G(d)/6-31G(d) energetics appear to be well-represented by both MP2/6-31G(d)/AM1 and MP2/6-31G(d)/STO-3G calculations on (5). The latter two methods are sufficiently modest that they may be used on (7).

11. While points 1–3 are partially based on SCF electron densities, they are also supported by geometries obtained with correlated (MP2) wave functions.

Acknowledgment. This work was supported in part by a grant from the donors of the Petroleum Research Fund, administered by the American Chemical Society. The calculations were performed on the North Dakota State University IBM 3090/200E, obtained in part by a Joint Study Agreement with IBM, on the Cray Y-MP at the San Diego Supercomputer Center, on a VAXstation 3200, and on an IBM RS/6000/530 obtained through a grant from the National Science Foundation. Dr. Thomas R. Cundari and Theresa L. Windsch are thanked for their careful reading of the manuscript.

Supplementary Material Available: Tables of total energies of molecules used and complete geometries (17 pages). Ordering information is given on any current masthead page.

References and Notes


(7) (a) Reference 6a, p. 939. (b) Reference 6b, p. 411.

(8) Reference 6b, p. 431–432.


(17) Stewart, J. J. P. QCPE, program 455.

(18) The option to add a molecular mechanics correction that increases the barrier to rotation around a peptide bond is not used.


(22) A nonlinear least-squares gradient minimizer that minimizes gradients using Barlett's method. See MOPAC manual and source code for further information.
On the Extraordinary Spectral Similarity of Nickel(II) Phthalocyanine and NiII(C11N7H2)2

Shinichiro Nakamura,*t Alberto Flamini,*t Vincenzo Fares,1 and Masafumi Adachi1

Research Center, Mitsubishi Kasei Corporation, Kamoshida-cho, Midori-ku, Yokohama 227, Japan, and Institute of Theory and Electronic Structure of Coordination Compound, CNR, Research Area of Rome, P.O. Box 10, 00016 Monterotondo Stazione, Rome, Italy (Received: June 12, 1992)

INDO/S type MO calculations have been carried out on nickel(II) phthalocyanine (NiPc) and bis(β-diminotetracyano­pyrrolo­zino)nickel(II), of formula Ni(C11N7H2)2, in order to account for the extraordinary similarity of their optical spectra. SCF calculations have been performed on the one-electron approximation at the closed shell Hartree-Fock level; extensive configuration interaction arising from single excitations has been evaluated. The calculated electronic state picture is as follows. In both systems the ground-state \( ^1\text{A}^2_u \) Mulliken population analysis indicates a formal Ni(II) character. In NiPc, four frontier orbitals, two \( \pi^* \) LUMO (\( \varepsilon_{\pi^*} \) in \( D_{2g} \) and \( b_2 \) in \( D_{2d} \)) and two \( \pi \) HOMO (\( \varepsilon_{\pi} \) in \( D_{2g} \) or \( b_2 \) in \( D_{2d} \)), consistent with previously reported extended Hückel calculations, account adequately for the characteristic Q (observed, 1.495 \( \mu \text{m}^{-1} \); calculated, 1.190 \( \mu \text{m}^{-1} \)) and B (observed, 3.268 \( \mu \text{m}^{-1} \); calculated, 3.112 \( \mu \text{m}^{-1} \)) bands, slightly mixed with each other by a small configuration interaction. In Ni(C11N7H2)2 there are four frontier orbitals, similar to those of NiPc (they are \( \pi \) ligand orbitals, located separately from other orbitals), but a switchover in energy between \( a_h \) and \( b_2 \) occurs and both orbitals of the two pairs (LUMO and HOMO) are close in energy, so that only two nearly degenerate transitions between them are allowed by symmetry; such intraligand transitions are of the same symmetry (\( ^1B_{2u} \)), \( b_{1u} \rightarrow b_{2u} \) and \( b_{2u} \rightarrow a_h \), both polarized along the shortest molecular axis (\( y \)), associated with additive large values of dipole moment and giving rise to the observed strong Q-like band (observed, 1.605 \( \mu \text{m}^{-1} \); calculated, 1.510 \( \mu \text{m}^{-1} \)). The B-like band (observed, 2.451 \( \mu \text{m}^{-1} \); calculated, 2.305 \( \mu \text{m}^{-1} \)), which is very little mixed with the Q-like band, given the observed relative intensity ratio, is tentatively assigned to a transition from lower \( \pi \) orbitals leading to a final state of \( 1B_{2u} \) symmetry. The polarization of the Q band, which is responsible for the intense blue color of complexes 1, has been observed under a polarizing microscope for the analogous zinc complex (Zn(C11N7H2)2·4THF), and it is in qualitative agreement with theoretical findings.

Introduction

Some of us have recently discovered a new class of dyes,1 the bis(β-diminotetracyano­pyrrolo­zino)metal(II) complexes, M(β-dtpy)2 (1), resulting from a metal-assisted intramolecular cyclization of the pyrroline anion (2).

Surprisingly, the absorption spectra of (1) show an extraordinary similarity with those of the corresponding metal phthalocy­anines (MPCs), characterized by (i) a sharp and very intense low energy band (Q band), accompanied by a vibronic progression of lower intensity to higher energy, (ii) a characteristic window in the 2.0-\( \mu \text{m}^{-1} \) region, and (iii) a highest energy band (B or Soret band), displaying poorly resolved shoulders. As an example, previously reported optical spectra of Ni(β-dtpy)2 and of NiPc are displayed in Figure 1.

Given the considerable interest for porphyrin-like complexes, including MPCs, as a dyestuff or catalyst or in biological applications,5 the electronic structure of such systems has been thoroughly studied by a semiepi­mical MO method,3,4 which successfully explained the characteristics of the absorption spectra, due to single \( \pi \rightarrow \pi^* \) transitions of \( \varepsilon_{\pi^*} \) symmetry, arising primarily from the macrocyclic dianion, as being based on an earlier four-orbital model.5 Recently, more semiepi­mical as well as Xα studies are reported.6

The overall similarity of the absorption spectra of the two systems, irrespective of their apparent difference in geometry and symmetry (1 has \( D_{2g} \) symmetry, MPC \( D_{4h} \) symmetry), is very interesting and provides an unique opportunity to study the electronic structure from a purely spectroscopic point of view as well as to obtain valuable information utilizable as guiding.