

Ab initio investigations of dipeptide structures

C. Dale Keefe*, Jason K. Pearson

Department of Physical and Applied Sciences, University College of Cape Breton, Sydney, Nova Scotia, B1P 6L2 Sydney, Canada

Received 9 January 2004; revised 9 January 2004; accepted 22 March 2004

Abstract

Finding a solution to the protein folding problem has been steadily pursued for decades. Unlocking the mysteries of the three-dimensional structure of these large and very important biological molecules could bring with it revolutionary methods of drug development among countless unforeseen benefits. In this study, ab initio methods are used to examine precise structural features of the peptide bond in a series of dipeptides in an effort to clarify the effects of local interactions in a peptide system. Bond lengths, angles, dihedral angles, and rotation barriers about the peptide bond are presented for a series of dipeptides with a discussion of the results.

© 2004 Elsevier B.V. All rights reserved.

Keywords: Dipeptide structure; Ab initio calculations; Peptide bond; Bond lengths; Bond angles

1. Introduction

Proteins are polymers of the 20 naturally occurring amino acids and are used for structural support, storage, transport of other substances, signaling from one part of the organism to another, movement, and defense against foreign substances among other things [1]. The range of the number of amino acid residues in a protein is typically 100–1800 while there have been proteins discovered with as many as 27,000 residues (titin, a muscle protein) [2]. The function of these proteins is dependant to a large extent upon their three-dimensional structure, [1] which is almost exclusively dependant on the linear sequence of amino acids in the protein (primary structure) [3]. This implies that inherent in the primary structure of a protein is all of the information required to decipher its three-dimensional structure [4] and consequently its function in the human body. It is, therefore, of great interest to determine the factors, which resolve the three-dimensional structure from the primary structure. The prediction of a protein's three-dimensional structure from the knowledge of only the linear sequence of amino acids has been described as the determination of the second half of the genetic code [5]. Because of this importance, this challenge has been steadily pursued for over a decade, however, there is still debate over the dominant factors of

protein folding; the process by which the three-dimensional structure of a protein is formed from the primary structure [6]. Understanding the intricacies of protein folding still presents a daunting task for the scientific community and a complete understanding at the atomic level is not likely in the near future [7].

Peptide structures have been investigated with ab initio calculations for almost three decades. For a review of the work up to 2001, the reader is referred to the paper by Csizmadia and co-workers [7]. Initial studies [8–17] were limited by technology to rigid geometries of diamides and dipeptides, which were shown to be insufficient during the early 1980s [18–28]. Recent papers [25,29–35] have studied the structures of diamides, dipeptides and other short chain peptides. These studies have limited the potential energy to functions of only the torsional angles within the amino acid residues by way of the so-called Ramachandran map [36,37]. To the best of our knowledge, no one has considered how the other geometrical parameters affect the energy and structure of peptides.

The focus of this study is to investigate numerous parameters involved in dipeptide structure prediction which is part of 'Phase One' [7] of the computational study of protein structure and the energetics of protein folding. Specifically, the planarity of the peptide bond is investigated in a series of dipeptides and internal barriers to rotation in alanine are reported. The dihedral angles of the amide *plane*, the geometry about the α -carbon atoms, the bond lengths,

* Corresponding author. Tel.: +1-902-563-1185; fax: +1-902-563-1880.
E-mail address: dale_keefe@ucb.ca (C. Dale Keefe).

and the bond angles of the amide *plane* in a peptide structure are also examined.

Using *ab initio* methods it is possible to probe the structure of dipeptides and examine structural features for the purpose of supporting or refuting current theories toward protein structure prediction. For example, the use of only two parameters per amino acid residue, ψ and ϕ , which are the dihedral angles defined in Fig. 1, as the only relevant parameters to consider in protein structure predictions is contingent upon the fact that the amide *plane* is in fact planar. By observing relevant dihedral angles in the optimized structures of small amino acid sequences, it can be determined if such planarity is actually expected. The use of ψ 's and ϕ 's as the only relevant parameters to consider in protein structure predictions is also contingent on the geometry of each α -carbon. While it is not expected that the α -carbon will have perfect tetrahedral geometry due to the fact that it is stereogenic, it is assumed that its bond angles throughout each amino acid in a sequence should be similar; if this is not the case it will play an important role in the final structure of the protein in addition to ψ and ϕ .

To investigate the structural parameters of small peptide molecules, dipeptides were constructed and their geometries were optimized at the Hartree–Fock level of theory with the 6-31G* basis set. The selection of dipeptides was based on earlier geometry optimizations of all 20 amino acid monomers. Eight structures were chosen from the 400 possible structures which were the smallest, optimized the fastest, and could be successfully completed on the Pentium® (Pentium is a trademark of Intel Corporation) and the Sun Blade™ (Sun Blade is a registered trademark of Sun Microsystems Inc.) 1000 systems available in this laboratory.

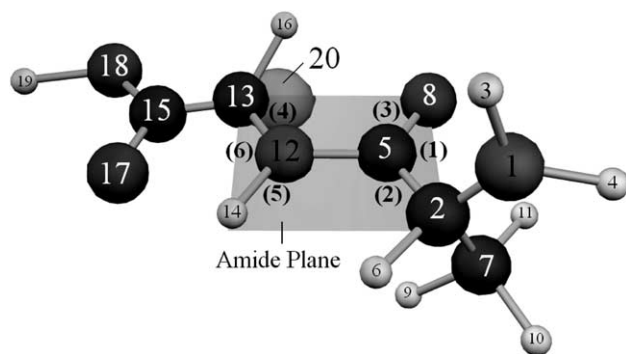


Fig. 1. General structure and numbering scheme of the dipeptides studied. The numbers in parenthesis indicate the numbering of the amide in-plane bond angles. Atoms 3, 4, 6, 9, 10, 11, 14, 16 and 19 are hydrogens; 1 and 12 nitrogens; 2, 5, 7, 13 and 15 carbons; 8, 17 and 18 oxygens and 20 represents the 1st atom of the R group. ϕ is the dihedral angle of atoms 5 and 15 about bond 12–13 and ψ is the dihedral angle of atoms 1 and 12 about bond 2–5. Carbons 2 and 13 are the α carbons.

2. Results and discussion

All calculations performed for this work were completed using the GAUSSIAN 98 (Revision A.11.4) [38] software package. The computer systems that were utilized include a Pentium® III 900 MHz with 1.2 GB of RAM, a Pentium® IV 2.0 GHz with 256 MB of RAM, and a Sun Blade™ 1000 system with 512 MB of RAM. Geometry optimizations on each of the 20 amino acids took from 18 min to 7 h 27 min each. For the dipeptides, the jobs ranged from 2 h 35 min to 8 h 55 min.

The dipeptides were constructed by keeping alanine fixed at the N-terminus position and varying the amino acid that was bonded to it. The varying C-terminus position is hereafter referred to as the ‘B’ position. Successively, the B position was filled with one of the eight chosen amino acids: glycine, alanine, serine, threonine, asparagine, aspartic acid, valine or cysteine. All of these residues were assumed to be neutral in the calculations. Once the dipeptide structures were optimized, the structural parameters were studied¹. The numbering scheme is shown in Fig. 1 and will be referred to throughout the remainder of this paper.

2.1. Bond lengths and bond angles

With the eight optimized structures obtained, the five bond lengths and six bond angles of the amide *plane* were examined. Shown in Tables 1 and 2, are the calculated bond lengths and angles, respectively, for the eight dipeptides studied. Table 1 indicates the atom type and number for both atoms in each bond and lists those bond lengths (in Å) for each case as the B group was varied. Table 2 is similar but identifies the angles (presented in degrees) with respect to the numbering scheme shown in Fig. 1. The maximum deviation in any of the bond lengths is 0.008 Å indicating very little variance in the bond lengths as the B group changes. The maximum deviation in the bond angles of the amide *plane* is 0.6° with the exception of bond angles five and six, which have maximum deviations of 1.9 and 1.3°, respectively. An explanation for this observation is offered in Section 2.3. When considering the bond lengths of the amide *plane*, it is clear that there is very little difference in any bond as the B group changes, indicating that these bond lengths are essentially fixed. Similar results are seen in the case of the bond angles of the amide *plane*. With the exception of angles five and six the geometry of the *plane* appears fixed.

2.2. α -Carbon geometry

The geometries about the α -carbons are important factors to consider as they can play a crucial role in the overall structure of a protein if they vary significantly throughout

¹ The optimized structures of the twenty amino acids and eight dipeptides are available on CDK's website: <http://faculty.uccb.ns.ca/dkeefe/research>.

Table 1
The calculated bond lengths of the amide plane for the eight dipeptides studied

| B amino acid | Bond length/Å | | | | |
|-------------------|---------------|-------|--------|---------|---------|
| | C2–C5 | C5–O8 | C5–N12 | N12–C13 | N12–H14 |
| Ala | 1.529 | 1.203 | 1.349 | 1.439 | 0.996 |
| Asp | 1.530 | 1.203 | 1.352 | 1.435 | 0.998 |
| Asn | 1.530 | 1.204 | 1.350 | 1.434 | 0.998 |
| Cys | 1.530 | 1.205 | 1.346 | 1.438 | 0.998 |
| Gly | 1.530 | 1.204 | 1.343 | 1.434 | 0.996 |
| Ser | 1.529 | 1.203 | 1.352 | 1.433 | 0.997 |
| Thr | 1.530 | 1.203 | 1.352 | 1.433 | 0.997 |
| Val | 1.532 | 1.205 | 1.344 | 1.444 | 0.997 |
| Average | 1.530 | 1.204 | 1.349 | 1.436 | 0.997 |
| Maximum deviation | 0.002 | 0.001 | 0.006 | 0.008 | 0.001 |

The numbering corresponds to that shown in Fig. 1.

a series of amino acid residues. Even slight deviations can sum to very large ones when considering that a protein could consist of thousands of residues. Ideally the bond angles about an sp^3 hybridized carbon would be 109.5° . This is not expected in these cases because of the stereogenic nature of the α -carbon atoms. However, what becomes important is the consistency of the α -carbons bond angles as the B group is varied. Significant changes in these angles would suggest that the α -carbons do not retain the same geometry and as such would need to be considered in larger protein structure predictions. These angles were measured for both residues of the dipeptide with respect to the R group of each. In other words, the angles formed between the first atom of the R group and each of the other three substituents on the α -carbon were examined. The deviations from the ideal 109.5° were noted as well as the variations with respect to varying the B groups. The results are given in Table 3, where all data is presented in degrees. The left-hand portion of Table 3 refers to the alanine residues of the dipeptides, which was kept at the N-terminus position throughout the entire study and the right-hand portion refers to the B amino acid residues.

Table 2
The calculated bond angles of the amide plane for the eight dipeptides studied

| B amino acid | Bond angle/degree | | | | | |
|-------------------|-------------------|---------|---------|---------|---------|---------|
| | (1) CCO | (2) CCN | (3) OCN | (4) CNC | (5) CNH | (6) HNC |
| Ala | 121.3 | 116.0 | 122.7 | 120.7 | 116.9 | 120.0 |
| Asp | 121.0 | 115.6 | 123.4 | 121.2 | 116.8 | 118.4 |
| Asn | 120.9 | 115.6 | 123.5 | 121.3 | 117.3 | 118.0 |
| Cys | 121.2 | 116.0 | 122.8 | 121.2 | 118.7 | 119.2 |
| Gly | 121.3 | 116.0 | 122.9 | 120.8 | 119.6 | 118.9 |
| Ser | 121.1 | 115.8 | 123.0 | 120.5 | 117.0 | 117.9 |
| Thr | 121.1 | 115.8 | 123.1 | 120.6 | 116.8 | 117.9 |
| Val | 121.0 | 115.8 | 123.2 | 121.6 | 118.5 | 119.4 |
| Average | 121.1 | 115.8 | 123.1 | 121.0 | 117.7 | 118.7 |
| Maximum deviation | 0.2 | 0.2 | 0.4 | 0.6 | 1.9 | 1.3 |

The numbers in parentheses correspond to the angles shown in Fig. 1.

The range of the angles for the alanine residue is very small. This suggests that the geometry around that α -carbon does not change significantly as its neighbor residue is varied. This was expected because the effect that groups have over a distance of four bonds away are typically small in any situation. The geometry about the α -carbon atom of the B group, however, does vary. There are ranges of 5, 3, and 3° between the minimum and maximum bond angles in the three cases listed indicating that the geometry around the α -carbon at a particular position changes with respect to the amino acid occupying that position. Since the results presented in Table 3 suggest that the geometry about the α -carbon atoms are not retained throughout an amino acid sequence, this factor must be taken into account when considering larger peptides. The geometry of the α -carbon will add extra degrees of freedom per amino acid residue to the determination of the overall protein structure.

Another interesting point to note in Table 3 is the geometry of the two alanine α -carbon atoms when they are both present in the dipeptide. This comparison reveals differences in bond angles about the α -carbon atoms of the two positions in the dipeptide despite the fact that both positions are occupied by the same amino acid residue. These differences can be attributed to the fact that there is a significant difference between the N of the amine group for the first amino acid and the N of the *plane* (formerly of an amine group) of the B amino acid and similarly with the carboxyl carbons.

These results suggest that it may be possible to build a database of angles for the various amino acids when they occupy different positions within small peptide structures and that they will not need to be part of the optimization for larger peptides.

2.3. Dihedral angles

Investigating dihedral angles in the dipeptide can provide valuable information regarding the peptide bond and the planarity of the amide *plane*. Specifically the dihedral angle

Table 3
Results calculated for the α -carbon bond angles in both residues of each dipeptide studied

| B amino acid | Bond angle/degree | | | | | |
|-------------------|--------------------------|--------------------------|------------------------------|-----------------------------|-----------------------------|---------------------------------|
| | Alanine | | | B amino acid | | |
| | CC $_{\alpha}$ N (7–2–1) | CC $_{\alpha}$ H (7–2–6) | CC $_{\alpha}$ plane (7–2–5) | RC $_{\alpha}$ C (20–13–15) | RC $_{\alpha}$ H (20–13–16) | RC $_{\alpha}$ plane (20–13–12) |
| Ala | 114.8 | 108.7 | 110.5 | 109.2 | 109.4 | 110.3 |
| Asp | 114.8 | 108.7 | 110.8 | 113.4 | 107.5 | 112.5 |
| Asn | 114.6 | 108.8 | 110.8 | 113.5 | 107.6 | 112.4 |
| Cys | 114.8 | 108.7 | 110.6 | 112.1 | 106.4 | 113.0 |
| Gly | 114.8 | 108.5 | 110.6 | 108.5 | 106.3 | 112.1 |
| Ser | 114.8 | 108.7 | 110.6 | 111.1 | 108.6 | 110.5 |
| Thr | 114.8 | 108.7 | 110.6 | 110.8 | 109.1 | 111.4 |
| Val | 114.6 | 108.5 | 110.6 | 113.0 | 108.8 | 112.0 |
| Average | 114.8 | 108.7 | 110.6 | 111.5 | 108.0 | 111.8 |
| Maximum Deviation | 0.2 | 0.2 | 0.2 | 3.0 | 1.7 | 1.5 |

The numbers in parentheses correspond to the atom numbers in Fig. 1.

between atom 13 and atom 14 of the amide *plane* with respect to the peptide bond (bond between atoms 5 and 12) can be considered and will be referred to as D_N hereafter. This dihedral angle should be 180° if indeed the amide *plane* is planar.

Other dihedral angles that were considered include: the angle between atom 12 and atom 8 of the amide *plane* with respect to the bond joining atoms 2 and 5 (referred to as Db1), the angle between atoms 2 and 13 with respect to the peptide bond (referred to as Db2), and the angle between atoms 2 and 14 with respect to the peptide bond (referred to as Db3). In a planar structure these angles should be 180° , 180° , and 0° , respectively.

The dihedral angles considered in this study are presented in Table 4. The angles are labeled as described above and in brackets are the values that these angles would have if the structure was planar. Clearly, none of the dihedral angles correspond to a planar amide *plane* with the exception of Db3 in the case of glycine in the B position. All other dihedrals exhibit deviations from their expected values with the largest differences seen in D_N . Deviations of almost 25° (refer to Table 5 for a list of the deviations from 180° in this parameter) occur in this dihedral angle, which obviously suggests that the geometry about the amide *plane* nitrogen (atom 12) is not planar. In addition to the deviations from 180° of D_N , Table 5 also presents ϕ for the peptide bond joining alanine with each of the listed B amino acids.

The R groups are presented in Table 5 to illustrate the trend that with larger R groups there are larger values of ϕ . It is reasonable to assume that this trend is due to steric interferences between the atoms of the R group and the hydrogen of the amide *plane* (atom 14). The trend then means that alanine in the B position will have its R group (methyl) just underneath the amide *plane* hydrogen while cysteine in the B position will have its much larger R group rotated farther away. This demonstrates a clear link between the side group R and the value of ϕ . This also provides

justification for the variations in the bond angles discussed in Section 2.1. There is, however, no such trend in the deviation of D_N from 180° . The R group size does not seem to directly affect the planarity of the structure.

Glycine, while it does not break the overall trend, exhibits a much larger decrease in ϕ . This can be explained by the much larger difference in size as you step down from a methyl group to a hydrogen atom as compared to differences of the other B amino acid R groups. The major

Table 4
The calculated dihedral angles of the amide *plane* for the eight dipeptides studied

| B amino acid | Dihedral angle/degree | | | |
|--------------|-----------------------|-----------|---------|-------------|
| | Db1 (180) | Db2 (180) | Db3 (0) | D_N (180) |
| Ala | 181.3 | 171.5 | 9.1 | 162.5 |
| Asp | 181.0 | 166.9 | 8.7 | 157.8 |
| Asn | 180.9 | 167.0 | 8.1 | 159.1 |
| Cys | 181.4 | 175.0 | 5.7 | 168.8 |
| Gly | 181.9 | 176.6 | 0.0 | 176.3 |
| Ser | 180.7 | 166.1 | 10.8 | 155.2 |
| Thr | 180.9 | 165.8 | 10.6 | 155.3 |
| Val | 181.9 | 174.2 | 2.5 | 171.7 |

The labels correspond to that described in Section 2.3.

Table 5
Calculated deviations from 180° in D_N and the corresponding value for ϕ in the eight dipeptides studied

| B amino acid | R | D_N deviation from 180° | ϕ (degrees) |
|--------------|-------------------|----------------------------------|------------------|
| Gly | H | –3.7 | –79.9 |
| Ala | CH $_3$ | –17.5 | 67.8 |
| Ser | CH $_2$ OH | –24.8 | 81.4 |
| Thr | CHCH $_3$ OH | –24.7 | 83.7 |
| Asp | CH $_2$ COOH | –22.2 | 104.3 |
| Asn | CH $_2$ CONH $_2$ | –20.9 | 104.2 |
| Val | CH(CH $_3$) $_2$ | –8.3 | 151.5 |
| Cys | CH $_2$ SH | –11.2 | 158.0 |

consideration here is the steric interaction of local species, or those directly bonded to the atom that is connected to the α -carbon. There is no other case where such a sharp increase in steric hindrance occurs. This large difference in steric effects for glycine actually places the R group hydrogen in line with the amide plane, which is the best position to minimize the interference of other groups in physical space.

It appears as though there also could be hydrogen bond-like attractive forces between the amide plane hydrogen (atom 14) and the carboxyl oxygen atoms (atoms 17² and 18) as well as between the amide plane oxygen (atom 8) and the hydrogen(s) of the B amino acid α -carbon in all cases. The inter-atomic distances of these species are in the range of 2.3–2.7 Å, which is consistent with weak hydrogen bonding in peptide systems [39].

2.4. Barriers to rotation in alanine

Also studied in this work were the internal barriers to rotation in alanine. Rotation of the R group (methyl) will provide insight into the freedom that a small R group has to rotate while in a protein. Rotating the carboxyl group will essentially be modeling the different values of ψ that the molecule can have and the barrier to internal rotation can be used as a simple model of the energy required to rotate about the α -carbon and carboxyl carbon bond. It is assumed that alanine will have the smallest barrier to internal rotation of the R group as compared with other amino acids (with the exception of glycine, having a hydrogen atom for an R group). Before the energy barrier was calculated the structure of alanine was optimized and then a rigid potential energy surface scan was performed by rotating the methyl group between 0 and 360° in increments of 5° (dihedral angle of atoms 10 and 1 with respect to bond 2–7). A rigid potential energy surface scan was also performed for the carboxyl group rotation between 0 and 360° in increments of 5° (dihedral angle of atoms 8 and 1 about bond 2–5).

The energy curves for both the methyl and carboxyl group rotations in alanine are presented in Figs. 2 and 3. Fig. 2 shows the expected three-fold symmetry of the methyl group, with each peak corresponding to the eclipsed conformer and each well corresponding to the staggered conformer. The barrier to this rotation was 17.1 kJ mol⁻¹, which suggests that this rotation is significantly hindered. This relatively large barrier indicates that rotation of the methyl group in alanine is very restricted and this interpretation can be extended to assume that rotation of any other R group (with the exception of glycine) would be restricted as well, if not more so than in alanine. This is because with the exception of glycine, every other amino acid R group is larger than methyl and thus there would be

² Atom 17 would likely play the major role in any hydrogen bonding effects because of the added electron density although at low pH both atoms 17 and 18 are equivalent.

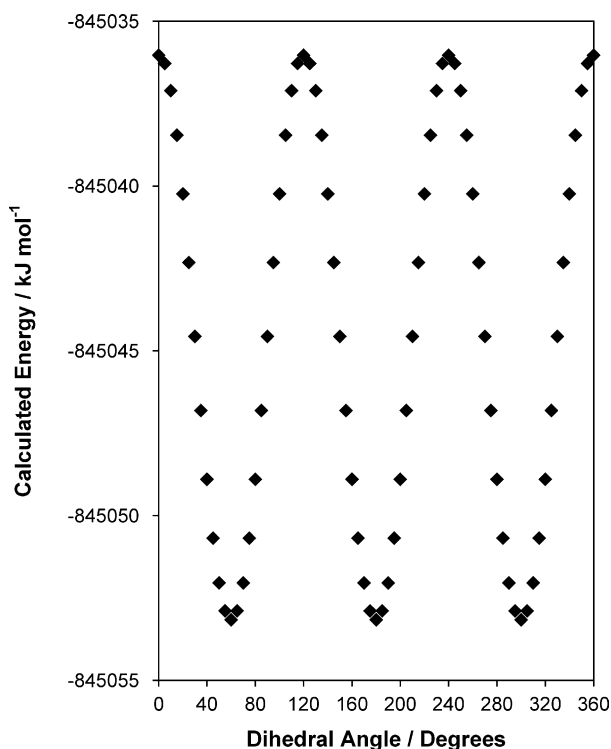


Fig. 2. Energy curve for the methyl group rotation in alanine.

an increase in steric interference upon rotation, and a subsequent increase in the energy barrier would be expected.

The energy curve for the rotation of the carboxyl group of alanine (Fig. 3) exhibits two energy wells which

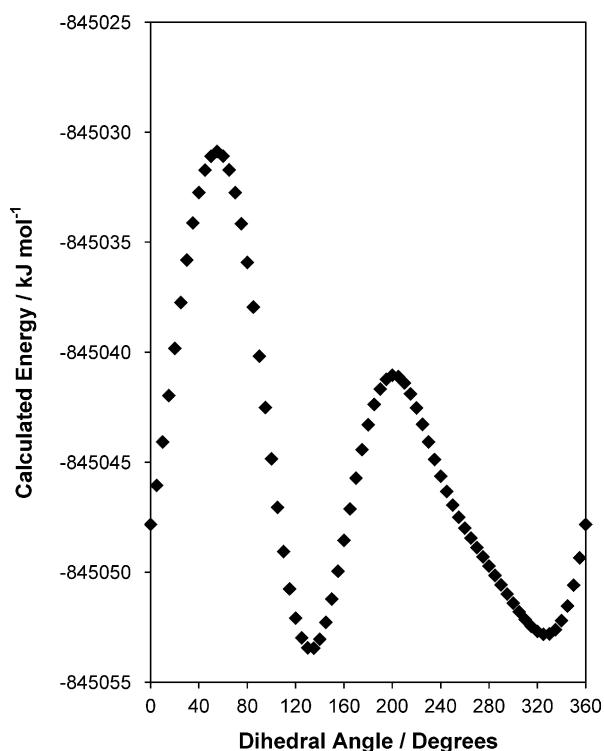


Fig. 3. Energy curve for the carboxyl group rotation of alanine.

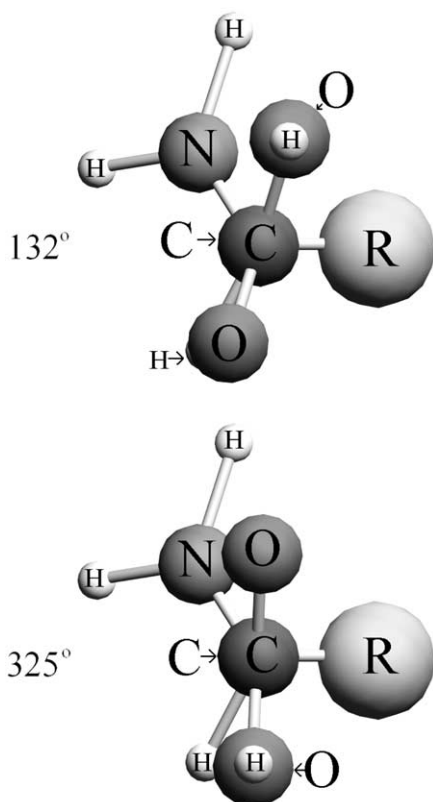


Fig. 4. The two positions of the carboxyl group in alanine that correspond to the energy minima. The 132° conformer is slightly lower in energy (0.19 kJ mol^{-1}).

correspond to 132 and 325° with the 132° rotation being slightly lower in energy. Fig. 4 shows graphically the positions of the carboxyl group in these wells. There are two barriers to rotation for this particular group as seen in Fig. 3. They are 22.6 and 12.4 kJ mol^{-1} . The larger barrier is associated with rotating the hydroxyl part of the carboxyl group past the methyl group and can be attributed to steric effects, since the methyl group is more bulky than the amine group. Again it can be seen that the barrier is relatively large. This suggests that rotation of the carboxyl group is restricted. As previously described, this rotation was a simple model of the rotation of ψ in larger peptide systems. Since the barrier indicates that this rotation may be constrained, it suggests that ψ would also be constrained, especially when considering the fact that in larger peptide systems the hydroxyl part of the carboxyl group would be even larger—it would be the joining point for the next residue.

This data is a demonstration of the rigidity of a peptide backbone. A limitation in the ability of ψ to vary would imply that the protein structure is essentially fixed. This rigidity supports the idea that a protein's function is closely related to its shape; it would not be desirable for a protein to easily change conformations if its function depended on a particular shape.

3. Conclusions

3.1. Interpretations

The various structural parameters studied in this paper provide valuable insight into the structural properties of small amino acid sequences and perhaps ultimately to protein structure in general. From the data provided it can be determined that the bond lengths and angles of the so called amide *plane* are essentially fixed for various dipeptides. This implies that in larger peptide systems such amide plane parameters need not be considered for structural optimization purposes. However, a condensation of the many degrees of freedom throughout a protein's backbone to just two per amino acid, ψ and ϕ , is questionable. The evidence presented in this paper supports the idea that the amide *plane* is not actually planar and that the geometry about the α -carbon atoms at the corners of the amide plane vary with respect to different amino acid substituents. Such variations place additional degrees of freedom in the scope of parameters that must be considered to reliably predict the path of a protein's backbone. A more detailed study needs to be completed before conclusions can be drawn as to exactly which parameters can be ignored in the optimizations and which are absolutely essential to include.

The rotation barrier study also presents some interesting results. Large barriers in alanine suggest rigidity in a protein's structure, a logical conclusion considering that the efficient function of a protein relies on its structure. This rigidity will be present not only in the backbone but also in the orientation of the residual R groups, as evidenced by the methyl rotation in alanine. A consideration that should be highlighted at this point was the methodology used to perform this portion of the study. During the rotation of both groups, all other parameters were held fixed. Conformational changes in the other atoms, however, may exist as the groups rotate. This could possibly lower the barrier to rotation in these cases. Because of the large barrier in both rotations considered, it is assumed that any lowering in energy from conformational changes would be negligible.

A persistent problem with *ab initio* calculations, particularly for the prediction of protein structure is the occurrence of multiple minima [6,40]. Multiple minima were observed in this study in the case of the dipeptide with glycine in the B position. The observation of the apparent break in the trend discussed in Section 2.4 prompted the search for additional 'optimum' structures. One was found and it was slightly lower in energy (0.19 kJ mol^{-1} lower). This lower energy structure was then taken as the optimum structure and its parameters were used for this paper. The difference in energy of 0.19 kJ mol^{-1} was considered small enough so as not to warrant a search for multiple minima within the other dipeptide structures or monomer structures.

Also, the physical conformations of both glycine dipeptide structures, while different, were not so much so as to change the observed trend (shown in Table 5). It is

important to state this because multiple minima with significantly different structures could lead to the compilation of a table such as Table 5 and a trend that is based on coincidence rather than on scientific merit. For instance, if every dipeptide studied had multiple minima with drastically different structures then the trend which was observed would be applicable to only those structures that were found, which may or may not be the most probable structures. However, the trend was upheld with both structures of the dipeptide containing glycine and it is obvious that glycine is certainly a special case due to its R group. With bigger R groups it is less likely that any multiple minima would have structures so different as to break the observed trend of increasing ϕ values with respect to increasing R group size.

3.2. Future work

There are a number of reasonable paths that this study could take in the future. First, it would be useful to obtain some additional information on the potential of hydrogen-bonding in these structures. While they would almost certainly be weak interactions, a conclusive rather than intuitive interpretation would be advantageous. In terms of the geometry about the α -carbon atoms in these dipeptides, a better study would involve the calculation of similar parameters in a tripeptide. This would afford the opportunity to observe an α -carbon between two peptide bonds as it is in all but two positions for an actual protein. As was seen in Section 2.3 the geometry about the terminal α -carbon atoms, even when they are of the same type (alanine), is different and so it is logical to assume that an α -carbon between two peptides could display a different geometry still.

Also, this work illustrates the potential downfalls of assuming only two degrees of freedom per amino acid for a peptide backbone, ψ and ϕ . A similar study, in which the same molecules are optimized yet only ψ and ϕ are allowed to vary, would likely provide interesting information about the validity of the simplification.

An investigation of environmental effects such as a water solvent would also be interesting to complete. It has been suggested that such effects play a large role in the determination of protein structure [41,42] and they will need to be understood to properly understand how proteins function within a biochemical environment.

Of course, definitive determinations about protein structure are not likely to be made based on the analysis of simple dipeptides. However, such analyses at an atomic level are crucial in the collaborative effort of the scientific community to solve the protein-folding problem. The information gained from the work on small molecules can be extended to larger ones and larger ones still until finally we possess the level of theory and computational power to predict with pinpoint accuracy the structure and perhaps function of any given amino acid sequence.

Acknowledgements

The authors would like to thank the Natural Sciences and Engineering Research Council of Canada for funding of this project. The authors would also like to thank the University College of Cape Breton microelectronics laboratory for donation of computer facilities.

References

- [1] N.A. Campbell, In *Biology*, 4th ed., Cummings Publishing Company, New York, 1996.
- [2] R.H. Garrett, In *Biochemistry*, 2nd ed., Saunders College Publishing, New York, 1999, Chapter 5.
- [3] R.H. Garrett, In *Biochemistry*, Saunders College Publishing, New York, 1999, Chapter 6.
- [4] C. Anfinsen, *The Molecular Basis of Evolution*, Wiley, New York, 1959.
- [5] L.M. Gierasch, J. King (Eds.), *Protein Folding, Deciphering the Second Half of the Genetic Code*, American Association for the Advancement of Science, Washington, 1990.
- [6] C. Hardin, T.V. Pogorelov, Z. Luthey-Schulten, *Curr. Opin. Struct. Biol.* 12 (2002) 176.
- [7] G.A. Chasse, A.M. Rodriguez, M.L. Mak, E. Deretey, A. Perczel, C.P. Sosa, R.D. Enriz, I.G. Csizmadia, *J. Mol. Struct. (Theochem)* 537 (2001) 319.
- [8] R. Ramani, R.J. Boyd, *Int. J. Quantum Chem. Quantum Biol. Symp.* 8 (1981) 117.
- [9] R. Ramani, R.J. Boyd, *Can. J. Chem.* 59 (1981) 3232.
- [10] R.J. Boyd, J.S. Perkyns, R. Ramani, *Can. J. Chem.* 61 (1983) 1082.
- [11] J.A. Ryan, J.L. Whitten, *J. Am. Chem. Soc.* 94 (1972) 2396.
- [12] D. Peters, J. Peters, *J. Mol. Struct.* 109 (1984) 137.
- [13] D. Peters, J. Peters, *J. Mol. Struct.* 68 (1980) 243.
- [14] D. Peters, J. Peters, *J. Mol. Struct.* 53 (1979) 103.
- [15] D. Peters, J. Peters, *J. Mol. Struct.* 69 (1980) 249.
- [16] L.R. Wright, R.F. Boskman, *J. Phys. Chem.* 86 (1982) 3956.
- [17] P.T. van Duijnen, B.T. Thole, *Biopolymers* 21 (1982) 1749.
- [18] L. Schafer, C.V. Alsenoy, J.N. Scardale, *J. Chem. Phys.* 76 (1982) 1439.
- [19] K. Siam, V.J. Klimkowski, C. van Alsenoy, J.D. Ewbank, L. Schafer, *J. Mol. Struct. (Theochem)* 152 (1987) 261.
- [20] K. Siam, S.Q. Kulp, J.D. Ewbank, L. Schafer, C. van Alsenoy, *J. Mol. Struct. (Theochem)* 184 (1989) 143.
- [21] S.J. Weiner, V.C. Singh, T.J. O'Donnell, P.A. Kollman, *J. Am. Chem. Soc.* 106 (1984) 6243.
- [22] A.M. Sapse, L.M. Fugler, D. Cowburn, *Int. J. Quantum Chem. Phys. Lett.* 29 (1986) 1241.
- [23] T.C. Chean, S. Krimm, *J. Mol. Struct. (Theochem)* 188 (1989) 15.
- [24] T.C. Chean, S. Krimm, *J. Mol. Struct.* 193 (1989) 1.
- [25] A. Perczel, J.G. Angyin, M.K. tar, W. Viviani, J.-L. Rivail, J.-F. Marcoccia, I.G. Csizmadia, *J. Am. Chem. Soc.* 113 (1991) 6256.
- [26] T. Head-Gordon, M. Head-Gordon, M.J. Frish, C. Brooks, J. Pople, *Int. J. Quantum Chem., Quantum Biol. Symp.* 16 (1989) 311.
- [27] T. Head-Gordon, M. Head-Gordon, M.J. Frish, C. Brooks, J. Pople, *J. Am. Chem. Soc.* 113 (1991) 5989.
- [28] R.F. Frey, J. Coffin, S.Q. Newton, M. Ramek, V.K.W. Cheng, F.A. Momany, L. Schafer, *J. Am. Chem. Soc.* 114 (1992) 5369.
- [29] A. Perczel, Ö. Farkas, I.G. Csizmadia, *J. Mol. Struct. (Theochem)* 118 (1996) 7809.
- [30] H.A. Baldoni, A.M. Rodriguez, M.A. Zamora, G.N. Zamarbide, R.D. Enriz, Ö. Farkas, P. Császár, L.L. Torday, C.P. Sosa, I. Jákli, A. Perczel, J.G. Papp, M. Hollosi, I.G. Csizmadia, *J. Mol. Struct. (Theochem)* 465 (1999) 79.

- [31] I. Bagyi, B. Balogh, A. Czajlik, O. Elias, Z. Gaspari, V. Gergely, I. Hudaky, P. Hudaky, A. Kalaszi, L. Karolyhazy, K. Keseru, R. Kiss, G. Krajszovszky, B. Lang, T. Nagy, A. Racz, A. Szentesi, T. Tabi, P. Tapolcsanyi, J. Vaik, J.C.P. Koo, G.A. Chass, O. Farkas, A. Perczel, P. Matyus, *J. Mol. Struct. (Theochem)* 625 (2003) 121.
- [32] Ö. Farkas, M.A. McAllister, J.H. Ma, A. Perczel, M. Hollósi, I.G. Csizmadia, *J. Mol. Struct. (Theochem)* 369 (1996) 105.
- [33] M.F. Masman, M.G. Amaya, A.M. Rodriguez, F.D. Suvire, G.A. Chasse, O. Farkas, A. Perczel, R.D. Enriz, *J. Mol. Struct. (Theochem)* 543 (2001) 203.
- [34] S. Shirazian, S. Gronert, *J. Mol. Struct. (Theochem)* 397 (1997) 107.
- [35] A.M. Tarditi, M.W. Klipfel, A.M. Rodriguez, F.D. Suvire, G.A. Chasse, O. Farkas, A. Perczel, R.D. Enriz, *J. Mol. Struct. (Theochem)* 545 (2001) 29.
- [36] G.N. Ramachandran, *Biopolymers* 6 (1963) 1494.
- [37] G.N. Ramachandran, C. Ramakrishnan, V. Sasisekharan, *J. Mol. Biol.* 7 (1963) 95.
- [38] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, V. G. Zakrzewski, J. A. Montgomery, Jr., R. E. Stratmann, J. C. Burant, S. Dapprich, J. M. Millam, A. D. Daniels, K. N. Kudin, M. C. Strain, O. Farkas, J. Tomasi, V. Barone, M. Cossi, R. Cammi, B. Mennucci, C. Pomelli, C. Adamo, S. Clifford, J. Ochterski, G. A. Petersson, P. Y. Ayala, Q. Cui, K. Morokuma, N. Rega, P. Salvador, J. J. Dannenberg, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. Cioslowski, J. V. Ortiz, A. G. Baboul, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. Gomperts, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, J. L. Andres, C. Gonzalez, M. Head-Gordon, E. S. Replogle, J. A. Pople. GAUSSIAN 98, Revision A.11.4; Gaussian, Inc.: Pittsburgh, PA, 2002.
- [39] Y.-L. Zhao, Y.-D. Wu, *J. Am. Chem. Soc.* 124 (2002) 1570.
- [40] J. Skolnick, A. Kolinski, *Comput. Sci. Eng.* 3 (2001) 40.
- [41] M. Blaber, X. Zhang, B.W. Matthews, *Science* 260 (1993) 1637.
- [42] M. Blaber, X. Zhang, J.D. Lidstrom, S.D. Peptot, W.A. Basse, B.W. Matthews, *J. Mol. Biol.* 235 (1994) 600.