Computational and Theoretical Chemistry 969 (2011) 76–82

Contents lists available at ScienceDirect

Computational and Theoretical Chemistry

journal homepage: www.elsevier.com/locate/comptc

A theoretical study of the structure and electron density of the peptide bond

Iskandar Saada, Jason K. Pearson *

Department of Chemistry, University of Prince Edward Island, Charlottetown PE, Canada C1A 4P3

ARTICLE INFO

Article history:
Received 2 May 2011
Accepted 10 May 2011
Available online 26 May 2011

Keywords:
Dipeptides
Structure
Peptide bond
Electron density
Bond critical points
Quantum theory of atoms in molecules

ABSTRACT

We present a systematic study of the geometries and bond critical point electron densities of each of the 400 possible dipeptide structures. Equilibrium conformations for each of the dipeptides were located using the MMFF94 force field and the resultant geometries were further optimized using the B3LYP/6-311+G (d,p) method in the gas phase. We find that intramolecular hydrogen bonding is generally responsible for the observed conformations and that the peptide plane is significantly distorted in 15% of the compounds. The steric bulk of the amino acid side chains does not appear to be a significant contributor to the geometry about the peptide bond but rather the potential for hydrogen bonding determines the conformational preference. We also find that the electron density at the bond critical point (ρ_c) of the peptide bond strongly correlates to the equilibrium bond length (r_e) and that this relationship compares well with others reported for similar bond types. Using a power law relationship between ρ_c and r_e (i.e. $\rho_c = \alpha r_e^{-\beta}$), we show that the regression coefficients (α, β) for heteronuclear bonds may be estimated from those of homonuclear bonds.

1. Introduction

Proteins are crucial molecular components of biological systems and are responsible for structural support, storage, signaling, movement, and defence against foreign substances among other things [1]. Because the biological function of any protein is intimately linked to its three-dimensional structure, major research efforts have been mounted to develop a variety of computational techniques for protein structure prediction [2] to complement the array of experimental methods for structure elucidation. However, because proteins are often composed of thousands of amino acid residues, their size precludes a rigorous *ab initio* treatment and significantly smaller systems are used as models when quantum chemical techniques are to be employed. For this reason dipeptides have served as valuable models for larger oligopeptides and proteins. Studying dipeptides can lead to insight regarding how local interactions can dictate global features of larger peptides and proteins.

Early applications of *ab initio* methods to peptides were pioneered by Boyd and coworkers [3–5] and also by Peters [6–9]. These studies were limited to rigid potential energy surface scans due to the available technology and subsequent studies have focused on predicting detailed relaxed potential energy surfaces for a variety of purposes. For a more detailed account of theoretical studies on peptides prior to 2001, the reader is referred to the review by Chasse et al. [10] and references therein. More recently, prediction of the structure and conformational preferences of dipeptides in addition to clarifying the link between their structure and that of larger amino acid sequences remains an active area of research [11–24].

Ab initio dipeptide structural studies have also been valuable for the purpose of obtaining accurate geometrical force field parameters for semi-empirical or molecular mechanics simulations of biomolecules [25–28,12,29–31]. Such force fields find widespread use in the computational simulation of a vast array of biologically relevant processes. Also, in addition to serving as model systems, dipeptides themselves have been shown to play key biological roles and these have been investigated *in silico* using a variety of computational techniques [32–40].

The prediction of the potential energy surfaces of peptide chains has been described as a vital component in the study of the energetics of protein folding [10]. Often, such energy surfaces are optimized and described in terms of the well-known Ramachandran angles ψ and ϕ [41,42], while neglecting variations in the bond lengths and other angles within the so-called peptide plane. Recently however, it has been reported that the amide *plane* of the peptide bond is rarely exactly planar in small dipeptides and that small variations in geometrical parameters around the peptide bond can lead to large errors in the predicted structures of large peptide chains when only ψ and ϕ are allowed to optimize [18,20,21]. It has been further postulated that such variations in peptide geometry are correlated to the signature functional group of the amino acids in the sequence. Unfortunately, these studies are based on a small subset of less than ten dipeptide structures

* Corresponding author. Tel.: +1 902 566 0934; fax: +1 902 566 0632.
E-mail address: jpearson@upei.ca (J.K. Pearson).

2210-271X/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.comptc.2011.05.023

and thus there remains a significant gap in the space of possible peptides that can be considered in this context.

Despite the many contributions to understanding the structure and properties of dipeptides, to the best of our knowledge, no previous study of these important molecules has considered all possible combinations of residues. Instead, the approach has invariably been to focus on a small subset of dipeptides either because there was specific interest in the subset or because the full set would have exceeded the scope of the study. This is perhaps not surprising considering the fact that there are 20 naturally occurring amino acids and thus for a given stereochemical configuration there are $20^2 = 400$ unique dipeptide structures, owing to the fact that each of the two terminal positions is nonequivalent. Furthermore, there remain a considerable number of dipeptides that have not been studied at all. If one is interested in determining general structural features of dipeptides and the factors that control their overall conformation, then it is apparent that a more complete structural survey is required.

Amino acids are chiral molecules consisting of a central carbon atom (the α carbon) covalently bound to a carboxylic acid group, an amino group, a hydrogen atom, and a distinctive functional group (R) of which there are 20 that naturally occur. There do exist several amino acid residues that are not considered among these 20, however they are far less common and will not be discussed further here. Amino acids link via a condensation reaction between the amine of the first residue and the carboxylic acid of the second. This results in a free carboxylic acid group terminating one end of the linear sequence (known as the C-terminus) and an amine group terminating the opposite end (known as the N-terminus). Fig. 1 illustrates the general dipeptide structure studied in the current work.

A survey of dipeptide properties based on the current literature is further complicated by the fact that previous studies have employed a variety of theoretical techniques and structural modifications. Hartree–Fock (HF) [11,25–27,43,32,12–16,33,18,17,19,34,36,37,22,21,38,24], Density Functional Theory (DFT) [30,15,17,44,45,22,21,20,23,24,40], and second order perturbation methods (MP2) [11,43,12–14,30,15,46,44,34,47,36,45,37,22,39,23,31,40] have all been applied to the study of dipeptide structure. Additionally, many studies consider single amino acid residues terminated by acetyl and methylamide groups via peptide bonds as dipeptides [11,25,26,43,17,46,47,12,13,29,30,16,33,31,27,14,44], while others consider two amino acid residues joined by a single peptide bond [18,34,45,32,15,39,40,24,21,20] as in Fig. 1. Even among these variations, one can choose to consider the dipeptides as neutral or zwitterionic [36,35,23,37,38] and may or may not employ explicit or implicit solvation techniques. For these reasons, *general* conclusions regarding this class of molecule are difficult to draw from the results. A key goal of the current study is to present *ab initio* structural data on the full set of dipeptides using a reliable and consistent theoretical model.

In the current work, we report optimized structural parameters for all possible dipeptides using a reliable and consistent theoretical model. This will afford the opportunity to interpret the structure of dipeptides in a general way. Specifically, we seek to answer the following questions: How significantly does the peptide plane deviate from planarity across the full spectrum of dipeptides? How does the signature R group of each amino acid contribute to the equilibrium structure of the peptide? What effects dominate the overall structure of dipeptides?.

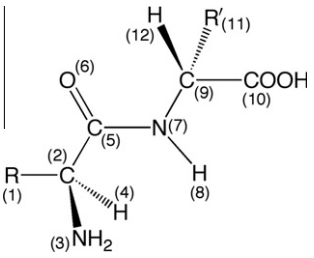

Fig. 1. General structure of dipeptides considered in this study.

We will also be concerned with how the electron density in the peptide bond is affected as a function of the variation in amino acid side chains. Within the context of Bader's Quantum Theory of Atoms in Molecules (QTAIM) [48], a chemical bond is identified by the presence of a bond path between two adjacent nuclei. Along any such path there exists a unique point where the electron density is a minimum in the direction of the bond but a maximum in all perpendicular directions. This point is known as the bond critical point (BCP) and the local properties of the density and Laplacian of the density at the BCP provide a unique "fingerprint" characterizing the chemical bond itself [49]. It was first pointed out by Boyd and coworkers that bond lengths can be reliably predicted based on the value of the electron density at the BCP and vice versa [50–52]. Since that time there has been a great deal of interest in investigating such relationships for a wide variety of bond types and molecular environments [53–56]. Such relationships can prove to be of great utility as one can estimate properties related to the density at the BCP (such as the bond order [57] or molecular similarity indices [49]) from bond length alone. Our set of 400 dipeptides provides one of the largest statistical sets of compounds to be used to investigate relationships between the electron density at the BCP (which we will abbreviate as ρ_c) and equilibrium bond length (r_e) to date. In addition, their biological relevance and relation to larger peptide sequences makes dipeptides an excellent test set to provide important data related to protein structure. In the current work, we study the correlation between ρ_c and r_e for the peptide bond and report a simple but highly effective predictive tool for these properties among a variety of bond types.

Where it is convenient to do so, we employ the standard three letter abbreviations when referring to amino acids and when referring to a particular dipeptide in the text we adopt a convention of listing the N terminal residue first. That is, the Ala-Gly dipeptide corresponds to a structure in which alanine is in the N terminal position and glycine is in the C terminal position.

2. Computational details

The pH of an aqueous *in vivo* solution of peptide chains usually dictates that they will exist in a zwitterionic form, and to accurately replicate that environment *in silico* would require explicit solvation and prohibitively cumbersome potential energy surfaces for the scope of the current study. Without the inclusion of many water molecules, a zwitterionic peptide chain will optimize to minimize the distance between its oppositely charged terminal groups, biasing the stationary point geometries. For this reason, all of our structures have been modeled as neutral compounds in the gas phase, a technique that has been extensively employed with success for similar systems in previous studies [18,34,45,32,15,39,40,24,21,20,38], including those where direct comparison to experimental structures was possible [37].

Equilibrium conformers of each of the 400 unique dipeptides were obtained using a molecular mechanics optimization employing the MMFF94 force field [58]. An extensive relaxed scan of the potential energy surface for each dipeptide was performed using the 'equilibrium conformer' search option implemented in the Spartan '08 software package [59] and the lowest energy conformation was chosen for further study. While it is generally not expected that this procedure will consistently locate the global