ELSEVIER

Journal of Molecular Structure: THEOCHEM 775 (2006) 9–17

THEO CHEM

www.elsevier.com/locate/theochem

Structures and conformational energies of amino acids in the zwitterionic, hydrogen-bonded state

Carl Henrik Görbitz *

Department of Chemistry, University of Oslo, N-0315 Oslo, Norway

Received 3 April 2006; received in revised form 30 June 2006; accepted 10 July 2006
Available online 21 July 2006

Abstract

Ab initio methods have been used to calculate relative energies for all side chain rotamers of the hydrophobic amino acids L-valine, L-leucine, L-isoleucine and L-norvaline, with supplementary calculations for glycine, L-alanine, L-serine and *S*-fluoroglycine. The amino acids were in the zwitterionic state, which was stabilized by explicit inclusion of three hydrogen bond donors and three hydrogen bond acceptors, thus forming complexes of seven molecules. Each complex was optimized at the HF/6-311++G** and B3LYP/6-311++G** levels of theory. The results are in excellent agreement with observations in crystal structures. A detailed analysis shows that the lowest energy (and most frequently observed) side chain rotamers invariably have the best set of intermolecular interactions overall, and in particular the most favourable hydrogen bonds to the three acceptor molecules. If just the isolated zwitterionic amino acids are considered (with geometries fixed as in the complexes), different sets of relative energies are obtained that do not fit the crystal structure distributions. The effect of the nature of the side chains on the total interaction energy has been addressed by considering not only their inductive effect on the hydrogen bonds involving the charged amino and carboxylate groups, but also the direct interactions between the side chains and the surrounding donors and acceptors.

Keywords: Ab initio calculations; Amino acids; Conformation analysis; Hydrogen bonds; Inductive effect

1. Introduction

The conformational diversity of amino acids, peptides and proteins is due not only to the rotation around main-chain covalent bonds, but also to the existence of various side chain rotamers. For all the common amino acid residues [except glycine (Gly) and alanine (Ala)] statistical surveys of crystal structures have revealed rotamer frequency distributions [1–3] that obviously reflect the relative conformational energies. It would clearly be desirable to have some perception of the magnitudes of the energies involved, but hardly any theoretical calculations are available dealing with this matter, except a recent study on methionine and lysine rotamers in proteins [4]. Even force field calculations are difficult due to the shortage of the required parameters [5]. At the heart of this remarkable lack of knowledge lies the fact that proton transfer between the –COOH and $–NH_2$ groups to produce $–COO^-$ and $–NH_3^+$ occurs in crystal structures and in aqueous solutions. For the isolated amino acids, on the other hand, low energy conformations normally incorporate one or more intramolecular hydrogen bonds between the uncharged functional groups, and zwitterions do not represent even local minima on the energy hypersurface. Accordingly, the abundance of intricate results from ab initio calculations on amino acids [6–11] are not directly relevant for observations in solution and in crystals. In some studies, limited to Gly, Ala and serine (Ser) [12–16], the solvent was taken into account by implementing continuum solvent models. Explicit inclusion of one to four water molecules [17–22], NaCl [23] or Zn^{2+} [24] in more traditional ab initio calculations has been used in a few investigations on the factors required to stabilize amino acid zwitterions. Apparently, none of these

* Tel.: +47 2285 5460; fax: +47 2285 5441.
E-mail address: c.h.gorbitz@kjemi.uio.no

0166-1280/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.theochem.2006.07.015

studies have dealt with the energy differences between the side chain rotamers, and the effect of side chains on the various terms of the total interaction energy is not known.

The aim of the work presented here was to fill part of this void by carrying out calculations on amino acids in complexes where the full hydrogen bonding potential has been utilized by inclusion of three hydrogen bond donors and three hydrogen bond acceptors. The investigation is focused on the amino acids Gly, Ala, valine (Val), leucine (Leu), isoleucine (Ile) and norvaline (Nva), which are simple in the sense that the side chains do not contain functional groups with strong hydrogen bond donors or acceptors. Nva is not one of the 20 common amino acids, but it was included since it forms the common C^{α}–C^{β}–C^{γ}–C^{δ} stem of the arginine (Arg), lysine (Lys), glutamine (Gln) and glutamic acid (Glu) side chains, thus providing useful information for all of these four amino acids.

All amino acids discussed in the text are present as the L-enantiomer. Specific conformations have received designations of the type Xaa(C^{α}–C^{β} rotamer, C^{β}–C^{γ} rotamer). Accordingly, Ile(*trans*/*gauche*+, *trans*) indicates that the $\chi^{1,1}$ (N–C^{α}–C^{β}–$C^{\gamma 1}$) rotamer is *trans*, $\chi^{1,2}$ (N–C^{α}–C^{β}–$C^{\gamma 2}$) is *gauche*+ and χ^{2}(C^{α}–C^{β}–$C^{\gamma 1}$–C^{δ}) *trans*.

2. Computational methods

2.1. Choice of model system

With the aim of using ab initio methods to study the conformational properties of zwitterionic amino acids relevant for crystal structures and aqueous solutions, an evident choice of a model system would be the selected amino acid hydrogen-bonded to six water molecules, three as donors and three as acceptors. Upon closer inspection, however, this proves not to be the best alternative. Since water molecules have three atoms, there are more degrees of freedom than for two-atom molecules, with the potential risk of unforeseen and unwanted steric conflict to occur in optimized structures, or, conversely, introduction of additional secondary contacts that would add extra interaction energy and mask energy differences between amino acid conformations. The most important liability of such a complex, however, is the strength of the six hydrogen bonds; they are not only weaker than the dominating $-NH_3^+ \cdots {}^-OOC-$ interactions in crystals, but in fact also weaker than peptide $\cdots$ water hydrogen bonds in aqueous solution due to the missing cooperative effect in such a small system [25]. This has been illustrated previously for the system $(H_2O)_3HCOO \cdots H \cdots H_2N–CH_3(H_2O)_2$ for which the neutral acid–base pair ($HCOOH \cdots H_2N–CH_3$) was the global energy minimum at the HF/6-311+G** level of theory [26]. The situation in aqueous solution and in the crystalline phase, corresponding to a global energy minimum for the ionic pair $HCOO^- \cdots {}^+H_3N–CH_3$, was obtained only after substitution of the water molecules with the stronger HF donors and NH_3 acceptors, in the complex $(HF)_3HCOO \cdots H \cdots H_2N–CH_3(NH_3)_2$.

For the present work ammonia proved to be a less suitable acceptor molecule due to potential steric conflict with the bulky amino acid side chains. Instead, BeO was selected as the acceptor. This small, two-atom molecule with few electrons (reducing computational time) is also an excellent hydrogen bond acceptor due to the very polar covalent bond.

All three donors were initially included as HF molecules, but in some optimizations attraction between the *syn*-HF molecule and a BeO molecule resulted in unwanted formation of a new O–Be–F–H species. This was effectively prevented by using a water molecule as the third donor. The final model for a complex is depicted in Fig. 1. Apart from the somewhat shorter hydrogen bonds, the geometry of the Ala complexes with three BeO acceptors is very similar to the complex with three NH_3 acceptors.

2.2. Ab initio optimizations

As the augmented Dunning correlation-consistent aug-cc-pVXZ ($X = D, T, Q, 5,$ and 6) basis sets do not apply to alkaline and alkaline-earth metals, the 6-311++G** basis set was thus used throughout this investigation. B3LYP and HF ab initio optimizations were carried out for all basic amino acid conformations with the GAUSSIAN98 computer program [27]. Calculations at the correlated level using second-order Møller–Plesset perturbation theory (MP2) are quite demanding for such large systems, and only the Gly complex was subject to geometry optimization [with a subsequent single point calculation using the 6-311++G(2df, 2pd) basis set].

It is imperial to note that the aim of the minimizations presented in this work was *not* to find the global energy minimum for each such complex *per se*. Rather, the role of the acceptors and donors was to form a stable, predictable and almost invariant environment for the amino acids, interfering to the smallest possible extent with the desired results of the calculations, i.e. the relative energies of the various side chain conformations. Following this approach, the six hydrogen bonds in the complexes were

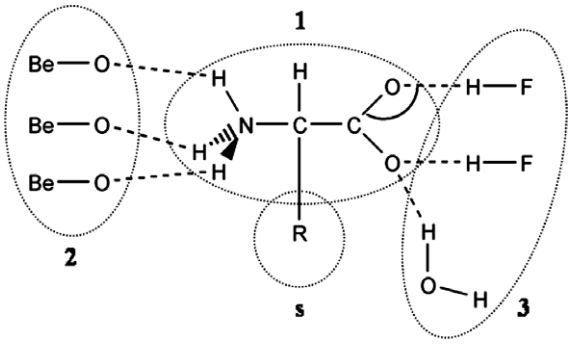

Fig. 1. The model system used in the ab initio optimizations. Dotted ellipses enclose the polar head of the amino acid (**1**), the side chain (**s**), the three acceptors (**2**) and the three donors (**3**). The C–O $\cdots$ H angle specified by an arc was fixed to 130° during optimization.