Conformational changes in glycine tri- and hexapeptide

A.V. Yakubovich[a], I.A. Solov'yov[a], A.V. Solov'yov[a,b], and W. Greiner

Frankfurt Institute for Advanced Studies, Johann Wolfgang Goethe University, Max von Laue Str. 1,
60438 Frankfurt am Main, Germany

Abstract. We have investigated the potential energy surfaces for glycine chains consisting of three and six amino acids. For these molecules we have calculated potential energy surfaces as a function of the Ramachandran angles φ and ψ, which are widely used for the characterization of the polypeptide chains. These particular degrees of freedom are essential for the characterization of proteins folding process. Calculations have been carried out within ab initio theoretical framework based on the density functional theory and accounting for all the electrons in the system. We have determined stable conformations and calculated the energy barriers for transitions between them. Using a thermodynamic approach, we have estimated the times of the characteristic transitions between these conformations. The results of our calculations have been compared with those obtained by other theoretical methods and with the available experimental data extracted from the Protein Data Base. This comparison demonstrates a reasonable correspondence of the most prominent minima on the calculated potential energy surfaces to the experimentally measured angles φ and ψ for the glycine chains appearing in native proteins. We have also investigated the influence of the secondary structure of polypeptide chains on the formation of the potential energy landscape. This analysis has been performed for the sheet and the helix conformations of chains of six amino acids.

PACS. 31.15.Ar Ab initio calculations – 31.15.Ew Density-functional theory – 31.50.Bc Potential energy surfaces for ground electronic states – 34.20.-b Interatomic and intermolecular potentials and forces, potential energy surfaces for collisions

1 Introduction

It is well-known that proteins consist of amino acids whose number may vary in the range from hundreds up to tens of thousands. Small fragments of proteins are usually called polypeptide chains or polypeptides. This work is devoted to a study of the conformational properties of glycine polypeptide chains.

Since recently, it became possible to study experimentally small fragments of proteins and polypeptides in the gas phase with the use of the MALDI mass spectroscopy [1–4] and the ESI mass spectroscopy [5,6]. From theoretical view point, investigation of small polypeptides is of significant interest because they can be treated by means of ab initio methods allowing accurate comparison of theoretical predictions with experiment. The results of ab initio calculations can be then utilized for the development of model approaches applicable for the description of larger and more complex protein structures.

Polypeptides are characterized by the primary and the secondary structure [7–10]. Different geometrical configu-

rations of a polypeptide are often called as the conformations. One can expect that chemical and physical properties of various conformations of complex molecules might differ significantly. The number of various conformations (isomeric states) grows rapidly with the growth of a system size. Thus, a search for the most stable conformations becomes an increasingly difficult problem for large molecules. With the help of the NMR spectroscopy and the X-rays diffraction analysis it has been shown [9], that the sheet and the helix structures are the most prominent elements of the protein secondary structure.

The main difference between the sheet and the helix structures is due to the difference of the dihedral angles formed by the atoms of the polypeptide chains in the two cases. These degrees of freedom are responsible for the transition of the molecule from one conformation to another. By increasing the temperature of the system, the degrees of freedom responsible for twisting of the polypeptide chain can be activated. The study of this transition and evaluation of its characteristic duration are of significant interest, because this problem is closely related to one of the most intriguing problems of the protein physics — the protein folding. To study this transition it is necessary to investigate the potential energy surface of amino

[a] *On leave from* the A.F. Ioffe Institute, St. Petersburg, Russia.
[b] e-mail: `solovyov@fias.uni-frankfurt.de`

acid chains with respect to their twisting. Besides the protein folding, the potential energy landscapes of polypeptides carry a lot of detail and useful information about the structure of these molecules.

In the present paper we have studied the potential energy surface for small glycine chains. These molecules were chosen because they are often present in native proteins as fragments, and also because they allow ab initio theoretical treatment due to their relatively small size.

Previously, only glycine and alanine dipeptides were studied in detail. Sometimes their analogues were used to reduce the computational costs (for example (S)-α-(formylamino)propanamide). In references [11–13] alanine and glycine dipeptides were investigated within the Hartree-Fock theory. In these papers the potential energy surfaces were calculated versus the twisting angles of the molecules. Different stable conformations of the dipeptides, corresponding to different molecular conformations, were determined. Each stable conformation of the molecule was additionally studied on the basis of the perturbation theory, which takes into account many-electron correlations in the system. In references [14–19] different conformations and their energies were determined within the framework of the density functional theory. In reference [19] dynamics of the alanine dipeptide analog was discussed and time of the transition between the two conformations of the alanine dipeptide was found. In [20] the glycine molecule was studied with accounting for the many electron correlations within the many-body perturbation theory. In this work the most important glycine isomers and the frequencies of the normal vibration modes were determined.

A number of papers were devoted to the study of tripeptides. In references [21–25] dynamics of the alanine and glycine tripeptides was studied by means of classical molecular dynamics and with the use of semi-empirical potentials (such as GROMOS, CHARMM and AMBER). In reference [26] within the framework of the Hartree-Fock theory several stable conformations of alanine and glycine tripeptides were found. In reference [27] the Raman and IR spectra for alanine and glycine tripeptides were measured in neutral, acidy and alkali environments.

Polypeptides have been studied less. We are aware of only several related papers. In particular, stable conformations of neutral and charged alanine hexapeptides were obtained with the use of empirical potentials and discussed in reference [28]. Experimental NMR study of various conformations of alanine heptapeptides at different temperatures was carried out in reference [29]. In reference [30] with the use of empirical molecular dynamics based on Monte-Carlo methods, a polypeptide consisting of 21 amino acids was described.

In the present paper we have performed an ab initio calculation of the multidimensional potential energy surface for the glycine polypeptide chains consisting of three and six amino acids. The potential energy surface versus twisting degrees of freedom of the polypeptide chain has been calculated. The calculations have been performed within ab initio theoretical framework based on the den-

sity functional theory (DFT) accounting for all the electrons in the system. Previously, this kind of calculations were performed only for dipeptides (see, e.g., [11,12,19]). For larger molecules, only a few conformations were considered (see citations above). We have calculated the energy barriers for the transitions between different molecular conformations and determined the energetically most favorable ones. Using a thermodynamic approach, we have estimated times of the characteristic transitions between the most energetically favorable conformations. The results of our calculation have been compared with other theoretical simulations and with the available experimental data. We have also analysed how the secondary structure of polypeptide chains influences the potential energy landscapes. In particular, the role of the secondary structure in the formation of stable conformations of the chains of six amino acids being in the sheet and in the helix conformations has been elucidated. Some preliminary results of our work were published as electronic preprints [31,32].

Our paper is organized as follows. In Section 2 we give a brief overview of theoretical methods used in our work. In Section 3 we present and discuss the results of our computations. In Section 4 we draw a conclusion to this paper. The atomic system of units, $|e| = m_e = \hbar = 1$, is used throughout the paper unless other units are indicated.

2 Theoretical methods

In the present paper we study the multidimensional potential energy surfaces for glycine polypeptides within the framework of the density functional theory. The potential energy surfaces are multidimensional functions of atomic coordinates. In our work the potential energy surfaces were considered as a function of the dihedral angles formed by the atoms of the polypeptide chain. For this calculation the Born-Oppenheimer approximation allowing to separate the motion of the electronic and ionic subsystems is used.

The density functional theory (DFT) is a common tool for the calculation of properties of quantum many body systems in which many electron correlations play an important role. The DFT formalism is well-known and can be found in many textbooks (see e.g. [33,34]). Therefore in our work we present only the basic equations and ideas of this method.

Electronic wave functions and energy levels within the framework of DFT are obtained from the Kohn-Sham equations, which read as (see e.g. [33,34]):

$$\left(\frac{\hat{p}^2}{2} + U_{ions} + V_H + V_{xc} \right) \psi_i = \varepsilon_i \psi_i, \qquad i = 1...N \quad (1)$$

where the first term represents the kinetic energy of the ith electron with the wavefunction ψ_i and the energy ε_i, U_{ions} describes the electron attraction to the ionic centers, V_H is the Hartree part of the interelectronic interaction [35], V_{xc} is the local exchange-correlation potential.

The exchange-correlation potential is defined as a functional derivative of the exchange-correlation energy
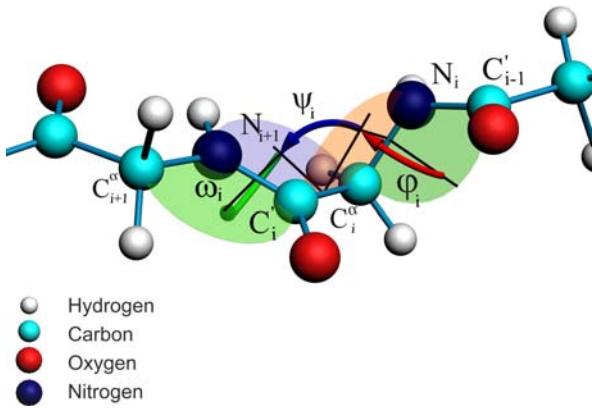
Fig. 1. Dihedral angles φ and ψ used to characterize the potential energy surface of the polypeptide chain. A colour version of the figure is available in electronic form at `http://www.eurphysj.org`.

functional:

$$V_{xc} = \frac{\delta E_{xc}[\rho]}{\delta \rho(\mathbf{r})}. \qquad (2)$$

Equation (2) is exact and follows from the Hohenberg theory [36]. However, no unique potential E_{xc}, universally applicable for different systems and conditions, has been found so far.

Approximate functionals employed by the DFT usually partition the exchange-correlation energy into two parts, referred to as the *exchange* and the *correlation* terms:

$$E_{xc}[\rho] = E_x(\rho) + E_c(\rho). \qquad (3)$$

Both terms are the functionals of the electron density, which can be of two distinctly different types: either a *local* functional depending only on the electron density ρ or a *gradient-corrected* functionals depending on both ρ and its gradient, $\nabla \rho$. A variety of exchange correlation functionals can be found in literature. In our work we have used the hybrid Becke-type three-parameter exchange functional [37] paired with the gradient-corrected Lee, Yang and Parr correlation functional ($B3LYP$) [38,39].

3 Results and discussion

3.1 Determination of the polypeptide twisting degrees of freedom

In this section we present the potential energy surfaces for the glycine polypeptide chains calculated versus dihedral angles φ and ψ defined in Figure 1. In particular, we focus on the chains consisting of three and six amino acids.

Both angles are defined by the four neighboring atoms in the polypeptide chain. The angle φ_i is defined as the dihedral angle between the planes formed by the atoms $(C'_{i-1} - N_i - C_i^\alpha)$ and $(N_i - C_i^\alpha - C'_i)$. The angle ψ_i is defined as the dihedral angle between $(N_i - C_i^\alpha - C'_i)$ and $(C_i^\alpha - C'_i - N_{i+1})$ planes. Beside the angles φ_i and ψ_i there is an angle ω_i, which is defined as the dihedral angle

between the $(C_i^\alpha - C'_i - N_{i+1})$ and $(C'_i - N_{i+1} - C_{i+1}^\alpha)$ planes. The atoms are numbered from the NH_2- terminal of the polypeptide. The angles φ_i, ψ_i and ω_i take all possible values within the interval $[-180°;180°]$. For the unambiguous definition we count the angles φ_i, ψ_i and ω_i clockwise, if one looks on the molecule from its NH_2- terminal (see Fig. 1). This way of angle counting is the most commonly used [10].

The angles φ_i and ψ_i can be defined for any amino acid in the chain except for the first and the last ones. Below we omit the subscripts and assume angles φ and ψ to correspond the middle amino acid of the polypeptide.

It is not feasible to study the dependence of potential energy on all possible angles φ and ψ for all amino acids, because the amount of computer powers required for such DFT computation would be enormous. In the present paper we consider only the twisting angles in the middle amino acid of the polypeptide, because we want to stress the topological conformity of the potential energy surfaces in the tripeptide and in the hexapeptide. One can expect that for the inner amino acids of a polypeptide the dependence of the potential energy surfaces on the twisting angles should be similar if the amino acids are loosely correlated. For glycines this condition is fulfilled, because glycines do not have side radicals and thus interact weakly with each other along the polypeptide chain. The study of correlation functions for different amino acids in a polypeptide chain is an interesting question, which is certainly worth to study for larger polypeptide chains consisting of larger amino acids. In this paper we do not further touch this question and leave it open for further investigation.

3.2 Optimized geometries of glycine polypeptides

In order to study the twisting of a polypeptide chain one needs first to define its initial structure. Although the number of its conformations increases with the growth of the molecule size, there are certain types of polypeptide structure, namely the sheet and the helix conformations, which are the most typical. In the present paper we have investigated twisting of the polypeptide chains of the sheet and the helix conformation. By varying the angles φ and ψ in the central amino acid one can create the structure of the polypeptide differing significantly from the pure sheet or helix conformations. If the structure of a polypeptide can be transformed to a helix or a sheet one by a trivial variation of φ and ψ, such polypeptides are referred below as belonging to the group of the helix or the sheet structure respectively.

In Figure 2 we present the optimized geometries of glycine polypeptide chains that have been used for the exploration of the potential energy surfaces. All geometries have been optimized with the use of the B3LYP functional. Figure 2a shows the glycine tripeptide structure. In the present work we choose in the sheet conformation, because the tripeptide is too short to form the helix conformation. Figures 2b and 2c show glycine hexapeptide in the sheet and the helix conformations respectively. The
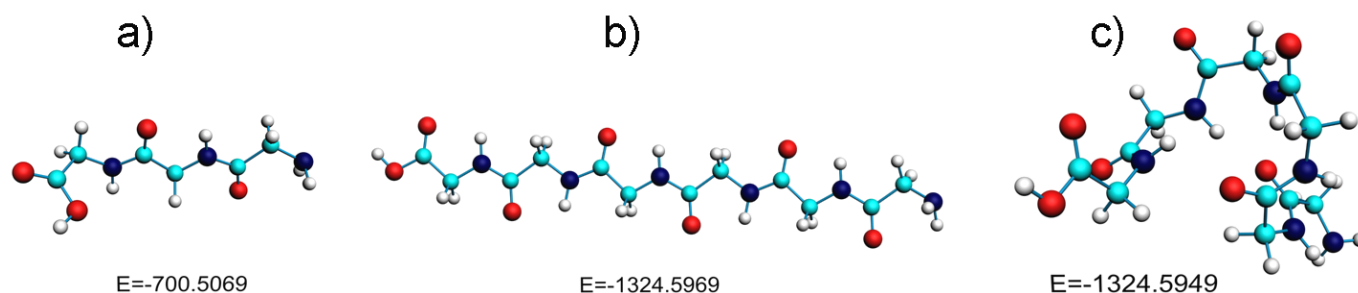
Fig. 2. Optimized geometries of glycine polypeptide chains calculated by the B3LYP/6-31++G(d,p) method: (a) glycine tripeptide; (b) glycine hexapeptide (sheet conformation); (c) glycine hexapeptide (helix conformation). A colour version of the figure is available in electronic form at http://www.eurphysj.org.
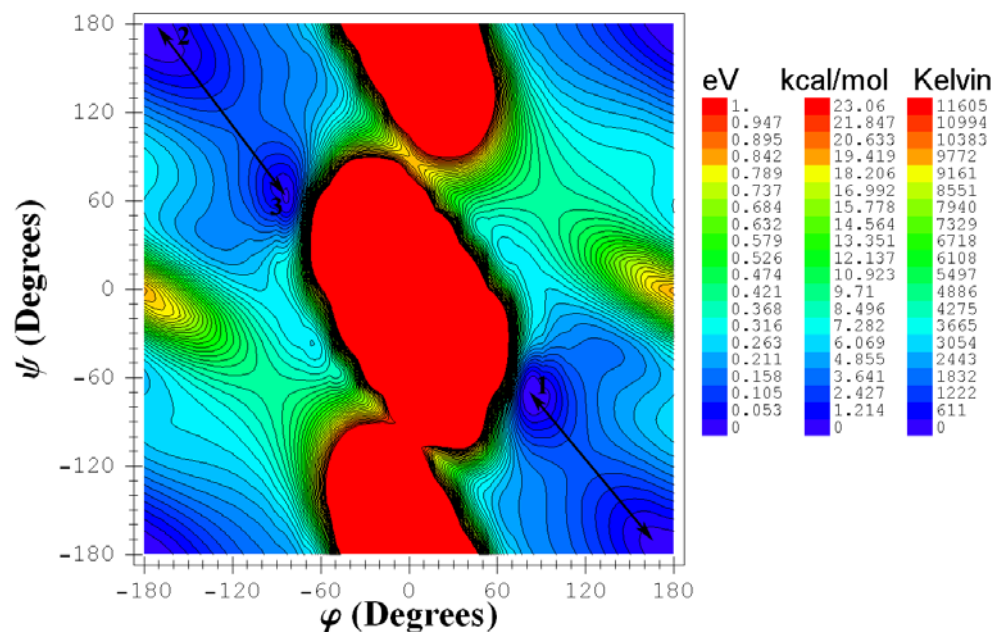

Fig. 3. Potential energy surface for the glycine tripeptide calculated by the B3LYP/6-31G(2d,p) method. Energies are given in eV, kcal/mol and Kelvin. Numbers mark energy minima on the potential energy surface. Arrows show the transition paths between different conformations of the molecule. A colour version of the figure is available in electronic form at http://www.eurphysj.org.

total energies (in atomic units) of the molecules are given below the images.

3.3 Potential energy surface for glycine tripeptide

In Figure 3 we present the potential energy surface for the glycine tripeptide calculated by the B3LYP/6-31G(2d,p) method. The energy scale is given in eV, kcal/mol and Kelvin. Energies on the plot are measured from the lowest energy minimum of the potential energy surface.

From the figure follows that there are several minima on the potential energy surface. They are numbered according to the value of the corresponding energy value. Each minimum corresponds to a certain conformation of the molecule. These conformations differ significantly from each other. In the case of glycine tripeptide there are only three conformations, shown in Figure 4. Dashed lines show the strongest hydrogen bonds in the system, which arise when the distance between hydrogen and oxygen atoms becomes less then 2.9 angstroms.

To calculate the potential energy surface the following procedure was adopted. Once the stable structure of the molecule has been determined and optimized, all but two (these are the angles φ and ψ in the central amino acid) degrees of freedom were frozen. Then the energy of the molecule was calculated by varying φ and ψ. This procedure was used to calculate all potential energy surfaces presented below in this section. It allows one to find efficiently the minima on the potential energy surface and to determine the main stable conformations of the molecule. The absolute energy values of different conformations of the tripeptide found by this method are not too accurate, because the method does not account for the relaxation of other degrees of freedom in the system. To calculate the potential energy surface with accounting for the relaxation one needs 20–30 times more of the computer time. Therefore, a systematic calculations with accounting for the relaxation have not been performed in our work. Instead, we have performed a complete optimization of the molecular conformations, corresponding to all minima on the calculated potential energy surface.

In Figure 4 we compare stable conformations of the glycine tripeptide calculated with and without accounting for the relaxation of all atoms in the system. As it is seen from this figure the angles φ and ψ differ by about
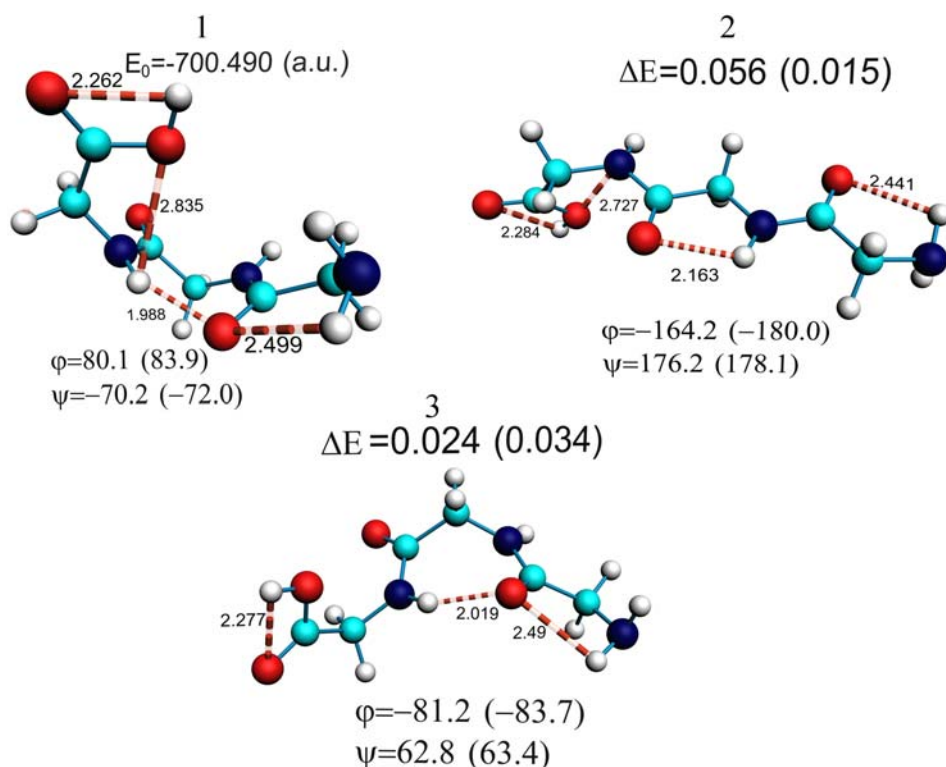
Fig. 4. Optimized conformations of the glycine tripeptide. Different geometries correspond to different minima on the potential energy surface (see contour plot in Fig. 3). Below each image we present angles φ and ψ, which have been obtained with accounting for relaxation of all degrees of freedom in the system. Values in brackets give the angles calculated without accounting for relaxation. Above each image, the energy of the corresponding conformation is given in eV. The energies are counted from the energy of conformation 1 (the energy of conformation 1 is given in a.u.). Values in brackets give the energies obtained without accounting for the relaxation of all degrees of freedom in the system. Dashed lines show the strongest hydrogen bonds. Their lengths are given in angstroms. A colour version of the figure is available in electronic form at http://www.eurphysj.org.

Table 1. Comparison of dihedral angles φ and ψ corresponding to different conformations of glycine tripeptide (column 3) with angles φ and ψ for glycine dipeptide from references [11,12] (column 1 and 2).

conformation	φ, Ref. [11]	ψ, Ref. [11]	φ, Ref. [12]	ψ, Ref. [12]	φ	ψ
1	-	-	76.0	−55.4	80.1	−70.2
2	−180.0	180.0	−157.2	159.8	−164.2	176.2
3	−85.2	67.4	−85.8	79.0	−81.2	62.8

10 percent in the two cases. This difference arises due to the coupling of φ and ψ with other degrees of freedom. Note the change of the sign of the relative energies of some conformations. This effect is due to the rearrangement of side atoms (radicals) in the polypeptide chain which lowers the energies of different conformations differently.

In our work the potential energy surface has been calculated and interpolated on the grid with the step of 18°. This step size is an optimal one, because the interpolation error is about 9°, i.e. comparable with the angle deviations caused by the relaxation of all degrees of freedom in the system.

In references [11,12] several stable conformations were found for alanine and glycine dipeptides. The values of angles φ and ψ for the stable conformations of dipeptide and tripeptide are close indicating that the third amino acid in tripeptide makes relatively small influence on the values of dihedral angles of two other amino acids. In earlier papers [11,12] dipeptides were studied within the framework of the Hartree-Fock theory. In reference [11], values of φ and ψ were obtained by the HF/6-31+G* method, and in reference [12] by HF/6-31G**. In Table 1 we compare the results of our calculation for tripeptide with the corresponding data obtained for dipeptides. Some discrepancy

between the values presented is due to the difference between the dipeptide and tripeptide (i.e. the third glycine in tripeptide affects the values of angles φ and ψ). However, another source of discrepancy might arise due to accounting for the many-electron correlations in the DFT and neglecting this effect in the Hartree-Fock theory used in references [11,12].

Figure 3 shows that some domains of the potential energy surface, where the potential energy of the molecule increases significantly, appear to be unfavorable for the formation of a stable molecular configuration. The growth of energy takes place when some atoms in the polypeptide chain approach each other of small distances. Accounting for the molecule relaxation results in the decrease of the system energy in such cases, but the resulting molecular configurations remain unstable. We call such domains on the potential energy surface as forbidden ones. In Figure 3 one can identify two forbidden regions in the vicinity of the points (0, 0) and (0, 180). At (0, 0) a pair of hydrogen and oxygen atoms approach to the distances much smaller than the characteristic H−O bond length. This leads to a strong interatomic repulsion caused by the exchange interaction of electrons. At (0, 180) the Coulomb repulsion of pair of oxygen atoms causes the similar effect.
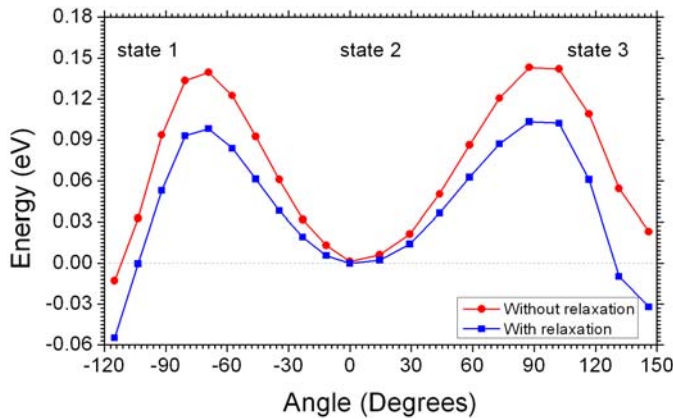
Fig. 5. Transition barriers between conformations $1 \leftrightarrow 2 \leftrightarrow 3$ of the glycine tripeptide. Circles and squares correspond to the barriers calculated without and with relaxation of all degrees of freedom in the system.

Figure 3 shows that there are three minima on the potential energy surface for glycine tripeptide. The transition barriers between the conformations $2 \leftrightarrow 1$ and $2 \leftrightarrow 3$ are shown in Figure 5. They have been calculated with and without relaxation of the atoms in the system. The corresponding transition paths are marked in Figure 3 by arrows. This comparison demonstrates that accounting for the relaxation significantly lowers the barrier height and influences the relative value of energy of the minima.

Let us now estimate the time needed for a system for the transition from one conformation to another. To do this we use the Arhenius equation, which reads as:

$$\frac{1}{\tau} = \Omega e^{-\frac{\Delta E}{kT}} \qquad (4)$$

where τ is the transition time, Ω is the factor, determining how frequently the system approaches the barrier, ΔE is the barrier height, T is the temperature of the system, k is the Bolzmann factor.

Figure 6 shows the barrier for the transition between the two main conformations of the alanine dipeptide analog ((S)-α- (formylamino)propanamide). It is seen that $\Delta E_{1 \to 2} = 0.047$ eV for the transition $1 \to 2$, while for the transition $2 \to 1$: $\Delta E_{2 \to 1} = 0.079$ eV. The frequency Ω for this molecule is equal to 42.87 cm^{-1}. Thus, at $T = 300$ K, we obtain $\tau_{2 \times Ala}^{1 \to 2} \sim 5$ ps and $\tau_{2 \times Ala}^{2 \to 1} \sim 17$ ps. This result is in excellent agreement with the molecular dynamics simulations results obtained in [19] predicting $\tau \sim 7$ ps for the transition $1 \to 2$ and $\tau \sim 19$ ps for the transition $2 \to 1$. This comparison demonstrates that our method is reliable enough and it can be used for the estimation of transition times between various conformations of the polypeptides.

Using the B3LYP/6-31G(2d,p) method we have calculated the frequencies of normal vibration modes for the glycine tripeptide. The characteristic frequency corresponding to twisting of the polypeptide chain is equal to 33.49 cm^{-1}. From figure 5 follows that $\Delta E_{2 \to 3} = 0.103$ eV for the transition $2 \to 3$, and $\Delta E_{3 \to 2} = 0.132$ eV for the transition $3 \to 2$. Thus, we obtain $\tau_{3 \times Gly}^{2 \to 3} \sim 54$ ps
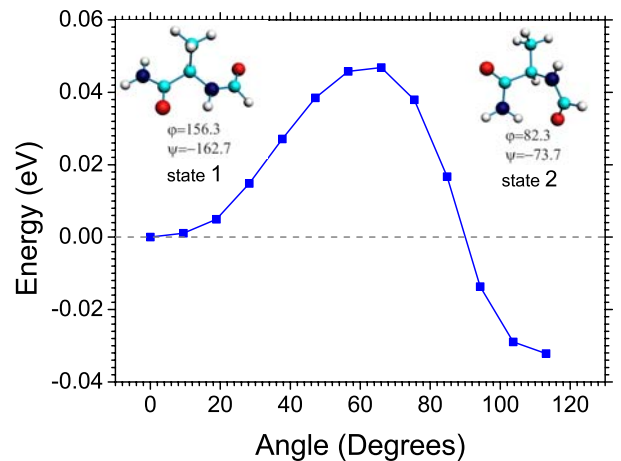

Fig. 6. Transition barriers between conformations $1 \leftrightarrow 2$ of alanine dipeptide analog calculated by the B3LYP/6-31+G(2d,p) method accounting for the relaxation of all degrees of freedom in the system. Structure of the conformations 1 and 2 is shown near each minimum.

and $\tau_{3 \times Gly}^{3 \to 2} \sim 164$ ps. Let us note that these transition times can be measured experimentally by means of NMR [10,43].

3.4 Potential energy surface for glycine hexapeptide with the sheet and the helix secondary structure

In Figure 7 we present contour plots of the potential energy surfaces for the glycine hexapeptide with the sheet (part a) and the helix (part b) secondary structure, respectively, versus dihedral angles φ and ψ. In both cases the forbidden regions arise because of the repulsion of oxygen and hydrogen atoms analogously to the glycine tripeptide case.

Minima 1–5 on the potential energy surface 7a correspond to different conformations of the glycine hexapeptide with the sheet secondary structure. Note that minima 1–3 are also present on the potential energy surface of the glycine tripeptide. Geometries of the conformations 1–5 are shown on the right-hand side of Figure 7a.

For the glycine hexapeptide with the sheet secondary structure additional minima 4–5 arise. The appearance of these minima is the result of the interaction of the outermost amino acids, which are absent in the case of tripeptide.

Energy barrier as a function of a scan variable (see Fig. 7a) for the transition between conformations 1 and 2 is shown in Figure 8. The energy dependence has been calculated with and without relaxation of all the atoms in the system. In the case of glycine hexapeptide with the sheet secondary structure the barrier height (0.128 eV) for the transition $1 \to 2$ appears to be close to the corresponding barrier height of the glycine tripeptide (0.103 eV), while the barrier height for the transition $2 \to 1$ is significantly lower (0.028 eV). The normal vibration mode frequency, corresponding to the twisting of the polypeptide chain is equal to 15.45 cm^{-1} and was calculated
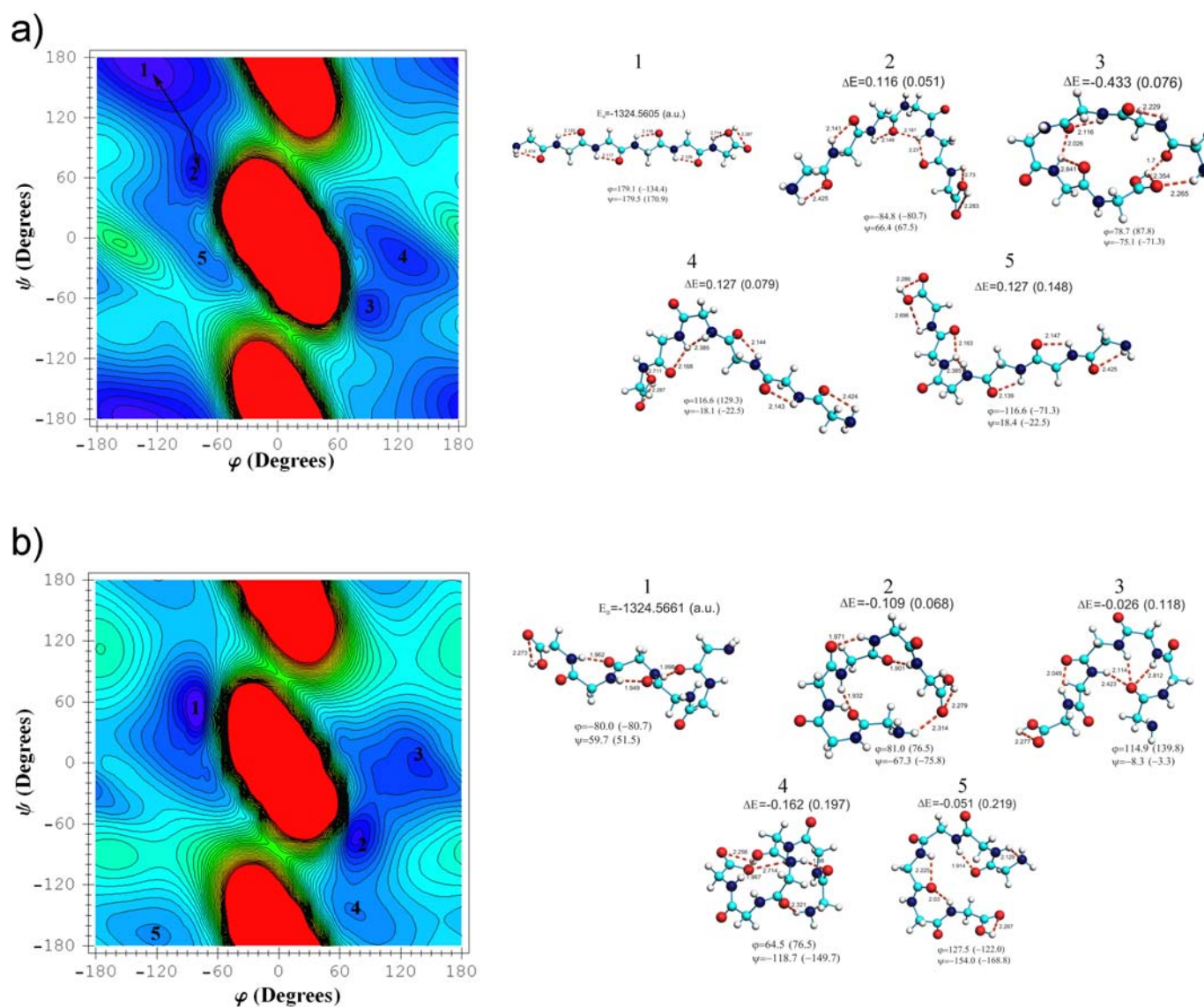
Fig. 7. Potential energy surface for the glycine hexapeptide with the sheet secondary structure (part a) and with the helix secondary structure (part b) calculated by the B3LYP/6-31G(2d,p) method. Energy scale is given in Figure 3. Numbers mark energy minima on the potential energy surface. Images of optimized conformations of the glycine hexapeptide are shown near the corresponding energy landscape. Values of angles φ and ψ, as well as the relative energies of the conformations are given analogously to that in Figure 4. A colour version of the figure is available in electronic form at http://www.eurphysj.org.
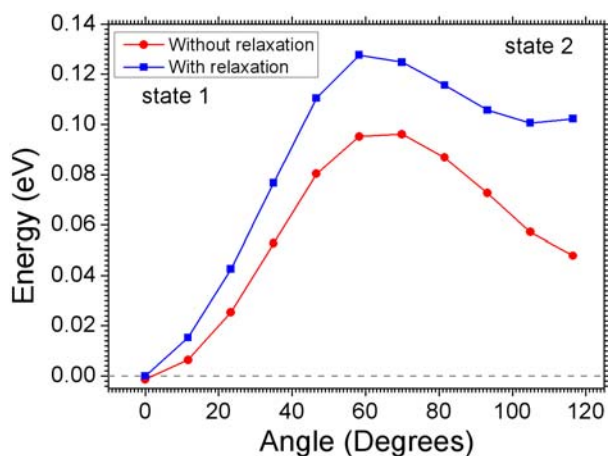

Fig. 8. Transition barriers between conformations $1 \leftrightarrow 2$ of the glycine hexapeptide with the sheet secondary structure. Circles and squares correspond to the barriers calculated without and with relaxation of all degrees of freedom in the system.

with the B3LYP/6-31G(2d,p) method. Using equation (4) one derives the transition times at room temperature: $\tau_{6 \times Gly}^{1 \to 2} \sim 305$ ps, $\tau_{6 \times Gly}^{2 \to 1} \sim 6$ ps.

It is well-known that the DFT method in its most simple formulation does not reproduce the attractive polarization (van der Waals) interaction well enough, for a discussion see, e.g., reference [40] and references therein. However the polarization interaction can partially be included in the DFT framework via the correlation functional. The degree of accounting for the polarization effects can be controlled by the choice of the number of the polarization functions included in the basis set.

The combination of the B3LYP functional with the 6-31G(2d,p) basis set, which we use for our computations accounts for the polarization interaction on the level of 50% at least. The most straightforward way of accounting for this energy correction more precisely would be the calculation performed within the framework of the perturbation theory. However, it is not feasible to perform such calculations for molecular systems like hexapeptides, because they would require enormous computer powers.

Another way of accounting for the van der Waals (vdW) interaction is based on the addition of phenomenological Lennard-Jones-type of terms to the total energy of the system [40]. Each of the potentials includes at least two parameters: the equilibrium separation distance of a pair of atoms, and the potential energy well depth. Unfortunately, there are no fixed values for these parameters which would be universally applicable in a wide scope of situations. Even for the same systems different authors choose different parameters (see e.g., [40–42]). Some arbitrariness in the in the choice of the vdW correction seems to be a general problem of the hybrid approaches.

A systematic study of the potential energy surfaces for polypeptide chains with vdW energy correction taken into account is an interesting problem, which however lies beyond the scope of the present paper, and is left open for further consideration. In this paper we only calculate the vdW interaction energies for a few key-points on the potential energy surface, in order to establish the level of accuracy of our calculations. Thus, we have estimated the vdW energy correction for the transition between the conformations 1 and 2 of the glycine hexapeptide using the set of constants suggested in [41]. The maximum energy value was obtained for the conformation 2, and is equal to 0.0093 eV, which gives the relative error of about 9%. For the state corresponding to the barrier maximum the vdW energy correction is equal to 0.0058 eV (relative error 4.5%). Note, that these energies depend strongly on the set of the chosen constants. For example, the vdW energy correction calculated with a set of constants suggested in [42] for conformation 2 is equal to 0.0076 eV, resulting in the relative error of 7.5%. These estimates demonstrate that the vdW polarization energy plays an important role giving the relative error to the calculated energies of the order of about 5–10%. Note, that this fact does not necessarily imply changes of the topology of the calculated potential energy surfaces on the same scale, be-

cause the relative variations of energies on the potential energy surface are much smaller.

Let us now consider glycine hexapeptide with the helix secondary structure. The potential energy surface for this polypeptide is shown in Figure 7b. The positions of minima on this surface are shifted significantly compared to the cases discussed above. This change takes place because of the influence of the secondary structure of the polypeptide on the potential energy surface. The geometries of the most stable conformations are shown on the right hand-side of Figure 7b.

It is worth noting that for some conformations of glycine hexapeptide the angles φ and ψ change significantly when the relaxation of all degrees of freedom in the system is accounted for (see for example conformations 1, 4, 5 in Fig. 7a and conformations 3, 4 in Fig. 7b). This means that the potential energy surface of the glycine hexapeptide in the vicinity of mentioned minima is very sensitive to the relaxation of all degrees of freedom. However, calculation of the potential energy surface with accounting for the relaxation of all degrees of freedom is unfeasible task. Indeed, one needs about 1000 hours of computer time (Pentium Xeon 2.4 GHz) for the calculation of the potential energy surface for the glycine hexapeptide. To perform an analogues calculation with accounting for the relaxation about 3 years of computer time would be needed. Nevertheless, the potential energy surface calculated without accounting for the relaxation carries a lot of useful information. Thus, one can predetermine stable conformations of polypeptide, which then can be used as starting configurations for further energy minimization.

3.5 Comparison of calculation results with experimental data

Nowadays, the structure of many proteins has been determined experimentally [9]. Knowing the protein structure one can find the angles φ and ψ for each amino acid in the protein.

In Figure 9a, we show a map of the allowed and forbidden conformations for glycine residues in poly-glycine chain taken from [45] (steric Ramachandran diagram). This map was obtained from pure geometrical considerations, in which the structure of the polypeptide was assumed to be fixed and defined by the interatomic van der Waals interaction radii. Depending on the distances between the atoms one could distinguish three regions: completely allowed, conventionally allowed and forbidden. The conformation is called completely allowed if all the distances between atoms of different amino acids are larger than some critical value $r_{ij} \geq r_{max}$. Conventionally allowed regions on the potential energy surface correspond to the conformations of the polypeptide, in which the distances between some atoms of different amino acids lie within the interval $r_{min} \leq r_{ij} < r_{max}$. All other conformations are referred to as forbidden. The values of r_{min} and r_{max} are defined by the types of interacting atoms
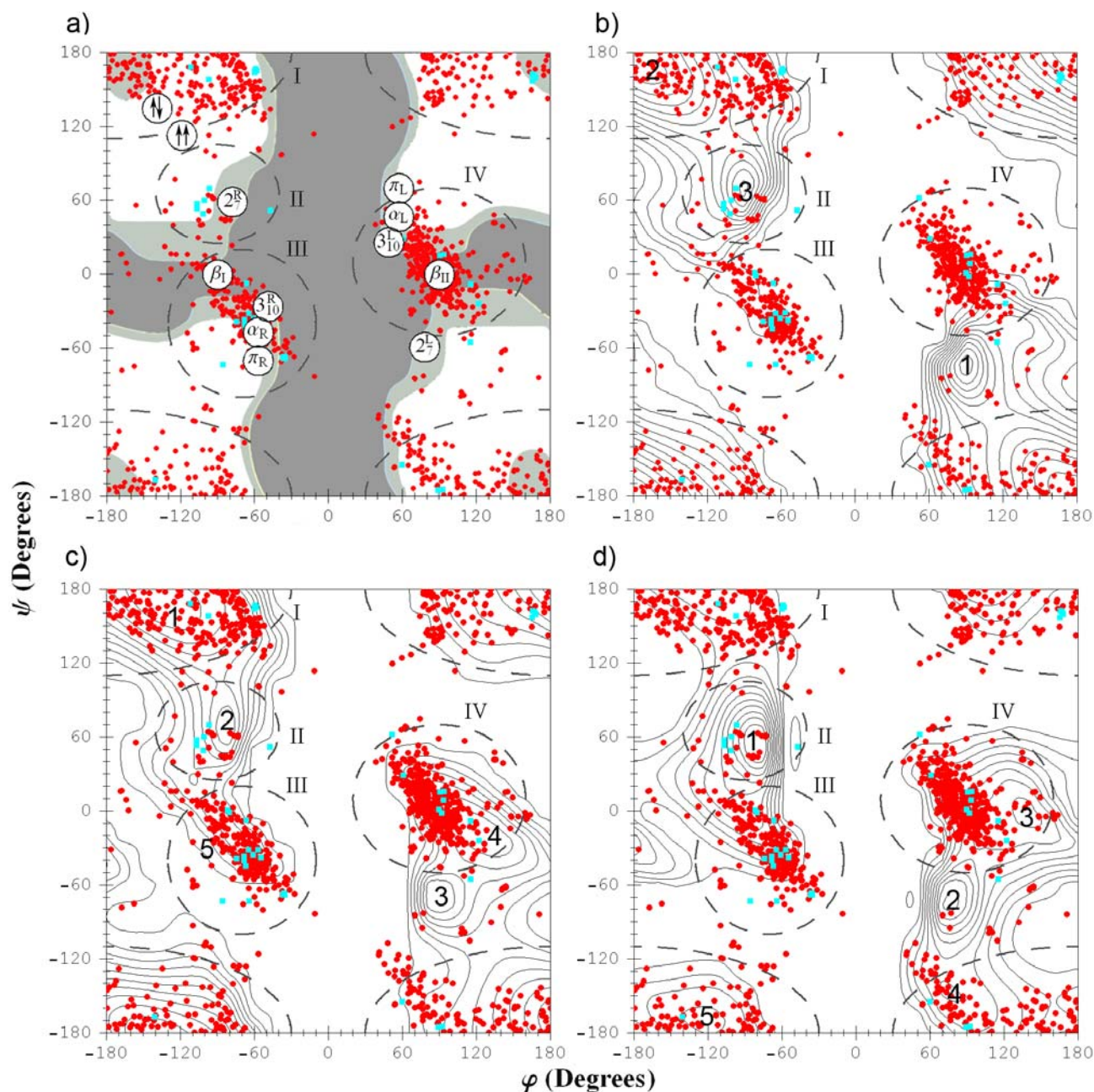
Fig. 9. Comparison of angles φ and ψ of glycine residues in protein structures selected from the Brookhaven Protein Data Bank [9,44] with the steric diagram for poly-glycine [45] (part a). Comparison of angles φ and ψ of glycine residues in protein structures selected from the Brookhaven Protein Data Bank [9,44] with the minima on the calculated potential energy surfaces for: glycine tripeptide (b); glycine hexapeptide in sheet conformation (c); glycine hexapeptide in helix conformation (d). Transparent rhomboids correspond to glycines surrounded with glycines, while filled circles correspond to glycines surrounded by other amino acids. Dashed ellipses mark the regions of higher concentration of the observed angles. A colour version of the figure is available in electronic form at http://www.eurphysj.org.

and can be found in the textbooks (see, e.g., [45]). In Figure 9a we mark the completely allowed regions with white, the conventionally allowed regions with light-gray and the forbidden regions with dark gray color. In this figure we mark the points, which correspond to the geometries of glycine, whose periodical iteration leads to the formation of chains with specific secondary structure. In Table 2 we compile the values of angles φ and ψ, which correspond to

the most prominent poly-glycine secondary structures. For the illustrative purposes we mark these points by white circles with the corresponding type of the secondary structure typed in. Thus, 2_7^R, 2_7^L are the right-handed and the left-handed 2_7 helix; 3_{10}^R, 3_{10}^L are the right-handed and the left-handed 3_{10} helix; α_R, α_L are the right-handed and the left-handed $\alpha-$helix (4_{13}); π_R, π_L are the right-handed and the left-handed $\pi-$helix (5_{16}); $\uparrow\uparrow$, $\uparrow\downarrow$ are the

Table 2. Angles φ and ψ corresponding to the most prominent poly-glycine secondary structures.

Structure type	φ (Deg.)	ψ (Deg.)
right-handed (left-handed) 2_7 helix	−78 (78)	59 (−59)
right-handed (left-handed) 3_{10} helix	−49 (49)	−26 (26)
right-handed (left-handed) α−helix (4_{13})	−57 (57)	−47 (47)
right-handed (left-handed) π−helix (5_{16})	−57 (57)	−70 (70)
parallel β sheet (↑↑)	−119	113
antiparallel β sheet (↑↓)	−139	135
β−turn of type I	−90	0
β−turn of type II	90	0

parallel and antiparallel β sheets. β_I, β_{II} correspond to the β−turns of types I and II respectively.

Note that not all of the structures listed above are present equally in proteins. In Figure 9a we show the distribution of the angles φ and ψ of glycine residues in protein structures selected from the Brookhaven Protein Data Bank [9,44]. It is possible to distinguish four main regions, in which most of the experimental points are located. In Figure 9 these regions are schematically shown with dashed ellipses. Note that these ellipses are used for illustrative purposes only, and serve for a better understanding of the experimental data. The regions in which most of the observed angles φ and ψ are located correspond to different secondary structure of the poly-glycine. Thus, region I corresponds to the parallel and antiparallel β−sheets. Region II corresponds to the right-handed 2_7^R helix. Region III corresponds to the right-handed α_R−helix, right-handed 3_{10}^R helix, right-handed π_R−helix and β−turn of type I. Region IV corresponds to the left-handed α_L−helix, left-handed 3_{10}^L helix and β−turn of type II. In some cases there are several types of secondary structure within one domain. In the present work we have not studied the secondary structure of proteins systematically enough to establish the univocal correspondence of the observed experimental points to different types of the secondary structure. Figure 9a demonstrates that π_L and 2_7^L helixes are rarely observed in experiment.

Note that some experimental points lie in the forbidden region of the steric Ramachandran diagram (see region IV in Fig. 9a). The quantum calculation shows that in fact this region is allowed and has several minima on the potential energy surface (see Figs. 9c and 9d). This comparison shows that for an accurate description of polypeptides it is important to take quantum properties of these systems into account.

Let us now compare the distribution of angles φ and ψ experimentally observed for proteins with the potential energy landscape calculated for glycine polypeptides, and establish correspondence of the secondary structure of the calculated conformations with the predictions of the simple Ramachandran model.

Region I corresponds to the minima 2, 1 and 5 on the potential energy surfaces of the glycine tripeptide (Fig. 9b), the glycine hexpeptide with the secondary structure of sheet (Fig. 9c), and the glycine hexapeptide with the secondary structure of helix (Fig. 9d) respectively.

Conformations 2 and 1 in Figures 9b and 9c correspond exactly to the glycine chains in the β-sheet conformation (see Figs. 4 and 7a). Conformation 5 in Figure 9d is a mixed state. Here the central amino acid has the conformation of sheet, while the outermost amino acids have the conformation of helix.

Region II corresponds to the minima 3, 2 and 1 on the potential energy surfaces 9b, 9c, and 9d, respectively. On the steric diagram for poly-glycine (see Fig. 9a) this region corresponds to the right-handed 2_7^R helix. The structure of conformations 3 and 2 on the surfaces 9b and 9c differs from the structure of this particular helix type. Only the central glycines, for which the angles φ and ψ in Figures 9b and 9c are defined, have the structure of 2_7^R helix. Thus, one can refer to conformations 3 and 2 as to the mixed states, where the central part of the polypeptide chain has the conformation of helix and the outermost parts have the conformation of sheet. Conformation 1 on the surface 9d is also a mixed state. Here one can distinguish one turn of 3_{10}^R helix and two turns of 2_7^R helix (see Fig. 7b).

Region III corresponds to the structure of right-handed α_R−helix, right-handed 3_{10}^R helix, right-handed π_R−helix and β−turn I. It corresponds to minimum 5 on the potential energy surface of the glycine hexapeptide with the secondary structure of sheet 9c. Conformation 5 can be characterized as partially formed β−turn, because the glycine, for which the dihedral angles φ and ψ in figure (Fig. 9c) are defined, has the geometry of β−turn and its neighbor forms a β−sheet (see Fig. 7a). There are no minima in region III on the potential energy surfaces presented in Figures 9b and 9d. This happens because in this case most probable is the structure of right-handed α_R−helix. To form one turn of the helix of this type it is necessary to link at least four amino acids, so the glycine tripeptide is too short for that. As well, six amino acids chain is too short to form a stable fragment of an α_R−helix, because it does not have enough hydrogen bonds to stabilize the structure. For the hexapeptide more probable are the elements of 3_{10}^R and 2_7^R helixes, because in these cases 2 and 3 helical turns respectively can be formed.

Region IV is represented by the structure of left-handed α_L−helix, left-handed 3_{10}^L helix and β−turn of type II. To form these structures it is necessary to have at least four amino acids, therefore there is no minima in this region on the potential energy surface of glycine tripeptide (Fig. 9b). Region IV corresponds to the conformations 4

and 3 on the surfaces 9c and 9d respectively. Conformation 4 on the surface 9c corresponds to partially formed $\beta-$turn, because the glycine, for which the dihedral angles φ and ψ in Figure 9c are plotted has the configuration of $\beta-$turn, but the neighboring aminoacid in the chain forms a $\beta-$sheet (see Fig. 7a). Conformation 3 on the surface 9d can be characterized as deformed turn of left-handed α_L-helix, which turns out to be energetically the most favorable in this region of the potential energy surface (see Fig. 7b).

Finally, we mention a few peculiarities of the calculated potential energy landscapes. At each of the potential energy surfaces discussed in this work one can see a minimum at $\varphi \sim 80^o$ and $\psi \sim -70^o$. On the steric diagram for poly-glycine this region corresponds to the left-handed 2_7^L helix. Conformations 1, 3 and 2 on the potential energy surfaces 9b, 9c, and 9d respectively partially represent this structure (see Figs. 4, 7a and 7b). The structure of conformation 4 on the potential energy surface 9d is similar to left-handed α_L-helix, but differs from it due to the short length of the polypeptide chain, resulting in significant variation of angles φ and ψ in all the residues along the chain.

4 Conclusion

In the present paper the multidimensional potential energy surfaces for amino acid chains consisting of three and six glycines has been investigated and the conformational properties of these systems with respect to the twisting of the polypeptide chain have been described. The calculations were carried out within ab initio theoretical framework based on the density functional theory accounting for all the electrons in the system. We have determined stable conformations and calculated the energy barriers for transitions between them. Using a thermodynamic approach, we have estimated times of the characteristic transitions between the conformations. It was demonstrated that the transition times lie within the picosecond region. Our estimates are compared with the available molecular-dynamics simulations results, and the correspondence between the results of the two different methods is reported. A strong barrier asymmetry between neighboring stable conformations on the potential energy surface was found.

We have compared for the first time values of angles φ and ψ for glycine residues experimentally observed in real proteins with the coordinates of minima on the potential energy surfaces. This comparison has showed that all profound minima on the potential energy surfaces correspond to the regions in which experimentally measured values of φ and ψ are located. We have also analysed how the secondary structure of polypeptide chains influences the formation of the potential energy landscapes. For the chains of six amino acids with the secondary structures of sheet and helix the influence of the secondary structure on the stable conformations of the molecule was demonstrated.

The results of this work can be utilized for modeling more complex molecular systems. For example, the suggested model for the estimation of the characteristic transition times can be used for longer polypeptide chains, also consisting of different amino acids and for estimates of time of proteins folding. It is also possible to use the results of the present work for testing the applicability and accuracy of different model approaches for the polypeptide description requiring much less computer time than ab initio calculations.

References

1. M. Karas, F. Hillenkamp, Anal. Chem. **60**, 2299 (1988)
2. F. Hillenkamp, M. Karas, Int. J. Mass Spect. **200**, 71 (2000)
3. M. Karas, U. Bahr, I. Fournier, M. Gluckmann, A. Pfenninger, J. Mass Spect. **226**, 239 (2003)
4. M. Wind, W. Lehmann, J. Anal. At. Spect. **19**, 20 (2004)
5. J. Fenn, M. Mann, C. Meng, S. Wong, C. Whitehouse, Science **246**, 64 (1989)
6. S. Bröndsted-Nielsen, J. Andersen, P. Hvelplund, B. Liu, S. Tomita, J. Phys. B: At. Mol. Opt. Phys. **37**, R25 (2004)
7. A. Finkelstein, O. Ptizin, *Physics of Proteins* (Moscow University Press "Universitet", 2002)
8. A. Mülberg, *Protein Folding* (St. Petersburg University Press, 2004)
9. H. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. Bhat, H. Weissig, I. Shindyalov, P. Bourne, Nucleic Acids Research **28**, 235 (2000)
10. A. Rubin, *Biophysics: Theoretical Biophysics* (Moscow University Press "Nauka", 2004)
11. T. Head-Gordon, M. Head-Gordon, M. Frisch, C.B. III, J. Pople, J. Am. Chem. Soc. **113**, 5989 (1991)
12. I. Gould, W. Cornell, I. Hillier, J. Am. Chem. Soc. **116**, 9250 (1994)
13. Z. Wang, Y. Duan, J. Comp. Chem. **25**, 1699 (2004)
14. A. Percel, O. Farkas, I. Jákli, I. Topol, I. Csizmadia, J. Comp. Chem. **24**, 1026 (2003)
15. I. Húdaky, P. Húdaky, A. Percel, J. Comp. Chem. **25**, 1522 (2004)
16. R. Improta, V. Barone, J. Comp. Chem. **25**, 1333 (2004)
17. R. Vargas, J. Garza, B. Hay, D. Dixon, J. Phys. Chem. A **106**, 3213 (2002)
18. R. Kashner, D. Hohl, J. Phys. Chem. A **102**, 5111 (1998)
19. D. Wei, H. Guo, D. Salahub, Phys. Rev. E **64**, 011907 (2001)
20. O. Bludský, J. Chocholoušová, J. Vacek, F. Huisken, P. Hobza, J. Chem. Phys. **113**, 4629 (2000)
21. S. Woutersen, Y. Mu, G. Stock, P. Hamm, Chem. Phys. **266**, 137 (2001)
22. S. Woutersen, R. Pfister, Y. Mu, D. Kosov, G. Stock, J. Chem. Phys. **117**, 6833 (2002)
23. Y. Mu, G. Stock, J. Phys. Chem. B **106**, 5294 (2002)

24. Y. Mu, D. Kosov, G. Stock, J. Phys. Chem. B **107**, 5064 (2003)
25. P. Nguyen, G. Stock, J. Chem. Phys. **119**, 11350 (2003)
26. H. Torii, M. Tasumi, J. Ram. Spect. **29**, 81 (1998)
27. R. Schweitzer-Stenner, F. Eker, Q. Huang, K. Griebenow, J. Am. Chem. Soc. **123**, 9628 (2001)
28. Y. Levy, O. Becker, J. Chem. Phys. **114**, 993 (2001)
29. Z. Shi, C. Olson, G. Rose, R. Baldwin, N. Kallenbach, PNAS **99**, 9190 (2002)
30. A. Garcia, Polymer **45**, 669 (2004)
31. A. Yakubovitch, I. Solov'yov, A. Solov'yov, W. Greiner, `arXiv:physics/0511026`, Eur. Phys. J. D (accepted, 2006)
32. A. Yakubovitch, I. Solov'yov, A. Solov'yov, W. Greiner, `arXiv:physics/0511036`, Phys. Rev. E (accepted, 2006)
33. *Atomic Clusters and Nanoparticles, NATO Advanced Study Institute, les Houches Session LXXIII*, les Houches, 2000, edited by C. Guet, P. Hobza, F. Spiegelman, F. David (EDP Sciences and Springer Verlag, Berlin, 2001)
34. *Latest Advances in Atomic Cluster Collisions Fission, Fusion, Electron, Ion and Photon Impact*, edited by A. Solov'yov, J.P. Connerade (World Scientific Press, 2004)
35. L. Lindgren, J. Morrison, *Atomic Many-Body Theory* (Springer-Verlag, New York, Heidelberg, Berlin, 1986)
36. P. Hohenberg, W. Kohn, Phys. Rev. **136**, B864 (1964)
37. A. Becke, Phys. Rev. A **38**, 3098 (1988)
38. C. Lee, W. Yang, R. Parr, Phys. Rev. B **37**, 785 (1988)
39. R. Parr, W. Yang, *Density-Functional Theory of Atoms and Molecules* (Oxford University Press, Oxford, New York, 1989)
40. S. Grimme, J. Comput. Chem. **25**, 1463 (2004)
41. M. Freindorf, J. Gao, J. Comp. Chem. **17**, 386 (1996)
42. M. Freindorf, Y. Shao, Th.R. Furlani, J. Kong, J. Comp. Chem. **26**, 1270 (2005)
43. A. Bax, Prot. Sci. **12**, 1 (2003)
44. S. Sheik, P. Sundararajan, A. Hussain, K. Sekar, Bioinformatics **18**, 1548 (2002)
45. D. Voet, J. Voet, *Biochemistry* (John Willey and Sons, Inc., USA, 2004)