

## Phase transition in polypeptides: a step towards the understanding of protein folding

A.V. Yakubovich<sup>a</sup>, I.A. Solov'yov<sup>a</sup>, A.V. Solov'yov<sup>a,b</sup>, and W. Greiner

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

Received 25 June 2006 / Received in final form 28 August 2006

Published online 8 November 2006 – © EDP Sciences, Società Italiana di Fisica, Springer-Verlag 2006

**Abstract.** We present a formalism which turns out to be very successful in the description of the polypeptide folding. We consider this process as a first-order phase transition and develop a theory which is free of model parameters and is based solely on fundamental physical principles. It describes essential thermodynamical properties of the system such as heat capacity, the phase transition temperature and others from the analysis of the polypeptide potential energy surface calculated within ab initio density functional theory and parameterized by two dihedral angles. This problem is viewed as a major breakthrough in the theoretical understanding of the protein folding process. Our conclusion is based on the comparison of the predictions of our theory with the results of several independent experiments.

**PACS.** 82.60.Fa Heat capacities and heats of phase transitions – 87.15.He Dynamics and conformational changes – 64.70.Nd Structural transitions in nanoscale materials – 64.60.-i General studies of phase transitions

We have developed a theory, which explains results of several experiments in which the folding process of polyalanine chains has been studied. This problem is of major current interest, because it deals with the simplified version of one of the most challenging interdisciplinary problems: the problem of protein folding. Although the number of published papers in this field during the recent years is enormous, the issue remains open, since there is still no unique approach capable of explaining all the aspects of this complex dynamical process (for review see [1,2]).

Alanine polypeptides (PPs) are an example of a system, which shows distinctive property of folding while it is simple enough for an advanced theoretical description. In this letter we focus on the theoretical description of the transition of an alanine chain from the  $\alpha$ -helix conformation to a random coil (RC) state. We treat this transition as a phase transition and develop a theory which is free of model parameters and is based solely on fundamental physical principles. This feature of our approach makes it very different from earlier theoretical considerations of this process [2–7], which all contained model parameters making them rather poor in the quantitative description of the folding process for concrete systems.

Let us consider a thermodynamically equilibrated ensemble of  $N$  equal PP chains. Each PP consists of  $n$  amino acids and can be present in one of numerous isomeric states having different energies. A group of isomeric states

with similar characteristic physical properties is called a *phase state* of the PP. Then, the *phase transition* is a transformation of the PP from one phase state to another.

The first order phase transition is characterized by an abrupt sudden change of the internal energy of the system with respect to its temperature, while heat capacity as a function of temperature has a sharp peak [8]. We study the manifestation of this peculiarity for our system. The heat capacity of the system is defined by the partition function, which can be expressed as follows:

$$Z = \left( \sum_{p \in \text{phase}} \sum_{j \in \text{state}} Z_{j_p} \right)^N. \quad (1)$$

Here  $Z_{j_p}$  is the partition function of the PP chain in a particular isomeric state  $j$  and phase  $p$ .

To construct  $Z_{j_p}$  one needs to know the Hamiltonian function of the PP chain. The Hamiltonian function is a sum of the potential, kinetic and vibrational energy terms. For a PP chain in a particular isomeric state  $j$  consisting of  $m$  atoms we obtain:

$$\begin{aligned} Z_{j_p} &= e^{-\frac{E_{j_p}}{kT}} \left[ \frac{V_{j_p} M^{3/2} \sqrt{I_{j_p}^{(1)} I_{j_p}^{(2)} I_{j_p}^{(3)}}}{4^{3m-6} \pi \hbar^{3m} \prod_{\alpha=1}^{3m-6} \omega_{j_p}^{(\alpha)}} \right] (kT)^{3m-3} \\ &= \gamma_{j_p} e^{-\frac{E_{j_p}}{kT}} (kT)^{3m-3}, \end{aligned} \quad (2)$$

where  $M$ ,  $V_{j_p}$ ,  $I_{j_p}^{(1,2,3)}$ ,  $\omega_{j_p}^{(\alpha)}$  and  $E_{j_p}$  are the mass, the specific volume, the three main momenta of inertia, the

<sup>a</sup> On leave from A.F. Ioffe Institute, St. Petersburg, Russia.

<sup>b</sup> e-mail: solovyov@fias.uni-frankfurt.de

frequencies of normal vibration modes, and the potential energy of the PP in the state  $j$  and phase  $p$ , respectively.  $k$  and  $T$  are the Boltzmann constant and the temperature respectively.  $\gamma_{j_p}$  denotes the factor in the square brackets. This equation has been derived within the harmonic approximation, which is a framework for the description of the molecule in the vicinity of its equilibrium points.

The phase transition is characterized by the transition temperature  $T_0$ , the temperature range of the phase transition  $\Delta W$  and the specific heat  $Q$ . It is possible to derive analytical expressions for these quantities if all isomeric states of a PP in a certain phase have the same energy. This model is usually referred in literature as the two-energy-level model [9]. It turns out to be very useful for the qualitative analysis of the phase transitions in PP chains. If one considers the phase transition between two such phases, the partition function (Eq. (1)) can then be rewritten in the following form:

$$Z \approx Z_0 \left[ 1 + \gamma \frac{\eta_2}{\eta_1} e^{-\frac{\Delta E}{kT}} \right]^N, \quad (3)$$

where  $Z_0$  is the partition function of the system in the first phase,  $\Delta E = E_2 - E_1$  is the energy difference between the states of the PP in two different phases,  $\eta_1$  and  $\eta_2$  are the numbers of isomeric states of the PP in the first and in the second phases respectively,  $\gamma = \gamma_2/\gamma_1$  is the coefficient depending on masses, specific volumes, normal vibration modes frequencies and momenta of inertia of the PP in the two phases.

The expression for the heat capacity in the framework of this model is given by:

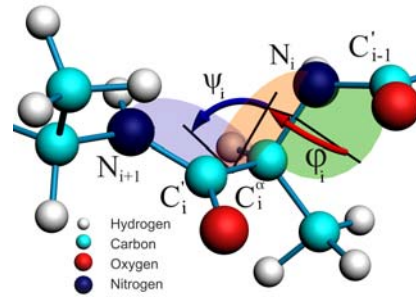
$$C(T) = kT \frac{\partial^2 T \ln Z}{\partial T^2} = \frac{N \gamma \frac{\eta_2}{\eta_1} \Delta E^2 e^{-\left(\frac{\Delta E}{kT}\right)}}{kT^2 \left( 1 + \gamma \frac{\eta_2}{\eta_1} e^{-\left(\frac{\Delta E}{kT}\right)} \right)^2}. \quad (4)$$

This function has a maximum, which corresponds to the phase transition of the system. The expressions for  $T_0$ , the corresponding maximum heat capacity value  $C_0$ ,  $\Delta W$  and  $Q$  are:

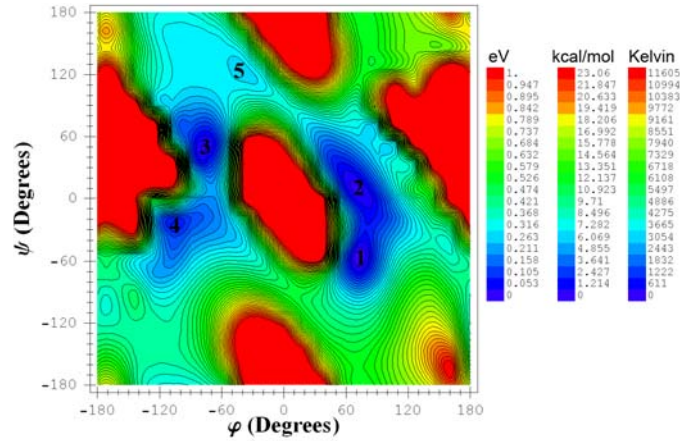
$$\begin{aligned} T_0 &\approx \frac{\Delta E}{k \ln \left( \gamma \frac{\eta_2}{\eta_1} \right)} = \frac{N \Delta E}{\Delta S}, \\ C_0 &\approx \frac{Nk}{4} \left[ \ln \left( \gamma \frac{\eta_2}{\eta_1} \right) \right]^2, \quad Q = \frac{\int C(T) dT}{m} = \frac{\Delta E}{\mu}, \\ \Delta W &\approx \sqrt{\frac{64 \ln 2}{\pi} \frac{\Delta E}{k \left[ \ln \left( \gamma \frac{\eta_2}{\eta_1} \right) \right]^2}} = \sqrt{\frac{64 \ln 2}{\pi} \frac{N^2 k \Delta E}{\Delta S^2}}. \end{aligned} \quad (5)$$

Here  $\Delta S = Nk \ln \eta_2 - Nk \ln \eta_1$  is the entropy change in the system and  $\mu$  is the mass of a single PP.

The secondary structure of a PP chain is characterized by a set of dihedral angles  $\varphi_i$  and  $\psi_i$  [10,11] (see Fig. 1). Both angles are defined by the four neighboring atoms in the PP chain. The angle  $\varphi_i$  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  $\psi_i$  is defined as the dihedral



**Fig. 1.** (Color online) Dihedral angles  $\varphi$  and  $\psi$  used for characterization of the secondary structure of a PP chain.



**Fig. 2.** (Color online) Potential energy surface for the alanine hexapeptide calculated by the B3LYP/6-31G(2d,p) method [11]. Energies are given with respect to the lowest energy minimum of the potential energy surface in eV, kcal/mol and Kelvin. Numbers mark energy minima on the potential energy surface.

angle between the  $(N_i - C_i^\alpha - C'_i)$  and  $(C_i^\alpha - C'_i - N_{i+1})$  planes. The atoms are numbered from the  $\text{NH}_2$ -terminal of the PP. The angles  $\varphi_i$  and  $\psi_i$  are defined in the interval  $[-180^\circ; 180^\circ]$ . For the unambiguous definition we count the angles  $\varphi_i$  and  $\psi_i$  clockwise, if one looks at the molecule from its  $\text{NH}_2$ -terminal (see Fig. 1). This definition is the most commonly used one [12].

The angles  $\varphi_i$  and  $\psi_i$  can be defined for all amino acids excluding the boundary ones. The variation of these angles describes the twisting of the PP chain. In this letter we demonstrate that these particular degrees of freedom are ones responsible for the dynamics of the PP chain experiencing the  $\alpha$ -helix  $\leftrightarrow$  RC phase transition.

Each amino acid in the PP chain has a number of stable conformations characterized by certain values of the angles  $\varphi$  and  $\psi$ . The potential energy surface of a PP chain calculated versus angles  $\varphi$  and  $\psi$  for any amino acid carries essential information about the conformations of this amino acid. Since we are focused on the  $\alpha$ -helix  $\leftrightarrow$  RC phase transition in polyalanine chains, in Figure 2 we present, for the sake of example, the potential energy surface for the alanine hexapeptide with the helix-like secondary structure calculated within ab initio theoretical framework using the density functional theory (DFT) accounting for all electrons in the system [11]. Such calculations allow us to conclude about the number

of conformations of each amino acid in the PP and to obtain the characteristic distribution of energies of these conformations. We demonstrate that with these data at hand one can explain reasonably well the dependence of the heat capacity of the system on temperature and thus to prove that the twisting degrees of freedom for a PP chain are responsible for the observed phase transition.

It is not feasible to study the dependence of potential energy on all possible angles  $\varphi$  and  $\psi$  for all amino acids, because the amount of computer powers required for such DFT computation would be enormous. However one can expect that for the inner amino acids of a PP the dependence of the potential energy surfaces on the twisting angles should be similar if the amino acids are loosely correlated. For alanines this condition is fulfilled, because alanines have neutral non-polar  $\text{CH}_3$ -side radicals and thus interact weakly with each other along the PP chain. The study of correlation functions for different amino acids in a PP chain is an interesting question, which is certainly worth to study for larger PP chains consisting of larger amino acids. In this paper we do not further touch this question and leave it open for further investigation.

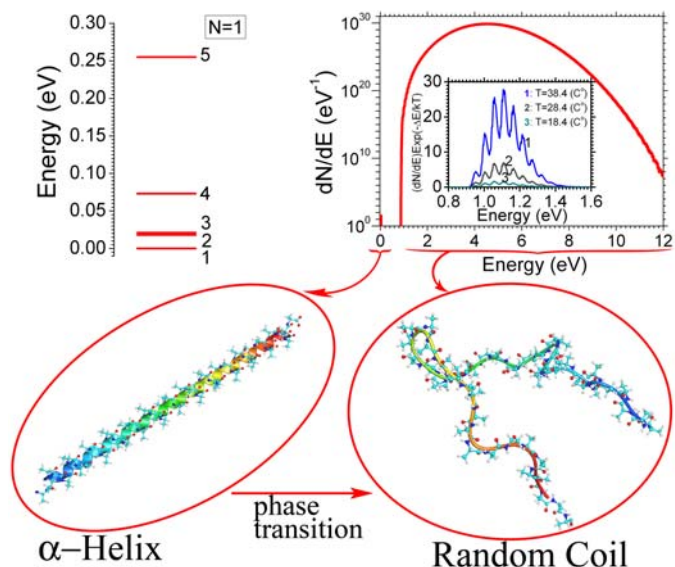
In our calculation we assume that for all amino acids in the PP chain the potential energy surfaces are topologically similar, which should be a reasonable assumption providing the correlation function between pairs of amino acids in the chain is small. In this letter we do not discuss the influence of correlation functions on the potential energy surface of PPs and leave this question open for further investigation and detalization.

Figure 2 shows that there are five minima on the potential energy surface corresponding to the stable conformations of alanine in the PP chain. In principle, one can expect that the potential energy landscape might experience some change with the growth of the PP chain length. However, the relevant minima, their number and the relative energies should vary only a little as it is seen from the comparison of the potential energy landscapes for the alanine tri and hexapeptide [11]. The exact numerical verification of this statement requires significant computer resources and will be definitely a subject for the future work. In the  $\alpha$ -helix phase the PP can be found in a single isomeric state, because in this case all amino acids in the PP are in the lowest energy conformation (see Fig. 3). In the RC phase the amino acids can occupy only the excited energy states (2–5). Thus, for a PP consisting of 50 amino acids, the number of different isomeric states of the PP is estimated to be of the order of  $4^{50}$ .

The partition function of the ensemble of  $N$  PPs, each consisting of  $n$  identical amino acids, reads:

$$Z = Z_0 \left[ 1 + \gamma \left( \sum_{j=2}^{\kappa} e^{-\frac{\epsilon_j}{kT}} \right) \right]^{nN} \quad (6)$$

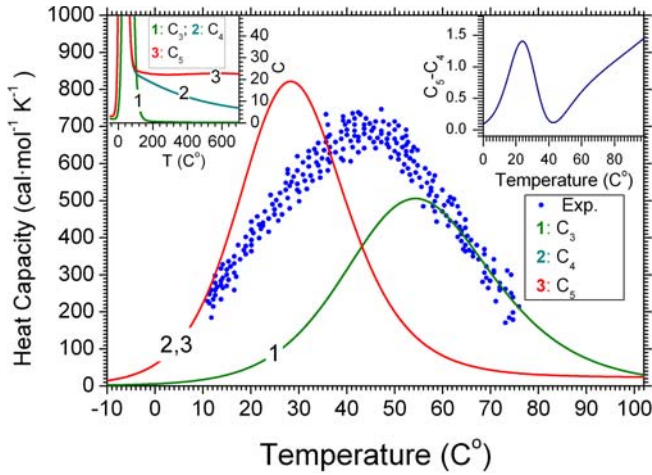
where  $\kappa$  is the number of conformations of a single amino acid (in the example considered  $\kappa = 5$ ) and  $\epsilon_j$  is the energy of the  $j$ th conformation.  $\gamma$  is the coefficient depending on the ratio of masses, specific volumes, frequencies of normal vibrations and momenta of inertia of the PP



**Fig. 3.** (Color online) Relative energies of different conformations of alanine PP obtained by twisting of a single alanine (top left). Dependence of density of states on their energy (top right). In the inset we show the dependence of  $(dN/dE) \exp(-E/kT)$  on energy at different temperatures. The characteristic structural change of the alanine PP in the phase transition (bottom).

in its two phases. For the  $\alpha$ -helix  $\leftrightarrow$  RC phase transition  $\gamma \approx 1$ , because the characteristic bond lengths in the PP are similar in both phases, the total number of atoms and the total masses of the system are exactly the same. This leads to the close values of the frequencies of normal vibrations and of the momenta of inertia in both phases.

One can expand the second term in equation (6) and end up with the following sum over all energy states of the PP in the RC state:  $\sum_i \zeta_i e^{-\frac{E_i}{kT}}$ , where  $E_i$  and  $\zeta_i$  are the energy and the degeneracy of the  $i$ th energetic state respectively. Each of the energies  $E_i$  is equal to a certain sum of energies  $\epsilon_j$  characterizing the amino acids conformations in the PP chain. Assuming that the energy of each state is smeared due to the thermal vibrations of the PP, it is possible to obtain the dependence of the energy density of states  $dN/dE$ . In Figure 3, in the right top plot we present the density of states as a function of their energy calculated for a PP consisting of 50 alanines. The second term in the brackets in equation (6) can be also rewritten as  $I(T) = \int (dN/dE) \exp(-E/kT) dE$ . This characteristic is important, because it becomes equal to one at the temperature of the phase transition. In the inset to the right top plot in Figure 3, we plot  $(dN/dE) \exp(-E/kT)$  as a function of energy. Curve 2 corresponds to the temperature of the phase transition, while curves 3 and 1 correspond to the temperatures at which the  $\alpha$ -helix and the RC states are dominant, respectively. From this plot it is also clear that the largest contribution to the integral  $I(T)$  occurs in the narrow energy range of 0.9–1.5 (eV), which means that the isomeric states of the PP with higher energies have only a minimal contribution to the partition



**Fig. 4.** (Color online) Dependence of the heat capacity on temperature. Curves 1, 2 and 3 correspond to the PP, in which three, four and five conformations of amino acids are taken into account respectively (the energies of these conformations are presented in Fig. 3 in the left top corner). Dots are the experimental results from reference [13]. In the right inset we present the difference between curves 3 and 2. In the left inset we present the behavior of the heat capacity curves in a wider range of temperatures.

function in the vicinity of the phase transition temperature.

For the description of the  $\alpha$ -helix $\leftrightarrow$ RC phase transition, it is also essential to take the boundary effects of the PP chain into account. The terminal amino acids are bound weaker than those inside the chain, and thus, in both phases they can change their conformation much easier than the central amino acids. This argument allows one to treat the terminal and the central amino acids as statistically independent. Therefore, the partition function of the whole PP can be written as a product of the partition functions of its terminal ( $Z_t$ ) and the central parts ( $Z_c$ ):  $Z = (Z_t Z_c)^N$ . Thus,

$$Z = Z_0 \left[ \sum_{j=1}^{\kappa} e^{-\frac{\epsilon_j}{kT}} \right]^{N\lambda} \left[ 1 + \left( \sum_{j=2}^{\kappa} e^{-\frac{\epsilon_j}{kT}} \right)^{n-\lambda} \right]^N. \quad (7)$$

Here  $\lambda$  is the number of amino acids in the terminal part of the PP. Their number varies in different models from one to three on each side of the PP [7,8]. We consider two boundary amino acids on each side ( $\lambda = 4$ ).

In Figure 4 we plot the heat capacity as a function of temperature calculated for the ensemble of the PP chains consisting of 50 alanines (46 central and 4 terminal). The maximum of the heat capacity defines the phase transition temperature. The pronounced maximum in the dependence of the heat capacity on temperature means that the  $\alpha$ -helix $\leftrightarrow$ RC transition is a first-order-like phase transition. Note that the 5th energy level practically does not influence on the temperature of the phase transition, because of its high energy (see curves 2 and 3 in Fig. 4).

The calorimetrically obtained experimental values for the Ac-Y(AEAAKA)<sub>8</sub>F-NH<sub>2</sub> alanine-rich peptide [13] are shown in Figure 4 by dots. Figure 4 shows that our prediction ( $T_0^{th} = 28.4$  °C for the pure alanine PP) is in a reasonable agreement with the experimental result ( $T_0^{exp} \approx 40$  °C) obtained for the alanine-rich PP. The calculated width of the peak and the maximal value of the heat capacity are also rather close to the experimental values. The discrepancies can be attributed to some extent to the differences in the systems studied.

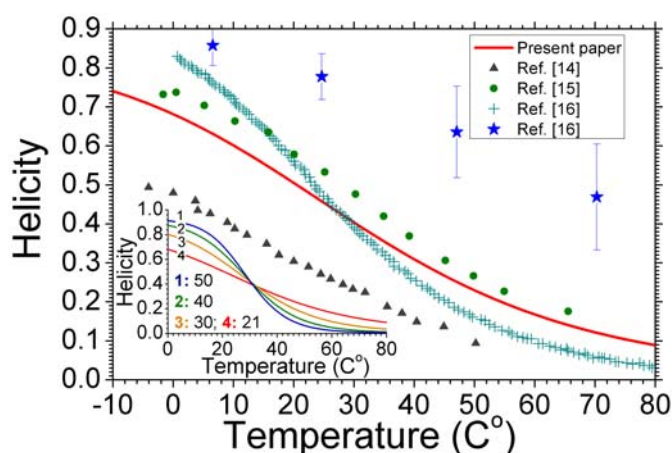
Another argument why our model predicts lower temperature of the phase transition than the actual experimental value is that the experiments were done for PPs in solution, while we perform the calculations for the system in vacuo. This fact allows one to give a qualitative justification to the discrepancy between theoretical and experimental temperatures of the phase transition. The temperature of the phase transition depends on the entropy change in the system (see Eq. (5)) and should decrease if the the entropy change grows. Let us assume that in vacuo and in solution the entropy changes are  $\Delta S^{vac}$  and  $\Delta S^{sol}$  respectively. The expression for  $\Delta S^{sol}$  reads as:

$$\begin{aligned} \Delta S_{sol} &= S_{coil}^{pp} + S_{coil}^{env} - S_{helix}^{pp} - S_{helix}^{env} \\ &= \Delta S^{pp} + \Delta S^{env}. \end{aligned} \quad (8)$$

Here  $S_{helix}^{pp}$  and  $S_{coil}^{pp}$  are the entropies of the PP in the  $\alpha$ -helix and RC states respectively;  $S_{helix}^{env}$  and  $S_{coil}^{env}$  are the entropies of the environment, corresponding to the  $\alpha$ -helix and RC phases of the PP;  $\Delta S^{pp}$  and  $\Delta S^{env}$  are the entropy changes of the PP and of the environment respectively.

The entropy of the environment is determined by the number of freely moving solvent molecules (for example H<sub>2</sub>O molecules). Some of the water molecules get attracted to the nitrogen and oxygen atoms of PP and form hydrogen bonds with them. These molecules do not contribute to the entropy of the environment but can influence the entropy of the PP. The number of surrounding water molecules is greater in the RC phase, because in this case many internal hydrogen bonds in the PP get broken, and additional hydrogen bonds with water molecules can be formed. Thus  $S_{helix}^{env} > S_{coil}^{env}$  leading to  $\Delta S^{env} < 0$ . Here we assumed that the solvent molecules in the  $\alpha$ -helix and the RC phases of the PP have the same phase state which is reasonable because the phase transition in pure solvent differs significantly from the phase transition temperature of the PP.

The entropy change of the PP is determined by the number of its possible isomeric states in both phases, which is determined by the number of minima on the potential energy surface. These minima arise mainly due to the hydrogen bonds which are formed between different amino acids in the PP chain. For the PP in solution the number of minima should be not greater than in vacuo, because the water clathrate which is formed around the PP prevents it from twisting. Thus  $\Delta S^{pp} \lesssim \Delta S^{vac}$  and  $\Delta S^{vac} > \Delta S^{sol}$  what leads to the shift in the phase transition temperature in vacuo towards smaller temperatures.



**Fig. 5.** (Color online) Dependence of the helicity on temperature (solid line). Triangles, dots and crosses are the experimental data from references [14–16], respectively. Stars are the results of molecular dynamics simulation taken from [16]. In the inset we present dependence of the helicity on temperature calculated for the alanine PP chains consisting of 21, 30, 40 and 50 amino acids.

Note, that the conformations of the amino acids with larger energies become more important at higher temperatures (see the left top inset in Fig. 4). At  $T = 600$  °C, the difference in the heat capacity between the PPs with four and five possible conformations of amino acids is more than  $10 \text{ cal mol}^{-1} \text{ K}^{-1}$ . In particular, this figure demonstrates that the high energy levels are responsible for the formation of the plateau beyond the heat capacity peak, and not the conformational fluctuations, as it was suggested in [9].

Another important characteristic of an ensemble of PP chains which can be measured experimentally, is the *helicity* of the system. The helicity of the system describes the fraction of PPs having the helix structure. The definition of helicity reads:

$$f_{\alpha} = \frac{(n - \lambda)}{n} Z_c^{-1} + \frac{\lambda}{n} Z_t^{-1}, \quad (9)$$

where  $Z_c$  and  $Z_t$  are the partition functions of the central and the terminal parts of the PP (see Eq. (7)).

In Figure 5 we present the dependence of helicity on temperature calculated for a polyalanine consisting of 21 alanines and compare it with the available experimental data and the results of other theoretical works [14–16] obtained for alanine-rich PPs. In [14,15] the experimental points were recorded for the MABA-[A]<sub>5</sub>-[AAARA]<sub>3</sub>-ANH<sub>2</sub> peptide by means of UV resonance Raman spectroscopy and by circular dichroism, respectively. In [16], the

Ac-WAAAH-[AAARA]<sub>3</sub>-A-NH<sub>2</sub> peptide was considered. Figure 5 demonstrates that the results of our calculation are in a reasonable agreement with the available experimental data. Note that the molecular dynamics simulation significantly overestimates the helicity and is in much worse agreement with experiment than the results of our theory [16].

With the growth of the chain length the steepness of the curve in the vicinity of the phase transition temperature increases, which is another evidence for the first order-like phase transition (see the inset in Fig. 5).

In this paper we have presented a new theoretical framework for the description of the  $\alpha$ -helix  $\leftrightarrow$  RC phase transition and explained the results of several independent experiments. The suggested method is rather general and with minor modifications can be applied to the description of other molecular systems experiencing phase transition, such as atomic clusters, liquids or gases. The detail investigation of all these systems is far beyond the scope of the present article and is left open for future considerations. Further development and detalization of our method is also in progress.

We acknowledge support of this work by the NoE EXCELL and are thankful to the CSC for providing us with computer facilities.

## References

1. C.M. Dobson, *Nature* **426**, 884 (2003)
2. J.E. Shea, Ch. Brooks III, *Annu. Rev. Phys. Chem.* **52**, 499 (2001)
3. T. Ooi, M. Oobatake, *Proc. Natl. Acad. Sci.* **88**, 2859 (1991)
4. N. Alves, U. Hansmann, *Phys. Rev. Lett.* **84**, 1836 (2000)
5. S. Lifson, A. Roig, *J. Chem. Phys.* **34**, 1963 (1961)
6. A. Irback, F. Sjunnesson, *Proteins* **56**, 110 (2004)
7. B. Zimm, J. Bragg, *J. Chem. Phys.* **31**, 526 (1959)
8. A. Finkelstein, O. Ptizin, *Physics of Proteins* (Moscow University Press “Universitet”, 2002)
9. N.V. Prabhu, K.A. Sharp, *Annu. Rev. Phys. Chem.* **56**, 521 (2005)
10. A.V. Yakubovich, I.A. Solov’yov, A.V. Solov’yov, W. Greiner, *Eur. Phys. J. D* **39**, 23 (2006)
11. I.A. Solov’yov, A.V. Yakubovich, A.V. Solov’yov, W. Greiner, *Phys. Rev. E* **73**, 021916 (2006)
12. A. Rubin, *Biophysics: Theoretical Biophysics* (Moscow University Press “Nauka”, 2004)
13. J. Scholtz et al., *Proc. Natl. Acad. Sci.* **88**, 2854 (1991)
14. I.K. Lednev et al., *J. Am. Chem. Soc.* **123**, 2388 (2001)
15. P. Thompson, W. Eaton, J. Hofrichter, *Biochem.* **36**, 9200 (1997)
16. G. Jas, K. Kuczeray, *Biophys. J.* **87**, 3786 (2004)