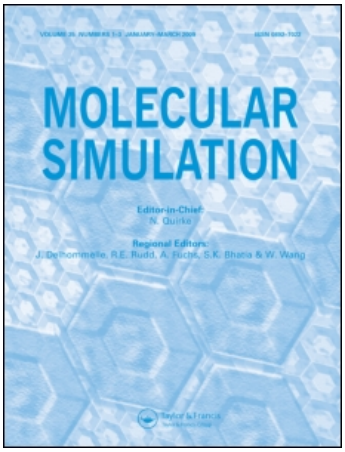

Molecular Simulation
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713644482

A Monte Carlo Study of the Relative Stability of Protein Helical Forms

M. C. Gordillo[a]; S. Lago[a]
[a] Departamento de Ciencias Ambientales, Facultad de Ciencias Experimentales, Universidad Pablo de Olavide, Sevilla, Spain

Online publication date: 26 October 2010

To cite this Article Gordillo, M. C. and Lago, S.(2003) 'A Monte Carlo Study of the Relative Stability of Protein Helical Forms', Molecular Simulation, 29: 6, 437 — 441
To link to this Article: DOI: 10.1080/0892702031000117234
URL: http://dx.doi.org/10.1080/0892702031000117234

Molecular Simulation, Vol. 29 (6–7), June 2003, pp. 437–441

A Monte Carlo Study of the Relative Stability of Protein Helical Forms

M.C. GORDILLO* and S. LAGO

Departamento de Ciencias Ambientales, Facultad de Ciencias Experimentales, Universidad Pablo de Olavide, Carretera de Utrera km 1, 41013 Sevilla, Spain

(Received November 2002; In final form February 2003)

We present simulation results on a simple model to describe the hydrogen bonding in proteins with helical structures. The approximation distinguishes between α helices, where each amino acid interacts with another one located four residues apart, 3_{10} structures, where the number of amino acids in between is three, and the π arrangement, in which that number is five. We found that the main features of the system are determined by the most stable structure (the α helix) and that the other type of hydrogen bonds appears just below the denaturation temperature of the peptide. The probability of finding a 3_{10}-type bond is greater at the beginning or at the end of the peptide chain, irrespectively of its length, while in short peptides the existence of those bonds increases appreciably the denaturation temperature, promoting stability. On the other hand, the temperature of denaturation decreases with the length of the peptide to reach a value independent of the number of amino acid residues.

Keywords: Monte Carlo; α-Helix; Peptides

INTRODUCTION

Proteins are ubiquitous components of the living matter. We usually study them by considering different hierarchical levels known as structures [1]. The primary structure is simply the account of amino acids forming the protein. Those amino acids are linked by amidic covalent bonds known as peptide bonds. In addition, they form intramolecular hydrogen bonds giving the so-called secondary structure. In a common arrangement, the i residue of the chain interacts with those in the positions $i+3$, $i+4$ or $i+5$ as a donor, and accepts the pair of electrons involved in the bond from the $i-3$, $i-4$ or $i-5$ sites. In any case, the peptide is an helix [2]. When all the bonds are of the type i, $i+4$ (and i, $i-4$), we would have an α-helix, while the molecules in which all the interactions are i, $i+3$ or i, $i+5$ are called 3_{10} and π helices, respectively. Although the first case seems to be possible in synthetic peptides (poly-alanines) [3], and in a very recent paper a case of a naturally occurring 3_{10}-only peptide has been described [4], an all π-helix has yet to be documented. However, the most common situation a majority of i, $i+4$ bonds with some i, $i+3$ in between. The i, $i+5$ links are very rare [5]. The effect of these last two types of bonds is to introduce a bend in the helical structure.

Obviously, the relative stability of the helical proteins depends on the particular amino acids forming the peptide [5]. However, the point of view adopted here is that a lot of important information about biological processes can be learned form simplified models, as some recent papers stress [6–9]. The model considered here is purely one dimensional with only one amino acid type and in which each site of the chain interacts with others through an energy term that implicitly takes into account the differences between i, $i+4$, i, $i+3$ and i, $i+5$ bonds [10]. In spite of the fact the existence of i, $i+3$ and i, $i+5$ bonds introduce bends in the structure, the one-dimensionality of our model implies that we cannot visualize directly such bends, only infer their presence. We have also studied the effect of the length of the peptide chains [11,12]. Despite their simplicity, we show that the model sheds light about the high frequency of the α-helix structure in real peptides.

*Corresponding author. E-mail: cgorbar@dex.upo.es

ISSN 0892-7022 print/ISSN 1029-0435 online © 2003 Taylor & Francis Ltd
DOI: 10.1080/0892702031000117234

METHOD

We describe a protein as a succession of N equivalent amino acids in which each of them can form only two hydrogen bonds. As mentioned above, we made the i residue interact only with the positions $i+3$, $i+4$ or $i+5$ as a donor, and with the $i-3$, $i-4$ or $i-5$ sites as an acceptor. The rest of the possible interactions are prohibited, since they are not present in the real helical structures.

Taking that into account, the energy of the protein can be written as:

$$E = \epsilon_{i,i+4}\, n_{i,i+4} + \epsilon_{i,i+3}\, n_{i,i+3} + \epsilon_{i,i+5}\, n_{i,i+5} \qquad (1)$$

where the different n's are the number of the respective type of links. To avoid double-counting, in the expression above we compute only the bonds in with the i site acts as a donor. $\epsilon_{i,\,i+4}$ was taken to be -1.1 in arbitrary units, while $\epsilon_{i,\,i+5}$ was fixed to -0.1, i.e. the difference was kept fixed to -1. To see how the particular value of $\epsilon_{i,\,i+3}$ determines the number of $i, i+3$ bonds, we varied it between -1 and -0.6. Since the appearance of a $i, i+3$ link is a relatively rare event, the energy stabilization is lower than in the case of $i, i+4$ bond, but greater than in the exceedingly rare $i, i+5$ arrangement. If $\epsilon_{i,\,i+3} = \epsilon_{i,\,i+5} = \infty$, this approach would be equivalent to the well-known Zimm–Bragg model [2], extensively used to study proteins. To solve the proposed model, we perform a Monte Carlo calculation in which we alternatively try to break or to create a bond starting from a random configuration. If the energy given by Eq. 1 above decreases, the new configuration (with a bond less or with an additional one) is taken to calculate the thermodynamic averages, if not, the proposed configuration is accepted with a probability given by the Boltzman factor. To decrease the possibilities of the system to be trapped in a metastable state at low temperatures, we performed a simulated annealing with an initial temperature of 2 and a final of 0.01, in the same arbitrary units used to define the ϵ's. For each temperature 50×10^6 equilibration Monte Carlo steps were performed. One Monte Carlo step is the result of trying to break or to form a link for each site of the chain, being the type of link ($i, i+3$; $i, i+4$; $i, i+5$) chosen at random. The averages of the thermodynamic quantities were calculated for 200,000 configurations. To avoid correlation between them, we consider only configurations separated by 1000 Monte Carlo steps. In any case, we checked that the system did not present ergodicity problems, being the simulated annealing technique enough to get the corresponding equilibrium states at any given temperature. To translate the ϵ values to energy differences, one can use a feature known experimentally, such as the almost constant denaturation temperature in big proteins (see below) [13].

The effect of the length of the peptide chain was taken into account by considering different values for N, the number of amino acids. These ranged from 20, to model a short peptide, to 150, for a short protein chain. We studied also $N = 50$ and 100 to cover the intermediate range. For longer chains, the results would be qualitatively similar than the displayed here.

RESULTS

Figure 1 shows the total number of hydrogen bonds for (from top to bottom) $N = 150$, 100, 50 and 20 amino acids in the chain as obtained from simulations. A couple of features are immediately apparent: in the lowest energy state, the total number of links is $N-4$ and there is something resembling a phase transition between $T = 0.25$ and 0.5, depending on the size and of the particular value of $\epsilon_{i,\,i+3}$. The results displayed in this figure corresponds to $\epsilon_{i,\,i+3} = -1$, but they would be very similar for different values of that parameter. The first feature is easily understandable if you consider the model given above; at low temperatures the system tends to maximize the number of most stable links, those of the type $i, i+4$, what implies that each site forms its corresponding bond. However, the last four residues in the chain cannot be donors in any bond, because there are not amino acids left. This can be seen experimentally in real proteins [14]. Incidentally, in the same way that we have four idle donors at the end of the protein, there are four impaired acceptors at the beginning of it. The fact that when the energy is lowest we always had a pure α helix explains why this structure is more common than the average in

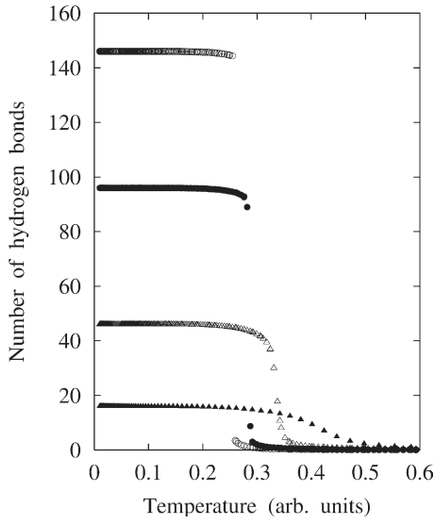

FIGURE 1 Total number of hydrogen bonds for different values of N versus temperature. From top to bottom $N = 150$, $N = 100$, $N = 50$, $N = 20$.

TABLE I T_c for different values of $\epsilon_{i,i+3}$ and the size of the system, N

	$N = 20$	$N = 50$	$N = 100$	$N = 150$
$\epsilon_{i,i+3}=-1.0$	0.424	0.338	0.287	0.265
$\epsilon_{i,i+3}=-0.9$	0.402	0.331	0.281	0.255
$\epsilon_{i,i+3}=-0.8$	0.381	0.326	0.281	0.241
$\epsilon_{i,i+3}=-0.7$	0.381	0.326	0.281	0.241
$\epsilon_{i,i+3}=-0.6$	0.381	0.326	0.281	0.241

See text for details.

the type I of antifreezing proteins found in fishes (Pleuronectidae) living in cold environments [15].

The total destruction of the links at high temperature is related to the experimental phenomenon of denaturation. In Table I and Fig. 1 we see that, in our model, the greater N, the smaller the temperature of the transition, a feature shared with the Zimm–Bragg model. However, T_c also changes with $\epsilon_{i,i+3}$, what is not predicted by that model, and can be seen in both Table I and Fig. 2. Here, we display the number of links for the two smaller chains considered ($N = 20$, triangles; and $N = 50$, circles). We can see immediately that the possibility of forming $i, i+3$ bonds increases the denaturation temperature with respect to that of the Zimm–Bragg model. This effect is greater in the smaller peptides and decreases when the difference between $\epsilon_{i,i+3}$ and $\epsilon_{i,i+4}$ increases. Thus, we observe that, for $N = 50$ and $\epsilon_{i,i+3} = -0.6$ both approaches give practically the same results, feature common to longer chains. However, for $N = 20$, the peptide would be more stable to denaturation if the possibility of creating other kind of links is allowed. This stabilization of small peptides is due to the fact that, when

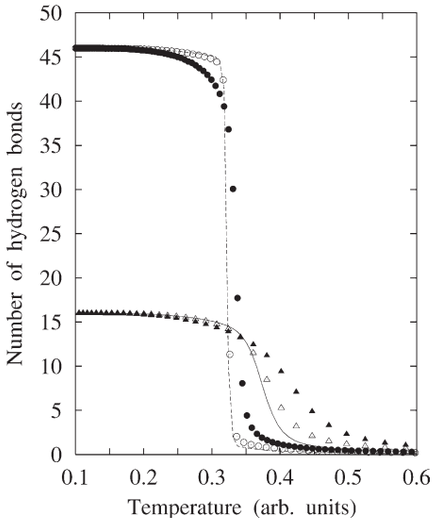

FIGURE 2 Same than in Fig. 1, but for $N = 20$ (triangles) and $N = 50$ (circles). The dark symbols correspond to our model when $\epsilon_{i,i+3} = -1$, and white ones to the case $\epsilon_{i,i+3} = -0.6$, while the lines are the results of the Zimm–Bragg approach.

the balance between the entropic term and the energy in the free energy is decided in favor of the first one at the transition, the possibility of having $i, i+3$ or $i, i+5$ links decreases de energy of the system. This implies that the inset of the disordered phase is postponed to higher temperatures. Moreover, the existence of those new linkages increases the number of bonded configurations accessible to the system, stabilizing the linked phase. However, when the number of residues increases, the effect of creating few bonds more is progressively diluted until it is negligible, even though this effect has been experimentally observed [16]. The same could be said when the energy stabilization decreases (greater values of ϵ). On the other hand, at low temperatures, the important term in the free energy is the energetic one, what precludes the existence of $i, i+3$ and $i, i+5$ bonds except around T_c. The experimental fact that T_c is practically constant for long proteins (~ 350 K), can be used to translate our arbitrary units for ϵ into standard ones, by equating that number with the T_c's for long systems. We can say then, that in our model, a $i, i+4$ bond is between 1 and 6 kJ/mol more stable than a $i, i+3$ one. To our knowledge, there is not a experimental value to compare this results to. A recent calculation for an all-atom model for polyalanine indicates that the energetic stabilization of a $i, i+4$ link in a solvent environment is between 7 and 12 kJ/mol [17]. On the other hand, a density-functional theory work on the same system suggest that that stabilization can be between ~ 15 and 36 kJ/mol per $i,i+4$ link [18]. If we use those values to deduce $\epsilon_{i,i+3}$ from the equivalence $\epsilon_{i,i+4} = -1.1$ and $\epsilon_{i,i+3} = -1.0, -0.9, \ldots, -0.6$ give above, we found that the energy difference between $\epsilon_{i,i+4}$ and $\epsilon_{i,i+3}$ ranges from 0.6 to 3.2 kJ/mol in the first case and 3.3 to 16.4 kJ/mol in the second. In any case, the $\epsilon_{i,i+3}$'s values are of the order of manitude that one expects, being the diference wth the $\epsilon_{i,i+4}$ ones about one third of the stabilization due to a $i,i+4$ bond.

The effects of the size in the amount of $i,i+3$ links can be established with the help of Fig. 3. There, we display the total number of these kinds of bonds for $\epsilon_{i,i+3} = -0.9$ and for the four values of N considered in this work. Obviously, the fractional numbers correspond to averages over many different configurations. As in previous figures, this case is representative of data for different $\epsilon_{i,i+3}$'s. A couple of facts are immediately apparent: first, when the average number of connections is one or less, and we are below the critical temperature, the results are independent of the size considered and, second, we have a significant number of $i, i+3$ links only around the transition temperatures (corresponding roughly to the maxima of the four given curves). Besides, for $N > 20$, the maximum value of $n_{i,i+3}$ decreases with the length of the peptide. This effect is

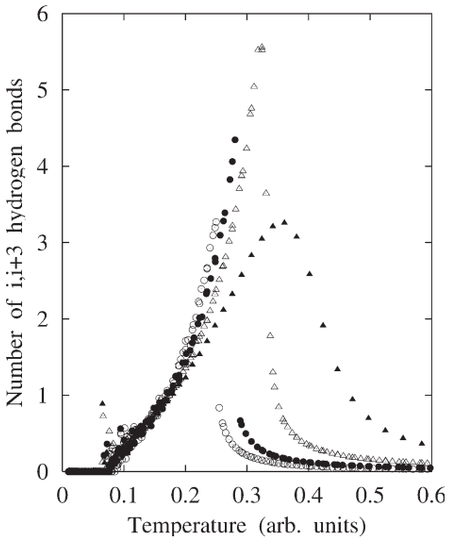

FIGURE 3 $n_{i,i+3}$ as a temperature function for $\epsilon_{i,i+3} = -0.9$ and different values of N. We display $N = 20$ (full triangles); $N = 50$, open triangles; $N = 100$, full circles; $N = 150$, open circles.

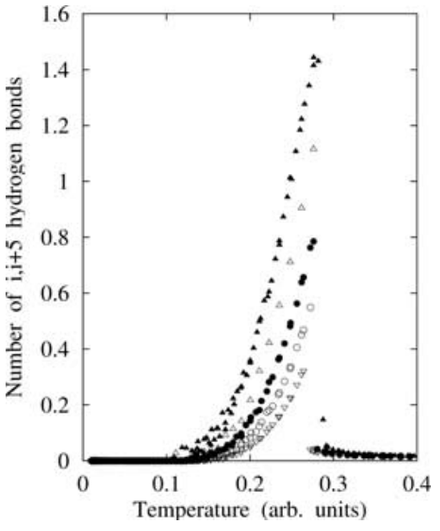

FIGURE 5 $n_{i,i+5}$ as a temperature function for $N = 100$ and different values of $\epsilon_{i,i+3}$. From top to bottom, $\epsilon_{i,i+3} = -1.0$; $\epsilon_{i,i+3} = -0.9$; $\epsilon_{i,i+3} = -0.8$; $\epsilon_{i,i+3} = -0.7$; $\epsilon_{i,i+3} = -0.6$.

a consequence of the progressive lack of effectivity of i, $i + 3$ links in the energetic stabilization of the system when the length changes mentioned above. Since in real proteins the existence of a i, $i + 3$ bond implies a bend in the structure, we can see that even in rather cold environments, the peptides are not lineal, a feature independent of the value of N.

Easily understandable is what happens when, for a given N, we change the parameter $\epsilon_{i,i+3}$. In Fig. 4 we can see it for chains with 100 amino acids. Obviously, the greater the difference between $\epsilon_{i,i+4}$ and $\epsilon_{i,i+3}$, the more difficult is the creation of $i,i + 3$ connections. The same can be said of the number of $i,i + 5$ bonds; they increase with a decreasing in $\epsilon_{i,i+5}$.

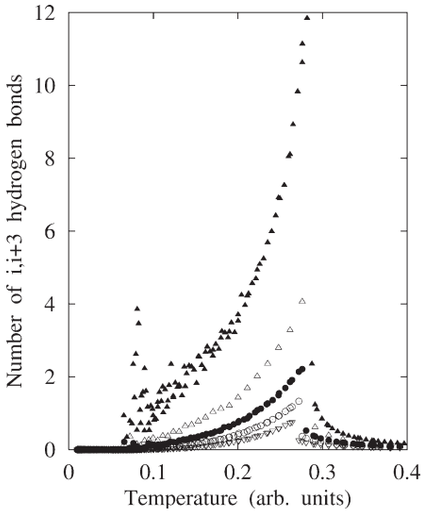

FIGURE 4 $n_{i,i+3}$ as a temperature function for $N = 100$ and different values of $\epsilon_{i,i+3}$. From top to bottom, $\epsilon_{i,i+3} = -1.0$; $\epsilon_{i,i+3} = -0.9$; $\epsilon_{i,i+3} = -0.8$; $\epsilon_{i,i+3} = -0.7$; $\epsilon_{i,i+3} = -0.6$.

In Fig. 5 we observe the same features than in Fig. 4 but for the number of $i,i + 5$ links. The only remarkable difference is the height of the peaks, basically an order of magnitude smaller than for the $i,i + 3$ bonds, consequence of bigger differences between $\epsilon_{i,i+4}$ and $\epsilon_{i,i+5}$. The dependence of $n_{i,i+5}$ on $\epsilon_{i,i+3}$ could be explained by thinking that at least part of the $i,i + 5$ links come from the transformation of a i, $i + 3$ one and not directly from the breaking of a i, $i + 4$ connection. In the same line, we do not have $i,i + 5$ bonds for $T < 0.1$, nor i, $i + 3$ ones for $T <\sim 0.07$. However, if we accept that the denaturation temperature in the case of the big peptides is around 350 K, both temperatures are well below the freezing point of water. This implies that in normal conditions we have always a certain (minimal) amount of i, $i + 5$ bonds.

Figure 6 is devoted to study the distribution of bonds in the peptides by indicating the probability of finding a i, $i + 3$ (full lines), $i,i + 4$ (dashed lines) and i, $i + 5$ (dotted lines) link as a function of the relative length of the chain. There, we display the results of $\epsilon_{i,i+3} = -1$, but the main features are common to the diffe0rent values of that parameter. The cases depicted are for $N = 20$ and $N = 100$, as representative of what happens when the length changes. The temperatures are the ones for what the maximum number of i, $i + 3$ bonds are found: 0.324 for $N = 20$ and 0.282 for $N = 100$, both very close to the denaturation temperature (see Table I). In both cases we found that the probability of having an i, $i + 3$ bond is a terminal position is very high, and that the frequency of the i, $i + 4$ links in those locations decreases accordingly. We observe also that the longer the chain, the greater the probability of a i, $i + 3$ link in a terminal position. Then, in the center

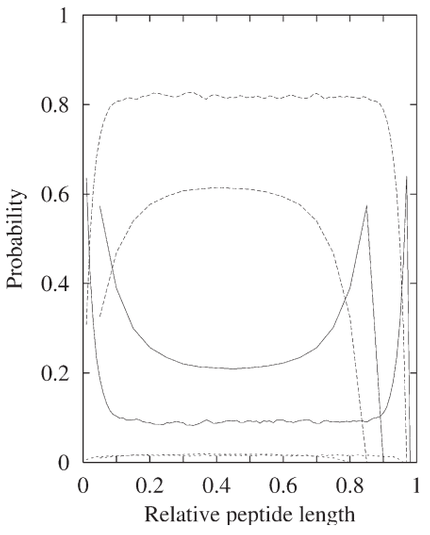

FIGURE 6 Probability of having an $i, i+3$ (full lines), $i, i+4$ (dashed lines), and $i, i+5$ (dotted lines) as a function of the relative length of the peptide chain for $N = 20$ (upper curve for $i, i+3$ bonds, lower curve for $i, i+4$ links) and $N = 100$ (lower curve for $i,i+3$ bonds, upper curve for $i, i+4$ ones) for the temperatures at which the number of $i, i+3$ bonds are maximal.

of the chain the number of $i, i+4$ bonds increases to reach a plateau, and the $i, i+3$ links diminishes, also to reach an almost constant value in the center. The $i, i+5$ linkages are kept in low numbers, even though we can observe that there are none at the extremities of the peptide. The effect of size could be ascertained too: the shorter the chain, the greater the number of $i, i+3$ bonds and the smaller the number of $i, i+4$ ones. On the other hand, the probability of having a $i, i+5$ bond is much less affected by the size of the chain. One can also see that at the end of the chains we have always several amino acids not linked.

CONCLUSIONS

We presented Monte Carlo results of a very simple model to describe proteins that could form helical structures. We observe a phase transition related to denaturation, whose temperature depends mainly on the stability of the $i, i+4$ bonds. Other type of links ($i, i+3$ and $i, i+5$) are present only around that denaturation temperature, stabilizing the linked phase by decreasing its energy and increasing the number of accessible configurations. However, this effect is only visible for small peptides, otherwise, it is diluted in the energetics of the long chains. Those $i, i+3$ links tend to accumulate in the extremes of the peptide chain, independently of the its size and on the energy difference between a $i, i+4$ link and a $i, i+3$ one. On the other hand, $i, i+5$ bonds are distributed more or less uniformly all over the protein.

Acknowledgements

This work has been supported by grants BQW2001-3615-C02-01 of the MCYT (Ministerio de Ciencia y Tecnología) and 01/1664 del Instituto de Salud Carlos III.

References

[1] Branden, C.I. and Tooze, J. (1991) *Introduction to Protein Structure* (Garland, New York).
[2] van Holde, K.E., Curtis Johnson, W. and Shing Ho, P. (1998) *Principles of Physical Biochemistry* (Prentice Hall, Englewood Cliffs, NJ).
[3] Speck, J.E., Olson, C.A., Shi, Z.S. and Kallenbach, N.R. (1998) "Alanine is an intrinsic α-helix stabilizing amino acid", *J. Am. Chem. Soc.* **121**, 5571.
[4] Biron, Z., Khare, S., Samson, A.O., Hayek, Y., Naider, F. and Anglister, J. (2002), *Biochemistry* **41**, 12687.
[5] Abagyan, R.A. (1997) In: van Gunsteren, W.F., Weiner, P.K. and Wilson, A.J., eds, *Computer Simulation of Biomolecular Systems* (Kluwer, Dordrecht) **Vol. 3**.
[6] Hansman, U.H. and Onuchic, J.N. (2001) "Thermodynamics and kinetics of folding of a small peptide", *J. Chem. Phys.* **115**, 1601.
[7] Chuang, J., Grosberg, A.Yu. and Kardar, M. (2001) "Free energy, self-averaging in protein-sized random heteropolymers", *Phys. Rev. Lett.* **87**, 78104.
[8] Micheletti, C., Banavar, J.R. and Maritain, A. (2001) "Conformations of proteins in equilibrium", *Phys. Rev. Lett.* **87**, 88102.
[9] Takano, M., Nagayama, K. and Suyama, A. (2002) "Investigating a link between all-atom model simulation and the Ising-based theory on the helix–coil transition: equilibrium statistical mechanics", *J. Chem. Phys.* **116**, 2219.
[10] Leach, A.R. (1996) *Molecular Modelling. Principles and Applications* (Prentice-Hall, Englewood Cliffs, NJ).
[11] Walther, D., Springer, C. and Cohen, F.E. (1998) "Helix–helix packing angle preferences for finite helix axes", *Proteins: Struct. Funct. Gen.* **33**, 457.
[12] Hansmann, U.H.E. and Okamoto, Y. (1999) "Finite-size scaling of helix–coil transitions in poly-alanine studied by multicanonical simulations", *J. Chem. Phys.* **110**, 1267.
[13] Ben Naim, A. (1982) *Statistical Thermodynamics of Chemistry and Biochemistry* (Plenum Press, New York).
[14] Kabsch, W. and Sander, C. (1983) "Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features", *Biopolymers* **22**, 2577.
[15] Fletcher, G.L., Goddard, S.V. and Wu, Y. (1998) "Antifreeze proteins and their genes: from basic research to business opportunity", *Chem. Tec.* **30**, 17.
[16] Dehner, A., Planker, E., Gemmecker, G., Broxterman, Q.B., Bisson, W., Fromaggio, F., Crisma, M., Toniolo, C. and Kresler, H. (2001) "Solution structure, dimerization and dynamics of a lipophilic α-3_{10}, C(α)-methylated peptide. Implications for folding of membrane proteins", *J. Am. Chem. Soc.* **123**, 6678.
[17] Peng, Y., Hansmann, U.H.E. and Alves, N.A. (2000) "Solution effects and the order of the helix–coil transition in polyalanine", *J. Chem. Phys.* **118**, 2374.
[18] Ireta, J., Neugebauer, J., Scheffler, M., Rojo, A. and Galvàn, M. (2003) "Density Functional Theory Study of the Cooperativity of Hydrogen Bonds in Finite and infinite α-helices", *J. Phys. Chem. B* **107**, 1432.