



# Studying folding $\leftrightarrow$ unfolding dynamics of solvated alanine polypeptides using molecular dynamics

Jonathan Hungerland<sup>1,a</sup> , Anders Frederiksen<sup>1,b</sup> , Luca Gerhards<sup>1,c</sup> , and Ilia A. Solov'yov<sup>1,2,3,d</sup> 

<sup>1</sup> Department of Physics, University of Oldenburg, Carl-von-Ossietzky Str. 9-11, 26129 Oldenburg, Germany

<sup>2</sup> Research Center for Neurosensory Sciences, University of Oldenburg, 26111 Oldenburg, Germany

<sup>3</sup> Center for Nanoscale Dynamics (CENAD), Institut für Physik, Carl von Ossietzky Universität Oldenburg, Ammerländer Heerstr. 114-118, 26129 Oldenburg, Germany

Received 10 June 2022 / Accepted 4 August 2022 / Published online 31 August 2022  
© The Author(s) 2022

**Abstract.** The atomistic description of the folding process of a structureless chain of amino acids to a functioning protein is still considered to be a challenge for computational biophysics. An in-depth understanding of protein-folding can be achieved through atomistic dynamics accounting for the solvent effects, combined with the theoretical description of conformational changes in shorter polypeptide chains. The present paper studies the folding transitions in short polypeptide chains consisting of 11 alanine residues in explicit solvent employing all-atom molecular dynamics. The multiple observed folding  $\leftrightarrow$  unfolding events of the peptide are interpreted as a dynamic process and rationalised through the analysis of the potential energy surface of the system. It is demonstrated that alanine polypeptide folding dynamics is governed by the backbone dihedral angles and involves the formation of spontaneously folded  $\alpha$ -helices which emerge and live for ca. 1–20 picoseconds. The helical content within the polypeptide at different temperatures was quantified through a statistical mechanics approach, which showed a reasonable agreement with the results of molecular dynamics simulations and experiments performed for alanine-rich peptides.

## 1 Introduction

Proteins constitute Nature's toolbox that, among other functions, transport and digest compounds, provide structural integrity, regulate chemical processes or serve in immune responses in a cell [1, 2]. It is striking to know that most of these functions are encoded by a set of 20 different amino acids that could fold into a specific three-dimensional structure, which defines the function of the folded protein [2]. The necessary effort to determine protein structures on an atomistic level is a challenging task, while observing the dynamics of protein folding with a similar accuracy is yet to be achieved [2]. Protein folding has interested researchers for decades as it is closely related to the fundamental understanding of protein biosynthesis and functionality [3–6].

Modern computational tools allow to perform atomistic simulations on the scale of hundreds of thousands of atoms and beyond [7–12]. However, the folding process of even the smallest proteins typically requires

simulations on the order of hundreds of microseconds [13–16]. These timescales are feasible with specialised computational infrastructure for atomistic molecular dynamics (MD) [17], but timescales of few microseconds are much more common [7–12, 18–21]. Computational architectures are thus still largely insufficient to perform all-atom protein folding simulations of arbitrary size routinely.

The problem of protein folding is conceptually related to the studies of conformational transitions in short polypeptide chains [4, 5, 22–24], and the dynamics of polypeptide chains can be described computationally at the relevant spatial and temporal scales. Several earlier studies employed *ab initio* and classical mechanics methods to investigate properties and dynamics of polypeptides in vacuum [22, 24–28]. It was shown that the backbone dihedral angles  $\phi$  and  $\psi$  (see Fig. 1C), the so-called twisting degrees of freedom, are the internal coordinates, which define the energy landscape of a polypeptide chain and permit describing the folding  $\leftrightarrow$  unfolding phase transition in polypeptides of varied length. The importance of backbone dihedral angles for the folding process is consistent with experimental data as the backbone dihedral angles are typically used to differentiate between a variety of structural conformations, i.e. the secondary structure motifs, in peptides and proteins [29–31].

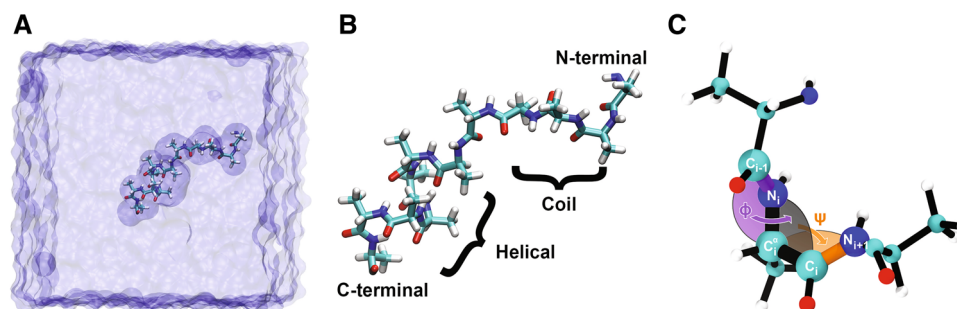
**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1140/epjd/s10053-022-00475-7>.

<sup>a</sup> e-mail: [jonathan.hungerland@uol.de](mailto:jonathan.hungerland@uol.de) (corresponding author)

<sup>b</sup> e-mail: [anders.frederiksen@uol.de](mailto:anders.frederiksen@uol.de)

<sup>c</sup> e-mail: [luca.gerhards@uol.de](mailto:luca.gerhards@uol.de)

<sup>d</sup> e-mail: [ilia.solovyov@uol.de](mailto:ilia.solovyov@uol.de) (corresponding author)



**Fig. 1** **A** Illustration of the A<sub>11</sub> peptide in a water box with some helical residues at the C-terminal and a coil structure at the N-terminal. Carbon, nitrogen, oxygen and hydrogen atoms are coloured in cyan, blue, red and light-grey, respectively. **B** Zoom-in on the peptide in **A** omitting water for clarity. **C** Definition of the backbone dihedral angles  $\phi$  (purple) and  $\psi$  (orange) of residue  $i$  inside an alanine polypeptide. The backbone atoms that determine these dihedrals are shown as larger spheres and are labelled. Atoms are coloured like in **A**

The present investigation employs an extensive all-atom MD approach to study the folding  $\leftrightarrow$  unfolding process in alanine polypeptides (see Fig. 1A) in explicit solvent. The results permit analysing the internal dynamics of the polypeptides including the temporal evolution of the backbone dihedral angles, helix-propagation and helix-stability, which provides atomistic insights into the helix-formation process highlighting a difference in the helix-propagation dynamics towards their N- and C-termini. Using steered MD, we have established the potential energy surface for an individual residue in a solvated alanine polypeptide. This potential energy surface was used to explain the observed polypeptide backbone dynamics and was employed for a statistical mechanics description of the temperature-dependent behaviour of the polypeptide's helicity for different peptide lengths.

## 2 Methods

MD simulations were performed using NAMD [7, 32] utilising the CHARMM36 force field with CMAP corrections for proteins [33–35] including explicit water solvent in a TIP3P model. Simulation preparation and analysis were done using VMD [36], where STRIDE [37] was used for the assessment of secondary structures of the polypeptides. The peptides were constructed using Pep McConst [38], interfaced through the online-platform VIKING [39].

### 2.1 Isobaric molecular dynamics of the A<sub>11</sub> peptide

A peptide consisting of 11 alanine residues (A<sub>11</sub>) was used to quantify the folding  $\leftrightarrow$  unfolding dynamics. It was constructed using Pep McConst [38] in a pure right-handed  $\alpha$ -helix conformation and solvated in a cubic water box spanning 50 Å in each direction adding up to a total of 11,661 atoms (see Fig. 1A). Periodic boundary conditions were imposed, and no ions were added corresponding to a salt concentration of 0 mol/l. The simulations were carried out using NAMD [7, 32] at a constant pressure of 1 bar using the Nosé–Hoover–Langevin pis-

ton pressure control [40, 41] with a period of 200 fs and a decay time of 50 fs. For the initial isobaric equilibration simulation of the system, the temperature was set at 300 K and controlled through the Langevin thermostat with a damping coefficient of 5 ps<sup>-1</sup>. A timestep of 1 fs was used to integrate the equations of motion for all atoms in the system. The non-bonded interactions with a smooth switching starting at 10 Å were cut off at 12 Å. Long-range Coulomb interactions were excluded to avoid artefacts from periodic boundary conditions for small simulation cells [42], especially in the case of the potential energy surface calculation.

The system was equilibrated for 250 ns. The equilibrated system was further simulated for 250 ns at 300 K to sample the data for analysis. Thereafter, the temperature of the system was increased by 2 K using the Langevin thermostat. The step-wise increase in temperature was repeated every 250 ns until the temperature of the system reached 350 K, resulting in 26 simulations at different temperature values and a total simulation time of 6.5  $\mu$ s.

### 2.2 Potential energy surface calculations

The potential energy surface was computed for a central residue inside an A<sub>5</sub> polypeptide constructed using PepMcConst [38]. The A<sub>5</sub> peptide was solvated in a cubic water box spanning 30 Å in each direction, resulting in a system containing 2930 atoms. Prior to the MD simulation, the system was minimised for 2000 conjugate gradient steps. The temperature in the simulation was set to 310 K, while all other parameters were identical to those from the simulation described above. To construct the potential energy landscape of a residue inside a polypeptide, the simulations were performed with constrained values of the dihedral angles  $\phi$  and  $\psi$  (see Fig. 1C) for all residues of the A<sub>5</sub> peptide. By constraining all backbone dihedral angles, the computed energy surface describes the residue-specific interactions in a segment of five or more residues with identical backbone dihedral angles. The potential energy surface for a residue was then computed as a function of these angles. Test calculations showed that five residues need to be constrained in order to produce potential energy

surfaces, which match the known statistical distributions of the backbone dihedral angles. The dihedral angles were constrained using a harmonic potential of the form:

$$U = k \left( \sum_{i=2}^5 (\phi_i - \phi_{\text{ref}})^2 + \sum_{i=1}^4 (\psi_i - \psi_{\text{ref}})^2 \right), \quad (1)$$

where the force constant  $k$  was assumed 200 kcal/mol in order to tightly control the dihedral angles in the simulation. The summations in Eq. (1) are performed over the residues in the peptide;  $(\phi_i, \psi_i)$  denote the dihedral angle values for the  $i$ th residue (see Fig. 1C). The summations account for the fact that  $\phi$  is not defined for the first residue (N-terminal residue), while  $\psi$  is not defined for the last residue (C-terminal residue) in the polypeptide. The reference dihedral angles  $(\phi_{\text{ref}}, \psi_{\text{ref}})$  were initially set to be  $\phi_{\text{ref}} = -60^\circ$  and  $\psi_{\text{ref}} = -60^\circ$  for the equilibration simulation, which was carried out for 10 ns in the isobaric ensemble and another 10 ns in the isochoric ensemble.

To obtain the potential energy surface of a residue inside the  $A_5$  peptide, the reference dihedral angles  $\phi_{\text{ref}}$  and  $\psi_{\text{ref}}$  were incrementally and independently altered by  $4^\circ$  every 2 ns. First, the reference dihedral angle  $\psi_{\text{ref}}$  was changed, while  $\phi_{\text{ref}}$  was kept fixed at  $-60^\circ$ , followed by a change of  $\psi_{\text{ref}}$  at a fixed  $\phi_{\text{ref}}$  value. The calculations resulted in multiple simulations where the values of the reference dihedral angles were separated by  $4^\circ$ . Each of these simulations was 2 ns long corresponding to a total simulation time of 16.2  $\mu$ s.

The potential energy for a residue was calculated for each combination of the angles  $\phi$  and  $\psi$  as the time-averaged energy of the central residue in the  $A_5$  peptide. The contribution of the constraining potentials was excluded from the potential energy calculations.

## 3 Results

### 3.1 Conformational dynamics in the $A_{11}$ peptide

Backbone dihedral angles  $\phi$  and  $\psi$  are the internal degrees of freedom that permit separating different conformations of the polypeptide [29, 30]. Figure 2A illustrates the probability density distribution of the dihedral angles recorded for the simulated  $A_{11}$  peptide. All residues inside the peptide, except the terminal ones, experience similar dynamics and have therefore contributed equally to the presented probability distribution. The distribution qualitatively resembles the distribution of dihedral angles found in experimental measurements of protein conformations [29, 30] by means of position and shape of the four most dominant backbone conformations. Regions with probability densities above  $10^{-5}$  degrees<sup>-2</sup> correspond to typical conformations that are adopted by the residues of proteins [31, 43]. These regions are used for the definition of the four important domains in the  $(\phi, \psi)$ -plane. The  $\alpha_R$

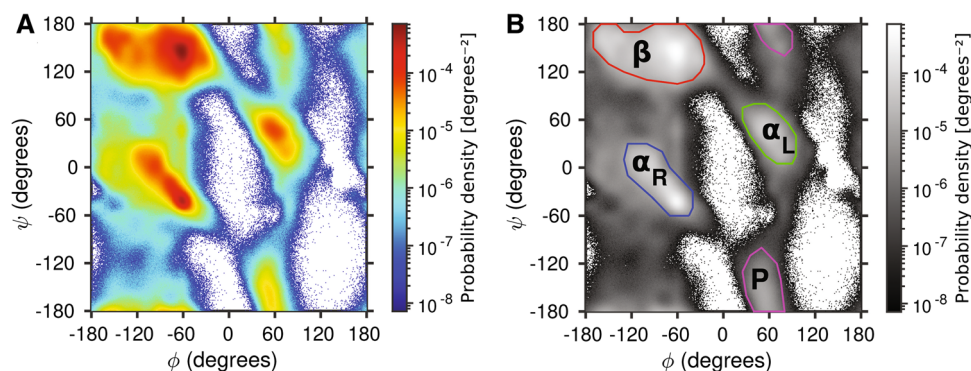
region embraces the right-handed  $\alpha$ -helices, the  $3_{10}$ -helices and the  $\pi$ -helices [29]. The  $\beta$  region includes the  $\beta$ -strands and the  $P_{II}$ -spirals [29]. The  $\alpha_L$  region is associated with the left-handed  $\alpha$ -helices [29], while the remaining  $P$  region describes the left-handed version of the  $P_{II}$ -spirals [29]. Certain combinations of the  $\phi$  and  $\psi$  angles are sterically forbidden for a polypeptide resulting in the white areas shown in Fig. 2A.

The studied  $A_{11}$  peptide was found in different conformations, which are characterised by the different backbone dihedral angles as defined in Fig. 2B.

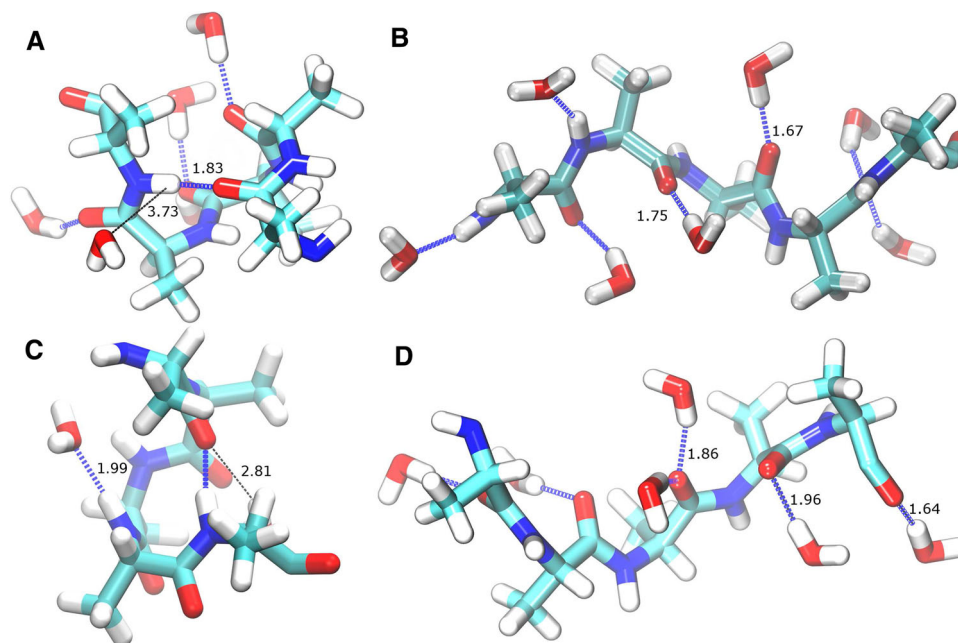
Figure 3 illustrates peptides associated with the four backbone conformations that are defined in Fig. 2 including the peptide-peptide hydrogen bonds for the two helical conformations. Here it can be seen that the two  $\alpha$  conformations (Fig. 3A, C) can be recognised by the peptide-peptide hydrogen bond, whereas as shown in Fig. 3B, D, the  $P$  and  $\beta$ -conformations do not have any peptide-peptide hydrogen bonds due to the stretched nature of these conformations. Figure 4A illustrates the probabilities for the residues in the peptide to occupy these different conformations. The figure shows that the most probable conformation is the  $\beta$ -conformation with an average probability computed for all residues (except the terminal ones) being 0.6 at any temperature within the studied range. The other occurrence probabilities in decreasing order are the  $\alpha_R$ ,  $\alpha_L$  and  $P$  conformations. Remarkably, the probabilities for the residues in a peptide to occupy the other conformations also do not feature any pronounced temperature dependence in the temperature range of 300–350 K. Fluctuations in the occurrence probabilities of  $\beta$  and  $\alpha_R$  conformations show a certain correlation because residues in the  $A_{11}$  peptide most often switch between these two conformations (see Fig. 4B). Figure 4A shows that a small residual probability of about 0.1 is observed, indicating that some residues in the peptide are found without any definite secondary structure (denoted as “other” in Fig. 4A). Table 1 compares the relative occupation of these states found in this work with results for  $A_2$  peptides. Here, the relative occupation of the helical conformations  $\alpha_R$  and  $\alpha_L$  is significantly larger than for a  $A_2$  peptide because the latter is too short to form stable helical structures.

Figure 4B shows the rate of conformational changes such as the  $\alpha_R \rightleftharpoons \beta$  transitions. The rate for the  $\alpha_R \rightleftharpoons \beta$  transitions is defined as the sum of the rates computed for the  $\alpha_R \rightarrow \beta$  and  $\beta \rightarrow \alpha_R$  transitions. All transition rates were found to be, with negligible deviation, in a detailed balance, as expected for the sufficiently sampled peptide at the equilibrium.

The transition rates in Fig. 4B show that the overall peptide is most often found in a coil-like state in which the residues explore the  $(\phi, \psi)$ -plane. This coil-like behaviour is reflected in a frequent change of conformations of residues in the peptide. Summing all conformational change rates reveals the temperature-dependent timescale of the conformational changes. At 300 K, a conformational change of some residue is observed about twice per nanosecond and at 350 K; it turns out to be about 10 changes per nanosecond.



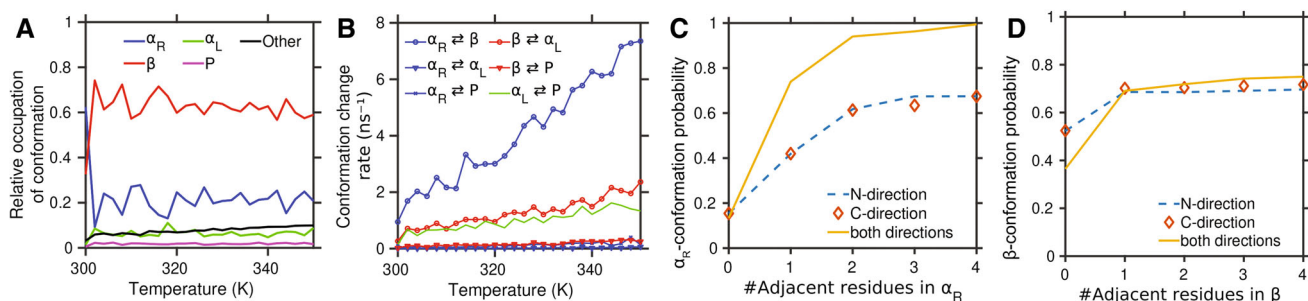
**Fig. 2** Possible conformations of alanine polypeptides. **A** Probability distribution of the backbone dihedral angles computed for the A<sub>11</sub> peptide in solution. The white area indicates the sterically forbidden regions of conformations that could not be realised. **B** Definition of the different dihedral conformations in the  $(\phi, \psi)$ -plane according to the regions with probability densities above  $10^{-5}$  degrees<sup>-2</sup> shown in **A**. Coloured contours encapsulate the dihedral angles associated with different peptide conformations (see text)



**Fig. 3** Examples for the four dominating backbone conformations of alanine polypeptides. Hydrogen bonds between backbone atoms and between the peptide and some exemplary water molecules are drawn with dashed lines, while their lengths are given in Ångström. The coordinates of the structures are taken from the potential energy surface calculations. The conformations are sorted as: **A**  $\alpha_R$  conformation, **B**  $\beta$  conformation, **C**  $\alpha_L$  conformation and **D**  $P$  conformation

The transitions  $\alpha_R \rightleftharpoons \beta$  occur most frequently and are temperature-dependent. The  $\beta \rightleftharpoons \alpha_L$  transitions or the  $\alpha_L \rightleftharpoons P$  transitions occur about once per nanosecond, while the transitions between the  $\alpha_R \rightleftharpoons \alpha_L$  conformations or between  $\alpha_R \rightleftharpoons P$  conformations are rather rare. Note that about 80% of the residues in the A<sub>11</sub> peptide are found in the left half of the  $(\phi, \psi)$ -plane, i.e. have  $\phi < 0$ . Therefore, conformational changes most likely occur clockwise in the  $(\phi, \psi)$ -plane, involving one of the  $\alpha_R \rightarrow \beta \rightarrow \alpha_L \rightarrow P$  transitions, or anti-clockwise, involving one of the  $P \rightarrow \alpha_L \rightarrow \beta \rightarrow \alpha_R$  transitions. Thus, the seemingly unordered dynamics of a solvated A<sub>11</sub> peptide follows certain rules and is in fact

not entirely random. Table 2 lists the expected transition rates between the conformations. A comparison with studies on A<sub>2</sub> peptides in Table 3 reveals that the A<sub>11</sub> peptide considered here performs torsional transitions at least one order of magnitude slower than the A<sub>2</sub> peptide. Furthermore, the fact that transition rates reported in the earlier studies [44–46] are not in detailed balance indicates that their respective transition rates were not converged. Thompson *et al.* experimentally observed helix propagation rates on the order of  $h = 0.1 \text{ ns}^{-1}$  [47], which is 1/5 of the  $k(\beta \rightarrow \alpha_R)$  rate computed here and thus fulfils the requirement  $h < \sum_x k(x \rightarrow \alpha_R)$ , where  $x$  denotes different con-



**Fig. 4** Statistics of  $A_{11}$  polypeptide dihedral conformations and their influence on the adjacent residues. **A** Average occupation of different conformations by a residue in the  $A_{11}$  peptide averaged over time at specific temperature. A dihedral angle outside any of the regions defined in Fig. 2B is referred to as “other”. **B** Average rate of conformational transitions as a function of temperature. The rate of conformational transitions was computed for all residues and was averaged over the simulation time span at a specific temperature. **C** Conditional probability of a residue to have the  $\alpha_R$  conformation once a certain number of adjacent residues also has the  $\alpha_R$  conformation. The lines differentiate the direction in which the neighbouring residues are considered: the N-direction counts towards the N-terminal of the peptide and the C-direction counts towards the C-terminal of the peptide. **D** Similar to C but for the  $\beta$  conformation

formations. Needless to say, this inequality must hold since a helix can only propagate as fast as a residue can change its backbone dihedral conformation but not all changes will lead to stable helix propagation. Previously, inconsistencies of experimental helix propagation rates [47–50] to computational rates [47, 51, 52] were found since the latter were at least two orders of magnitude faster, which is not the case here.

Figure 4C, D illustrates the correlation of residues in the  $A_{11}$  peptide in terms of conditional probabilities of observing a residue in the  $\alpha_R$  or  $\beta$  conformation depending on the conformation of the neighbouring residues. Figure 4C shows the probability for a residue to have the  $\alpha_R$  conformation once the adjacent residues are also found in the  $\alpha_R$  conformation. Excluding the terminal residues in the peptide, each datapoint corresponds to residues with a different number of neighbouring residues in a certain conformation. Residues 3–9 were used to evaluate the conditional probabilities where a residue had 0 or 1 adjacent residues in a certain conformation, whereas for 4 adjacent residues only the middle residue 6 was used, due to the necessity of having sufficient neighbours.

For an arbitrary residue in the  $A_{11}$  peptide, the blue dotted line and the orange squares in Fig. 4C reveal that there is a probability of about 0.2 that the residue is found in the  $\alpha_R$  conformation once either the preceding or the following residue has a different conformation. The probability for a residue to be in the  $\alpha_R$  conformation increases to 0.4 once one neighbouring residue is also found in the  $\alpha_R$  conformation, and the probability becomes 0.6 once two preceding or following residues are found in the  $\alpha_R$  conformation. The yellow line in Fig. 4C shows that the probability of a residue to be in the  $\alpha_R$  conformation is at 0.2 once neither of its neighbouring residues are found in the  $\alpha_R$  conformation. However, the probability increases to 0.75 if the neighbouring residues from both sides of a residue are found in the  $\alpha_R$  conformation.

The results in Fig. 4C show the effect of the  $\alpha_R$  conformation of adjacent residues in the N-terminal direction or the C-terminal direction to a given residue. The effect converges once 2 adjacent residues have an  $\alpha_R$  conformation which indicates that at least 3 residues are necessary to initialise an  $\alpha$ -helix. As soon as 3 residues nucleate into a helix, the helical structure can extend further with an equal probability in any direction. The probability of a residue to be found in the  $\beta$  conformation (Fig. 4D) shows a relatively smaller correlation to adjacent residues in the  $\beta$  conformation and already saturates as soon as a residue has a single adjacent residue in the  $\beta$  conformation. In the case that no residues in both directions are found in the  $\beta$  conformation, the probability for the middle residue is at 0.4, which is even lower than the respective probability in only N-direction or C-direction (0.5). If both neighbours appear not to be in the  $\beta$  conformation, they are most likely to be in the  $\alpha_R$  conformation, in which case the middle residue is found in the  $\alpha_R$  with a probability of 0.8 (see Fig. 4C). The strong correlation of residues in  $\alpha_R$  conformation thus decreases the probability of finding the middle residue in the  $\beta$  conformation when both of its neighbours are not in the  $\beta$  conformation.

Figure 5A–D shows distributions of the so-called dihedral difference computed for residues 3–9 of the  $A_{11}$  polypeptide. The dihedral difference quantifies the proximity of the backbone dihedral angles of different residues. Once two neighbouring residues appear in the same conformation, the dihedral difference is small but is expected to increase once the residues adopt different conformations. All residues in a helical fragment of the peptide have similar backbone dihedral angles (within  $\alpha_R$ ), so the dihedral differences within that fragment are expected to be minor (i.e. ca.  $20^\circ$ ).

The dihedral difference between two sets of dihedral angles  $(\phi_i, \psi_i)$  and  $(\phi_j, \psi_j)$  can formally be defined as a Cartesian two-dimensional distance which accounts for the periodicity of the angles  $\phi$  and  $\psi$  as follows

$$d(i, j) = \min_{k, l \in \{0, 1\}} \dots$$

**Table 1** Relative population of different states from this work and from other works

Refs.	$\alpha_R$	$\beta$	$\alpha_L$	$P$	Other
Bolhuis [44]	0.0946	0.8962	0.0042	0.0048	
Chodera [45]	0.1028	0.8492	0.0017	0.0013	
This work	0.2070	0.6310	0.0655	0.0183	0.0782

For this work, the means of data in Fig. 5A except the point at 300 K are used. Reference [44] presents data from Multiple State Transition Interface Sampling of an A<sub>2</sub> peptide, and Ref. [45] presents data from Replica Exchange MD of an A<sub>2</sub> peptide

**Table 2** Transition rates  $k$  between different states from this work given in units of ns<sup>-1</sup> for a temperature of 300 K

$k(\alpha_R \rightleftharpoons \beta)$	$k(\alpha_R \rightleftharpoons \alpha_L)$	$k(\alpha_R \rightleftharpoons P)$	$k(\beta \rightleftharpoons \alpha_L)$	$k(\beta \rightleftharpoons P)$	$k(\alpha_L \rightleftharpoons P)$
1.0916	-0.0056	0.0164	0.5047	0.0566	0.4835

The data in Fig. 5B were taken except for the point at 300 K and fitted with a linear function, which was then evaluated at 300 K.  $k(\alpha_R \rightleftharpoons \alpha_L)$  is an artifact of the fit due to low transition rates. Can be taken as 0

**Table 3** Transition rates  $k$  between different states from this and other works given in units of ns<sup>-1</sup>

Refs.	$k(\alpha_R \rightarrow \beta)$	$k(\beta \rightarrow \alpha_R)$	$k(\alpha_R \rightarrow P)$	$k(P \rightarrow \alpha_R)$
Bolhuis [44]	25.64	3.057	0.6027	17.76
Chodera [45]	33.5	4.6	0.1	18.5
Chekmarev [46]	35.1	1.10	0.906	13.9
This work	0.5455	0.5461	0.0087	0.0077

References [44–46] feature calculations using the A<sub>2</sub> peptide. The respective authors employed Multiple State Transition Interface Sampling [44], Replica Exchange MD [45] and Brownian dynamics within a potential of mean force [46]

$$\left( \sqrt{(|\phi_i - \phi_j| - k \cdot 360^\circ)^2 + (|\psi_i - \psi_j| - l \cdot 360^\circ)^2} \right). \quad (2)$$

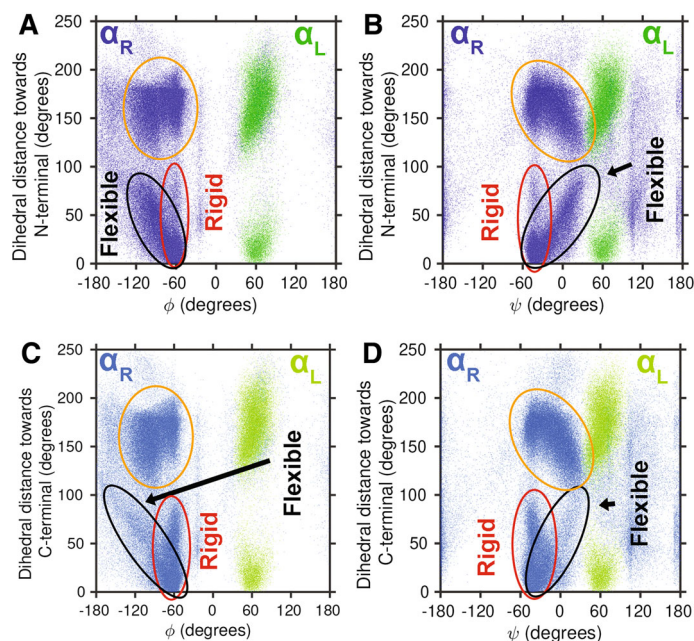
Here,  $\phi_i$  and  $\psi_i$  are the dihedral angles measured for residue  $i$ , while  $\phi_j$  and  $\psi_j$  are the dihedral angles measured for residue  $j$ . The definition of  $d(i, j)$  respects the periodicity of the dihedral angles  $\phi$  and  $\psi$  and ensures that, for example, the dihedral difference between  $(170^\circ, 0^\circ)$  and  $(-170^\circ, 0^\circ)$  yields a value of  $20^\circ$ . The dihedral difference for a residue  $i$  calculated towards the N-terminal is defined as  $d(i, i-1)$ , while the dihedral difference calculated towards the C-terminal is  $d(i, i+1)$  since the counting of residues in a polypeptide always goes from the N-terminal to the C-terminal direction.

Figure 5A–D shows the dihedral differences calculated for residues 3–9 where the residues were found in the  $\alpha_R$  (blue),  $\alpha_L$  (green) conformations or without any particular structural motif. The datapoints for the latter scenario were plotted using the same colour (blue) as the datapoints for residues with the  $\alpha_R$  conformation to ensure that the observations about the formation of  $\alpha_R$  motifs in the peptide are a consequence of the internal dynamics of the polypeptide and not a consequence of a very particular definition of the  $\alpha_R$  conformation. Similar results for other polypeptide conformations, as well as three-dimensional plots showing the  $d(i, i \pm 1)$  dependence on  $\phi$  and  $\psi$ , are presented in Figs. S1 and S2.

The data points shown in blue in Fig. 5 correspond to residues found in the  $\alpha_R$  conformation, thereby

being part of a helical fragment in the A<sub>11</sub> peptide. Helix propagation is the process where a helical fragment extends because residues neighbouring the fragment also adopt the  $\alpha_R$  conformation. During the helix propagation, the dihedral difference between the last helical residue  $i$  and the first non-helical residue decreases. This decrease can either be accompanied by fluctuations in the backbone dihedral angles of the last helical residue (flexible helix-propagation, black ellipsoids in Fig. 5) or keeping the backbone dihedral angles of the last helical residue more consistent (rigid helix-propagation, red ellipsoids in Fig. 5). For helix-propagation in the N-terminal-direction (see Fig. 5A, B), the flexible helix-propagation is more pronounced, while helix propagation in the C-terminal direction (see Fig. 5C, D) is dominated by the rigid helix-propagation. It is thus clear that there is a difference in the typical dynamics of helix propagation, although helix propagation is expected to occur with an equal probability in either direction (see Fig. 4C). The asymmetry in helix propagation is related to the asymmetry of the amino acids' backbone. For the left-handed helices ( $\alpha_L$  conformation, green datapoints), the effect has not been observed.

The orange ellipsoids in Fig. 5A–D highlight the data points where a certain residue  $i$  has a helical conformation, while its neighbouring residue in N-terminal direction (see Fig. 5A, B) or in C-terminal direction (see Fig. 5C, D) does not. In these cases, a residue  $i$  with the  $\alpha_R$  conformation has the dihedral distance to the



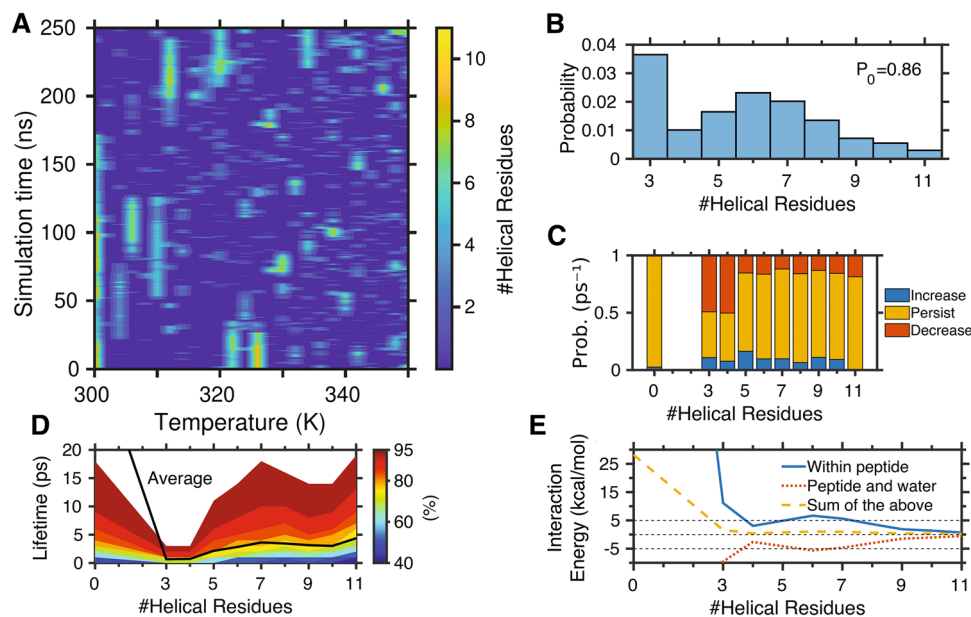
**Fig. 5** Distribution of dihedral differences. **A** Dihedral difference  $d(i, i - 1)$ , Eq. (2), plotted over  $\phi$  for a residue  $i$  with respect to its neighbouring residues in the N-terminal direction. The residues  $i \in \{3, \dots, 9\}$  from the  $A_{11}$  peptide were considered. MD frames, where the residue  $i$  is either in the  $\alpha_R$  (blue),  $\alpha_L$  (green) or outside any of the conformations defined in Fig. 2B (same blue as  $\alpha_R$ , see text for an explanation), are included in the plot. Ellipsoids mark datapoints that correspond to non-helical neighbouring residues (orange), flexible helix-propagation (black) and rigid helix-propagation (red). **B** Similar to **A**, but showing  $d(i, i - 1)$  as a function of  $\psi$ . **C** Dihedral difference  $d(i, i + 1)$  as defined in Eq. (2) plotted over  $\phi$ . Otherwise similar to **A**. **D** Similar to **C**, but showing  $d(i, i - 1)$  as a function of  $\psi$

neighbouring residue  $j$  of  $d(i, j) > 100^\circ$ . Such a condition can be illustrated by considering circles in the  $(\phi, \psi)$ -plane centered somewhere at the  $\alpha_R$  region and having a radius of  $100^\circ$  (see Fig. S3). When  $(\phi_i, \psi_i)$  is the center of such a circle, then  $(\phi_j, \psi_j)$  lying outside of the circle is logically equivalent to  $d(i, j) > 100^\circ$ . Thus, the datapoints which are encapsulated by the orange ellipsoid lie outside any of these circles.

A suitable description of the overall conformation of a peptide can be achieved through defining its helical content. The simulated  $A_{11}$  peptide was initialised as an ideal  $\alpha$ -helix. The initial helical structure becomes distorted after around 170 ns at 300 K (420 ns if equilibration is included) and, over the course of the performed MD simulation, multiple folding and unfolding events were observed (see Fig. 6A). The performed analysis illustrates that the  $A_{11}$  peptide does not form long-lived stable helices in the studied temperature range. With increasing temperature, helical fragments occurred more often but also broke apart on a shorter timescale. This behaviour corresponds to an increased flexibility of the peptide, which can be quantified through the increase in its conformational changes (see Fig. 4B). Figure 6B–D shows the statistics of helical stability, namely the probability of observing a certain number of helical residues (see Fig. 6B), the probabilities for changes in the length of the helical fragment over one picosecond (see Fig. 6C) and the expected lifetime of helical fragments (see Fig. 6D). Figure 6E shows the average potential energies of helical fragments of

certain length. The analysis in Fig. 6B–E included the simulation results obtained at temperatures 302–350 K, so the initial helical structure which persisted through a part of the production simulation at 300 K was not considered. Helical fragments most frequently are formed from 3 residues (see Fig. 6E), although such constructs turn out to be very short-living, having a probability of 0.5 to break apart after 1 ps (see Fig. 6C) and a very short lifetime even up to the 95th percentile (see Fig. 6D). Similarly, helical fragments consisting of 4 residues were also found to be rather unstable. With at least 5 residues in a helical formation, the number of helical residues in the peptide had a reasonable chance to persist (see Fig. 6C), and remarkably the  $A_{11}$  peptide with 5 helical residues does not further stabilise when the helical content is further increased.

Configurations with longer helices occur less frequently compared to those having 3 helical residues since there are less ways to realise a longer helix in the  $A_{11}$  peptide. Peptides with 6 helical residues turn out to be significantly more stable than fragments with 3 or 4 helical residues (see Fig. 6C, D), which is reflected in the probability to find helical fragments with 6 helical residues (see Fig. 6B). The interaction energies in Fig. 6E show another peculiarity of intermediately sized helices. A peptide with 6 helical residues typically has stronger interactions within the peptide and weaker interaction with the solvent. This fact indicates that an  $A_{11}$  peptide with 6 helical residues is found in a conformation in which the contact with the solvent is avoided,



**Fig. 6** Statistics on the number of helical residues in the  $A_{11}$  peptide. **A** The number of helical residues in the  $A_{11}$  peptide, plotted over temperature and simulation time. The temperature was increased stepwise from 300 K to 350 K every 250 ns, so the plot shows a single simulation where the temperature was increased. **B** Probability distribution of the number of helical residues in the  $A_{11}$  peptide. The probability not to contain any helical residues,  $P_0$ , is indicated. Since at least 3 residues are necessary to form a helical structure, it is not possible to find only 1 or 2 helical residues, i.e.  $P_1 = P_2 = 0$ . **C** Probability of a peptide with a certain number of helical residues to increase, persist or decrease in the helical content within the next 1 ps. **D** Lifetime of a certain number of helical residues in the peptide. The average value is plotted in black, while the 40–95 percentiles are shown with colour. **E** Average interaction energies within the peptide (blue, solid line), between the peptide and the solvent (orange, dotted line) and the sum of these interactions (yellow, dashed line) as a function of the helical content of the peptide

an effect that can be attributed to the hydrophobicity of alanine and a possible proximity of the oppositely charged terminal residues.

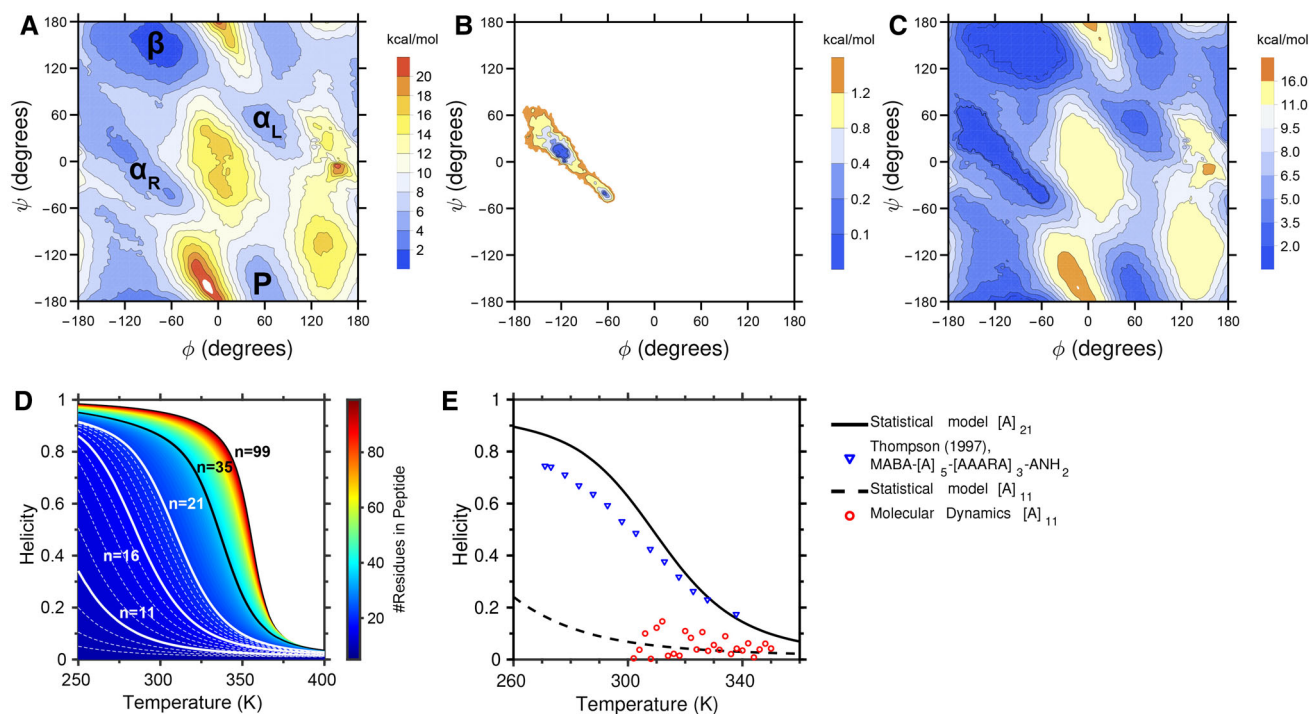
The sum of interaction energies in Fig. 6E shows that there is an energetic advantage for the peptide to form the helical structures. There is, however, no advantage to form longer helices, indicating that the  $A_{11}$  peptide is mostly observed in the random coil-like state already at 300 K.

### 3.2 Statistical mechanics model

Steered MD simulations of the  $A_5$  peptide were performed in explicit water to calculate the potential energy surface of a residue in the polypeptide. The approach used here computes the energy surface of the middle residue in the variable space in which all five residues adopt the same backbone dihedral angles. It is therefore assumed that the secondary structures of interest (energy minima on the surface) show similar backbone dihedral angles for all the residues involved and that these structures are locally stabilised by interactions with residues not further apart than two residues. An  $A_5$  peptide was chosen because two adjacent helical residues in each direction are sufficient for the convergence of the conditional probabilities to be helical (see Fig. 4C) and of helix stability (see Fig. 6C). The resulting potential energy surface of the central residue is shown in Fig. 7A, which describes the poten-

tial energy of a residue inside a solvated alanine peptide computed as a function of the twisting dihedral angles  $\phi$  and  $\psi$ . The potential energy surface was separated into the contributions describing a bound alanine, as illustrated in Figs. S4 and S5. The energy landscape in Fig. 7A corresponds to the probability density of the  $A_{11}$  peptide conformations shown in Fig. 2A, observed for the unconstrained MD simulations with explicit water. The agreement shows that the approach for calculating the potential energy surface is reasonable and permits to establish the potential energy of a residue with respect to its twisting degrees of freedom. The result in Fig. 7A shows that a potential energy of not more than 6 kcal/mol is sufficient for a residue to freely switch between the  $\alpha_R$  and  $\beta$  conformation, while a potential energy of 8 kcal/mol is needed for the  $\beta \rightleftharpoons \alpha_L$  transition and a potential energy of 10 kcal/mol is necessary for the  $\alpha_L \rightleftharpoons P$  transition. It is therefore natural that the frequency of the conformational changes  $\alpha_R \rightleftharpoons \beta$  in Fig. 4B is the largest, followed by the  $\beta \rightleftharpoons \alpha_L$  transitions and finally by the  $\alpha_L \rightleftharpoons P$  transitions. Other transitions are suppressed due to energy barriers of 12 kcal/mol and above. The potential energy barriers here are higher by a factor of 2–3 compared to previous results reported for implicitly solvated  $A_2$  [46]. In the present investigation, the larger change in potential energy is due to the simultaneous steered movement of multiple residues.





**Fig. 7** Potential energy surface of the middle residue inside the solvated  $A_5$  peptide calculated using steered MD and its statistical model interpretation. **A** Average potential energy surface of the middle residue inside the solvated  $A_5$  peptide computed as a function of the two twisting dihedral angles  $\phi$  and  $\psi$  with the potential energies counted relatively to the global energy minimum. **B** Potential energy surface describing the bound state of the residue inside a peptide, where energies are calculated with respect to the local minimum in the  $\alpha_R$  conformation as defined in Fig. 2. The highest energy of the bound state is characterised by the hydrogen bond energy  $E_{HB}$ . **C** Potential energy surface of the unbound state of a residue inside a peptide. The lowest energy of the unbound state matches the hydrogen bond energy  $E_{HB}$ . **D** Helicity of peptides of different lengths. Helicity was calculated within the framework of the statistical model using the potential energy surfaces shown in **B** and **C**, which represent the bound and the unbound states, respectively. The numbers indicate the length of the peptides. **E** Comparison of the temperature dependence of the helicity calculated using the statistical model (lines). The statistical model is compared to the experimentally measured helicity of an alanine-rich peptide consisting of 21 amino acids (blue triangles) [47] and to the average helicity obtained from the MD simulations of the  $A_{11}$  peptide (red circles)

The folding  $\leftrightarrow$  unfolding dynamics of the polypeptide can be described through a two-state statistical mechanics model using the potential energy surface shown in Fig. 7A [26,27]. The potential energy surface needs to be separated into the potential energy surfaces of the bound state corresponding to an  $\alpha_R$  helix conformation and the unbound state corresponding to the random coil state, where residues constantly switch conformations. For such a description, the minimum of the  $\alpha_R$  conformation should first be set to 0 kcal/mol. The energy necessary to break a single hydrogen bond  $E_{HB}$  can be used as a threshold value to separate the bound and unbound conformations; the procedure is illustrated in Fig. S6. The potential energy surface, within a threshold value  $E_{HB}$  counted from the  $\alpha_R$  minimum, therefore describes the bound state (see Fig. 7B), while the potential energy greater than  $E_{HB}$  is representing the unbound state (see Fig. 7C). The statistical mechanics model is sensitive to the choice of  $E_{HB}$ . This stems mainly from the helix initiation factor  $\beta$  defined as [26]:

$$\beta = \exp\left(-\frac{3E_{HB}}{k_B T}\right), \quad (3)$$

where  $k_B$  is the Boltzmann factor and  $T$  is the temperature of the system. Hydrogen bond energies in solvated peptides in simulation and experiment are expected to be in the range of 0.5–1.5 kcal/mol and are significantly weaker than the hydrogen bonds in vacuum [53,54]. A characteristic value of the hydrogen bond energy of  $E_{HB} = 1.65$  kcal/mol was used here, corresponding to the value computed in explicit water using the CHARMM force field [53]. The partition function for a residue of the peptide that has one of the two possible conformations can be constructed as:

$$Z_b = \iint_{K_{E_{HB}}} \exp\left(-\frac{\epsilon_b(\phi, \psi)}{k_B T}\right) d\phi d\psi, \quad (4)$$

$$Z_u = \int_{-180}^{180} \int_{-180}^{180} \exp\left(-\frac{\epsilon_u(\phi, \psi)}{k_B T}\right) d\phi d\psi. \quad (5)$$

The partition function of a residue in the bound state,  $Z_b$ , follows Boltzmann statistics and includes the potential energy of the bound state  $\epsilon_b$  shown in Fig. 7B. The partition function is built from an integral over  $K_{E_{HB}}$  which is the area accessible from the minimum in  $\alpha_R$

limited by the maximal energy  $E_{\text{HB}}$ . In Fig. 7B, the area  $K_{E_{\text{HB}}}$  is coloured. For the partition function of the unbound state,  $Z_u$ , the integration is performed over the whole span of the dihedral angles  $\phi$  and  $\psi$ , considering the energies  $\epsilon_u$  shown in Fig. 7C. The partition functions  $Z_b$  and  $Z_u$  of a residue permit constructing the partition function of the peptide of length  $n$ . Employing the approximation of a singular helical fragment, the partition function of a peptide can be constructed for the purpose of calculating the expectation value of helicity. The theory developed by Yakubovich *et al.* [26] was adopted here to consider 3 residues necessary for helix nucleation and to allow a fully helical peptide. The partition function can then be written as:

$$\mathbb{Z} = Z_u^n + \beta \sum_{k=3}^n (n-k+1) Z_b^k Z_u^{n-k}. \quad (6)$$

Here,  $\beta$ ,  $Z_b$  and  $Z_u$  are defined in Eqs. (3)–(5). The first term,  $Z_u^n$ , accounts for the possibility of a peptide to have all residues in the unbound state. The summation is performed over  $k$ , the number of residues in the helical fragment, which is required to be at least 3 to be recognised as a helix and can maximally be  $n$ , i.e. the number of amino acids in the peptide. The term  $(n-k+1)$  counts the number of ways to fit a helix of length  $k$  inside a peptide of length  $n$ . Using the partition function in Eq. (6), it is possible to derive the expectation value for the fraction of helical residues  $f_H$  in a peptide as:

$$f_H = \frac{1}{\mathbb{Z}} \beta \sum_{k=3}^n \frac{k}{n} (n-k+1) Z_b^k Z_u^{n-k}. \quad (7)$$

The fraction  $k/n$  defines the helical fraction of  $k$  residues in a peptide of length  $n$ . Other terms are the same as defined in Eq. (6). The Zimm–Bragg [55] parameters can be inferred from this model with the helix-propensity  $s$  given as  $s = Z_b/Z_u$  and the helix nucleation parameter  $\sigma = \beta s^t$  where  $t$  depends on the number of residues necessary to initiate a helix [27]. Note that thermodynamical characteristics are calculated with a different approximation for the partition function [26].

Equation (7) is used to compute the expected helicity for peptides of different lengths (see Fig. 7D). The helicity drops with increasing temperature due to the thermal fluctuations in the system. For peptides containing 8–21, 35 and 99 amino acids, the temperature dependence of the helicity was plotted explicitly, while other peptide lengths are indicated by a colour gradient. For small peptides, the helix-initiation factor  $\beta$  dominates the resulting helicity and for longer peptides the difference in entropy between  $Z_b$  and  $Z_u$  becomes the most relevant contribution which causes a steep slope. Longer peptides are expected to have a higher helical fraction which drops at latest around 350 K. Helicity-dependencies for longer peptides become closely packed; the addition of another residue causes

only a minor change in the behaviour of the whole peptide.

The helicity calculated within the two-state statistical model for an  $A_{21}$  peptide was compared with experimental measurements in Fig. 7E, showing a reasonable agreement. Between 270 and 300 K, the statistical model overestimates helicity compared to experiment [47] probably due to the higher helix propensity of alanine (A) compared to arginine (R) [56]. The latter was present in the alanine-rich peptide used in the experiment. It is also possible that the present model needs to be extended such that it also includes the  $\beta$  conformations of the amino acids in a similar way to the  $\alpha_R$  conformation studied here. These additional states in the partition function would not contribute to the helicity and would reduce the helical fraction overall. A study on four slightly different 24 residues long peptides containing 16 alanines found helical fractions of  $f_H = \{0.61, 0.61, 0.54, 0.53\}$  at 277 K [57], which is close to the prediction of the statistical model for an  $A_{16}$  peptide of  $f_H = 0.6$  but significantly lower than the prediction for the  $A_{24}$  peptide of  $f_H = 0.9$ .

Replica Exchange MD performed on an  $A_{21}$  peptide resulted in an approximately linear dependence of helicity on temperature and  $f_H = 0.5$  at 350 K, but the deviations can probably be attributed to the parameterisation of the  $\phi$  and  $\psi$  angles [58]. Comparison of the helicity values for the  $A_{11}$  peptide with the helicity data gathered from MD shows that the results of MD simulations yield helicity values that fluctuate around the value obtained from the statistical model. The comparison illustrates that in order to gather consistent results about the helicity for the  $A_{11}$  peptide, simulations on the order of at least few microseconds are necessary for each temperature of interest. The necessary simulation time for sufficient sampling is expected to increase further for longer peptides due to the larger conformational space.

The potential energy surface of a residue utilised in the presented two-state statistical model employing steered MD of the  $A_5$  peptide only includes interactions with the solvent and its two neighbouring residues in either N- and C-terminal directions. The correlation of neighbouring residues in the  $\alpha_R$  conformation (see Fig. 4) influences the next two adjacent residues making this approximation reasonable. For polypeptides composed of different amino acids, such an approach could, however, yield insufficient results. Interactions of longer side-chains would make it necessary to simulate longer peptides in order to calculate the correct potential energy surfaces.

Comparison of helicity calculated with the use of the statistical model with experimental measurements or simulations of long polypeptides demands further caution. For sufficiently long polypeptides, tertiary structures would emerge, which are not included in the two-state model. Longer peptides will also include several helical fragments, which would alter the statistics and the results. Multiple helical states are not considered in the present study because of the following reasons. Firstly, any additional helical fragment, which

can be considered independent from the other one, should be at least 3 residues apart. The first peptide in which two helical fragments might occur is therefore the A<sub>9</sub> peptide, having the secondary structure 3×Helix - 3×Coil - 3×Helix. This represents only a single state with two helical fragments compared to  $\sum_{k=3}^9 (9-k+1) = 28$  states with a single helical motif, following Eq. (6). Secondly, the boundary amino acids of helical residues are less flexible, which suppresses the weight of multiple helical fragments due to the loss in entropy [27]. Thirdly, the helix-initiation factor needs to be accounted for at least twice when a peptide has more than one helical fragment, which also reduces the statistical weight of such helix-composite structures. To conclude, the presented theory in single-helix approximation should be sufficient for peptide lengths in which tertiary structures are not very dominant.

## 4 Conclusion

The dynamics of a solvated alanine polypeptide was studied with the use of atomistic MD simulations and a statistical approach. It was shown that the backbone dihedral angles of the residues in the polypeptide chain cluster in conformations, which are commonly associated with certain secondary structure motifs in proteins. The dynamics of the backbone dihedral angles was characterised by conformational transitions where the  $\alpha_R \rightleftharpoons \beta$  transitions turned out to be the dominating ones, while the  $\beta \rightleftharpoons \alpha_L$  transitions and  $\alpha_L \rightleftharpoons P$  transitions occurred less frequently. The results of the simulations revealed that prolongation of helical fragments in the direction of the N-terminal or C-terminal differs with respect to the statistical occupation of different folding pathways. Furthermore, helix-propagation in the C-terminal direction was described as rigid helix propagation which might extrapolate to more complex peptides as well and thus provide an explanation why proteinogenesis from the N-terminal to the C-terminal could be advantageous [59].

Once folded, helical fragments showed a short lifetime on the order of ca. 1-20 picoseconds. The investigation revealed that the helix stability increased significantly once the helical fragment becomes extended from 3 to 5 residues. Given the assumption that during the folding process secondary structures form first which then form tertiary structures, 5 or less residues are expected to be insufficiently stable and cannot support the formation of tertiary structures. This indicates a possible identification of those parts of a protein which nucleate the folding process: only helical fragments longer than 5 residues could be considered as nucleation candidates for the formation of tertiary structures.

The four conformations that encompass 90% of the probability density in the  $(\phi, \psi)$ -plane coincide with local minima in the potential energies computed here. These minima are separated by energy barriers of less than 4 kcal/mol. The potential energy surfaces for alanine residues were used to construct a partition function that permitted to calculate the fractional helicity

of the peptide. The predictions of the statistical model revealed that most residues in short peptides, with up to 15 amino acids in length, are expected to be in a flexible coil-like conformation at a temperature of 300 K and that alanine polypeptides of an arbitrary length are not able to hold a stable  $\alpha$ -helix conformation at temperatures about ca. 350 K. Persistent helices beyond that temperature are possible, however, these helices are expected to be stabilised by the formation of tertiary structures.

**Acknowledgements** The authors would like to thank the Volkswagen Foundation (Lichtenberg professorship awarded to I.A.S.), the Deutsche Forschungsgemeinschaft (GRK1885 Molecular Basis of Sensory Biology and SFB 1372 Magnetoreception and Navigation in Vertebrates, no. 395940726), the Ministry for Science and Culture of Lower Saxony (Simulations meet experiments on the nanoscale: opening up the quantum world to artificial intelligence (SMART)). Computational resources for the simulations were provided by the North-German Supercomputing Alliance (HLRN) and by the CARL Cluster at the Carl-von-Ossietzky University, Oldenburg, supported by the DFG and the Ministry for Science and Culture of Lower Saxony.

## Author contributions

JH, IAS, AF did conceptualisation; JH done data curation, validation, writing—original draft and investigation; JH and IAS performed formal analysis and software; IAS provided funding acquisition, project administration and resources; JH, AF, LG, IAS were involved in methodology and writing—review and editing; IAS, LG, and AF supervised the study; JH and LG did visualisation.

**Funding** Open Access funding enabled and organized by Projekt DEAL.

**Data Availability Statement** The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request. This manuscript has data included as electronic supplementary material.

## Declarations

**Conflict of interest** The authors have no conflict of interests to declare that are relevant to the content of this article.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If

material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

1. D. Vasudevan, S. Sreekumari, K. Vaidyanathan, *Textbook of Biochemistry for Medical Students*, 7th edn., chap. 4 (JPBMP, 2013), pp. 33–46
2. D.J. Rigden, *From Protein Structure to Function with Bioinformatics* (Springer, Netherlands, 2017)
3. K. Wolff, Protein structure prediction and folding dynamics. Ph.D. thesis, Technische Universität, Darmstadt (2010)
4. K.A. Dill, S.B. Ozkan, M.S. Shell, T.R. Weikl, The protein folding problem. *Annu. Rev. Biophys.* **37**, 289–316 (2008)
5. A.R. Fersht, V. Daggett, Protein folding and unfolding at atomic resolution. *Cell* **108**, 573–582 (2002)
6. M. Levitt, M. Gerstein, E. Huang, S. Subbiah, J. Tsai, Protein folding: the endgame. *Annu. Rev. Biochem.* **66**, 549–579 (1997)
7. J.C. Phillips, D.J. Hardy, J.D.C. Maia, J.E. Stone, J.V. Ribeiro, R.C. Bernardi, R. Buch, G. Fiorin, J. Hénin, W. Jiang, R. McGreevy, M.C.R. Melo, B.K. Radak, R.D. Skeel, A. Singharoy, Y. Wang, B. Roux, A. Aksimentiev, Z. Luthey-Schulten, L.V. Kalé, K. Schulten, C. Chipot, E. Tajkhorshid, Scalable molecular dynamics on CPU and GPU architectures with NAMD. *Chem. Phys.* **153**, 044130 (2020)
8. S.M. Kimø, I. Friis, I.A. Solov'yov, Atomistic insights into cryptochrome interprotein interactions. *Biophys. J.* **115**, 616–628 (2018)
9. A.B. Salo, P. Husen, I.A. Solov'yov, Charge transfer at the Q<sub>o</sub>-site of the cytochrome bc<sub>1</sub> complex leads to superoxide production. *J. Phys. Chem. B* **121**, 1771–1782 (2016)
10. S.M. Frahs, J.C. Reeck, K.M. Yocham, A. Frederiksen, K. Fujimoto, C.M. Scott, R.S. Beard, R.J. Brown, T.J. Lujan, I.A. Solov'yov, D. Estrada, J.T. Oxford, Pre-chondrogenic ATDC5 cell attachment and differentiation on graphene foam; modulation by surface functionalization with fibronectin. *ACS Appl. Mater. Interfaces* **11**, 41906–41924 (2019)
11. M.C.R. Melo, R.C. Bernardi, C. de la Fuente-Nunez, Z. Luthey-Schulten, Generalized correlation-based dynamical network analysis: a new high-performance approach for identifying allosteric communications in molecular dynamics trajectories. *Chem. Phys.* **153**, 134104 (2020)
12. J.A. Hadden, J.R. Perilla, C.J. Schlicksup, B. Venkatakrishnan, A. Zlotnick, K. Schulten, All-atom molecular dynamics of the HBV capsid reveals insights into biological function and cryo-EM resolution limits. *eLife* **7**, e32478 (2018)
13. K. Lindorff-Larsen, S. Piana, R.O. Dror, D.E. Shaw, How fast-folding proteins fold. *Science* **334**, 517–520 (2011)
14. D.E. Shaw, P. Maragakis, K. Lindorff-Larsen, S. Piana, R.O. Dror, M.P. Eastwood, J.A. Bank, J.M. Jumper, J.K. Salmon, Y. Shan, W. Wriggers, Atomic-level characterization of the structural dynamics of proteins. *Science* **330**, 341–346 (2010)
15. P.L. Freddolino, C.B. Harrison, Y. Liu, K. Schulten, Challenges in protein-folding simulations. *Nat. Phys.* **6**, 751–758 (2010)
16. V.A. Voelz, G.R. Bowman, K. Beauchamp, V.S. Pande, Molecular simulation of ab initio protein folding for a millisecond folder NTL9(1–39). *J. Am. Chem. Soc.* **132**, 1526–1528 (2010)
17. D.E. Shaw, R.O. Dror, J.K. Salmon, J.P. Grossman, K.M. Mackenzie, J.A. Bank, C. Young, M.M. Denneroff, B. Batson, K.J. Bowers, E. Chow, M.P. Eastwood, D.J. Ierardi, J.L. Klepeis, J.S. Kuskin, R.H. Larson, K. Lindorff-Larsen, P. Maragakis, M.A. Moraes, S. Piana, Y. Shan, B. Towles, in *Proceedings of the Conference on High Performance Computing Networking, Storage and Analysis, SC '09* (ACM, New York, 2009), pp. 1–11
18. J.D. Prajapati, C. Mele, M.A. Aksoyoglu, M. Winterhalter, U. Kleinekathöfer, Computational modeling of ion transport in bulk and through a nanopore using the drude polarizable force field. *J. Chem. Inf. Model.* **60**, 3188–3203 (2020)
19. D.W. Kneller, O. Gerlits, L.L. Daemen, A. Pavlova, J.C. Gumbart, Y. Cheng, A. Kovalevsky, Joint neutron/molecular dynamics vibrational spectroscopy reveals softening of HIV-1 protease upon binding of a tight inhibitor. *Phys. Chem. Chem. Phys.* **24**, 3586–3597 (2022)
20. A. Volkhardt, H. Grubmüller, Estimating ruggedness of free-energy landscapes of small globular proteins from principal component analysis of molecular dynamics trajectories. *Phys. Rev. E* **105**, 044404 (2022)
21. S.L. Schaefer, H. Jung, G. Hummer, Binding of SARS-CoV-2 fusion peptide to host endosome and plasma membrane. *J. Phys. Chem. B* **125**, 7732–7741 (2021)
22. A.V. Yakubovich, *Theory of Phase Transitions in Polypeptides and Proteins* (Springer, Berlin, 2011)
23. G.G. Maisuradze, A. Liwo, P. Senet, H.A. Scheraga, Local vs global motions in protein folding. *J. Chem. Theory Comput.* **9**, 2907–2921 (2013)
24. A.V. Yakubovich, I.A. Solov'yov, A.V. Solov'yov, W. Greiner, Conformational changes in glycine tri- and hexapeptide. *Eur. Phys. J. D At. Mol. Opt. Plasma Phys.* **39**(1), 23–34 (2006)
25. A.V. Yakubovich, I.A. Solov'yov, A.V. Solov'yov, W. Greiner, Phase transition in polypeptides: a step towards the understanding of protein folding. *Eur. Phys. J. D.* **40**, 363–367 (2006)
26. A.V. Yakubovich, I.A. Solov'yov, A.V. Solov'yov, W. Greiner, Ab initio theory of helix↔coil phase transition. *Eur. Phys. J. D.* **46**, 215–225 (2007)
27. I.A. Solov'yov, A.V. Yakubovich, A.V. Solov'yov, W. Greiner, α-helix↔random coil phase transition: analysis of ab initio theory predictions. *Eur. Phys. J. D.* **46**, 227–240 (2007)
28. A.V. Yakubovich, I.A. Solov'yov, A.V. Solov'yov, W. Greiner, Phase transitions in polypeptides: analysis of energy fluctuations. *Eur. Phys. J. D.* **51**, 25–32 (2008)
29. S.A. Hollingsworth, P.A. Karplus, A fresh look at the ramachandran plot and the occurrence of standard

- structures in proteins. *Biomol. Concepts* **1**, 271–283 (2010)
30. O. Carugo, K. Djinović-Carugo, Half a century of Ramachandran plots. *Acta Crystallogr. D* **69**, 1333–1341 (2013)
  31. J. Song, H. Tan, M. Wang, G.I. Webb, T. Akutsu, TANGLE: Two-level support vector regression approach for protein backbone torsion angle prediction from primary sequences. *PLoS ONE* **7**, e30361 (2012)
  32. J.C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R.D. Skeel, L. Kalé, K. Schulten, Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **26**, 1781–1802 (2005)
  33. R.B. Best, X. Zhu, J. Shim, P.E.M. Lopes, J. Mittal, M. Feig, A.D. MacKerell, Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone  $\phi$ ,  $\psi$  and side-chain  $\chi_1$  and  $\chi_2$  dihedral angles. *J. Chem. Theory Comput.* **8**, 3257–3273 (2012)
  34. A.D. MacKerell, D. Bashford, M. Bellott, R.L. Dunbrack, J.D. Evanseck, M.J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F.T.K. Lau, C. Mattos, S. Michnick, T. Ngo, D.T. Nguyen, B. Prodhom, W.E. Reiher, B. Roux, M. Schlenkrich, J.C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiórkiewicz-Kuczera, D. Yin, M. Karplus, All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **102**, 3586–3616 (1998)
  35. A.D. Mackerell, M. Feig, C.L. Brooks, Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J. Comput. Chem.* **25**, 1400–1415 (2004)
  36. W. Humphrey, A. Dalke, K. Schulten, VMD: visual molecular dynamics. *J. Mol. Graph.* **14**, 33–38 (1996)
  37. D. Frishman, P. Argos, Knowledge-based protein secondary structure assignment. *Proteins Struct. Funct. Genet.* **23**, 566–579 (1995)
  38. F. Schuhmann, V. Korol, I.A. Solov'yov, Introducing pep McConst—a user-friendly peptide modeler for biophysical applications. *J. Comput. Chem.* **42**, 572–580 (2021)
  39. V. Korol, P. Husen, E. Sjulstok, C. Nielsen, I. Friis, A. Frederiksen, A.B. Salo, I.A. Solov'yov, Introducing VIKING: a novel online platform for multiscale modeling. *ACS Omega* **5**, 1254–1260 (2019)
  40. G.J. Martyna, D.J. Tobias, M.L. Klein, Constant pressure molecular dynamics algorithms. *Chem. Phys.* **101**, 4177–4189 (1994)
  41. S.E. Feller, Y. Zhang, R.W. Pastor, B.R. Brooks, Constant pressure molecular dynamics simulation: the Langevin piston method. *Chem. Phys.* **103**, 4613–4621 (1995)
  42. W. Weber, P.H. Hünenberger, J.A. McCammon, Molecular dynamics simulations of a polyalanine octapeptide under Ewald boundary conditions: influence of artificial periodicity on peptide conformation. *J. Phys. Chem. B* **104**, 3668–3675 (2000)
  43. D. Ting, G. Wang, M. Shapovalov, R. Mitra, M.I. Jordan, R.L. Dunbrack, Neighbor-dependent Ramachandran probability distributions of amino acids developed from a hierarchical Dirichlet process model. *PLoS Comput. Biol.* **6**, e1000763 (2010)
  44. W.N. Du, K.A. Marino, P.G. Bolhuis, Multiple state transition interface sampling of alanine dipeptide in explicit solvent. *J. Chem. Phys.* **135**, 145102 (2011)
  45. J.D. Chodera, W.C. Swope, J.W. Pitera, K.A. Dill, Long-time protein folding dynamics from short-time molecular dynamics simulations. *Multiscale Model. Simul.* **5**, 1214–1226 (2006)
  46. D.S. Chekmarev, T. Ishida, R.M. Levy, Long-time conformational transitions of alanine dipeptide in aqueous solution: continuous and discrete-state kinetic models. *J. Phys. Chem. B* **108**, 19487–19495 (2004)
  47. P.A. Thompson, W.A. Eaton, J. Hofrichter, Laser temperature jump study of the helix=coil kinetics of an alanine peptide interpreted with a kinetic zipper model. *Biochemistry* **36**, 9200–9210 (1997)
  48. B. Bösterling, J. Engel, Kinetic studies on the helix-coil transition of fluorescent labeled poly(-l-lysine) by the temperature-jump technique. *Biophys. Chem.* **9**, 201–209 (1979)
  49. B. Gruenewald, C. Nicola, A. Lustig, G. Schwarz, H. Klump, Kinetics of the helix-coil transition of a polypeptide with non-ionic side groups, derived from ultrasonic relaxation measurements. *Biophys. Chem.* **9**, 137–147 (1979)
  50. S. Williams, T.P. Causgrove, R. Gilmanishin, K.S. Fang, R.H. Callender, W.H. Woodruff, R.B. Dyer, Fast events in protein folding: helix melting and formation in a small peptide. *Biochemistry* **35**, 691–697 (1996)
  51. P. Ferrara, J. Apostolakis, A. Cafisch, Thermodynamics and kinetics of folding of two model peptides investigated by molecular dynamics simulations. *J. Phys. Chem. B* **104**, 5000–5010 (2000)
  52. V. Daggett, M. Levitt, Molecular dynamics simulations of helix denaturation. *J. Mol. Biol.* **223**, 1121–1138 (1992)
  53. N. Arora, B. Jayaram, Strength of hydrogen bonds in  $\alpha$  helices. *J. Comput. Chem.* **18**, 1245–1252 (1997)
  54. S.Y. Sheu, D.Y. Yang, H.L. Selzle, E.W. Schlag, Energetics of hydrogen bonds in peptides. *PNAS* **100**, 12683–12687 (2003)
  55. B.H. Zimm, J.K. Bragg, Theory of the phase transition between helix and random coil in polypeptide chains. *Chem. Phys.* **31**, 526–535 (1959)
  56. K. Manikandan, S. Ramakumar, The occurrence of C–H...O hydrogen bonds in  $\alpha$ -helices and helix termini in globular proteins. *Proteins Struct. Funct. Genet.* **56**, 768–781 (2004)
  57. C.A. Olson, Z. Shi, N.R. Kallenbach, Polar interactions with aromatic side chains in  $\alpha$ -helical peptides: Ch–O H-bonding and cation- $\pi$  interactions. *J. Am. Chem. Soc.* **123**, 6451–6452 (2001)
  58. A.E. García, K.Y. Sanbonmatsu,  $\alpha$ -helical stabilization by side chain shielding of backbone hydrogen bonds. *PNAS* **99**, 2782–2787 (2002)
  59. J.J. Ellis, F.P. Huard, C.M. Deane, S. Srivastava, G.R. Wood, Directionality in protein fold prediction. *BMC Bioinform.* **11**, 1–16 (2010)