MD Simulations of N-terminal Peptides from Lactate Dehydrogenase

David van der Spoel, Hans J. Vogel and Herman J.C. Berendsen

Molecular dynamics (MD) simulations of N-terminal peptides from Lactate Dehydrogenase (LDH) with increasing length and individual secondary structure elements were used to study their stability in relation to folding. Ten simulations of 1-2 ns of different peptides in water starting from the coordinates of the crystal structure were performed. The stability of the peptides was compared qualitatively by analyzing the root mean square deviation (RMSD) from the crystal structure, radius of gyration, secondary and tertiary structure and solvent accessible surface area. In agreement with earlier MD studies, relatively short (<15 amino acids) peptides containing individual secondary structure elements were generally found to be unstable; the hydrophobic $\alpha_1$-helix of the nucleotide binding fold displayed a significantly higher stability, however. Our simulations further showed that the first $\beta\alpha\beta$ supersecondary unit of the characteristic dinucleotide binding fold (Rossmann fold) of LDH is somewhat more stable than other units of similar length and that the $\alpha_2$-helix, which unfolds by itself, is stabilized by binding to this unit. This suggests that the first $\beta\alpha\beta$ unit could function as an N-terminal folding nucleus, upon which the remainder of the polypeptide chain can be assembled. Indeed, simulations with longer units ($\beta\alpha\beta\alpha\beta$ and $\beta\alpha\beta\alpha\beta\beta\alpha$) showed that all structural elements of these units are rather stable. The outcome of our studies is in line with suggestions that folding of the N-terminal portion of LDH in vivo can be a cotranslational process which takes place during the ribosomal peptide synthesis.
6.1 Introduction

The folding of a protein into its correct three-dimensional structure is a topic of continuing interest [33]. Following the original demonstration by Anfinsen [15], that a denatured protein can regain its original structure when the denaturing agent is removed, it is clear that the information required for the folding of the protein is retained in its amino acid sequence. Since this discovery it has been customary to study protein folding by following the renaturation of the denatured polypeptide, which comprises the entire amino acid sequence. These studies have led to the widely held view that so-called "molten globules" exist as obligatory intermediates along the in-vitro folding pathway of many proteins [22-24, 239]. Such intermediates have been characterized in detail primarily by NMR hydrogen exchange measurements and have been found for various segments in the amino acid sequence [240-242]. This view is no longer generally accepted however, as recent experiments have suggested that the partially folded forms of some proteins could be trapped unproductive forms rather than obligatory folding intermediates [25, 26].

Other studies aimed at solving the protein folding problem have relied on the assumption that the overall folding of a protein is dictated by the intrinsic folding properties of short stretches of the amino acid sequence. Consequently, short linear peptides resembling parts of the amino acid sequence of various proteins have been produced synthetically, and have been studied by high resolution NMR and CD spectroscopy for their folding properties [66, 212, 243]. While this work has clearly established the transient formation of elements of secondary structure in such peptides, the role of these secondary structure elements in directing the course of folding remains to be determined [212, 243].

The question remains whether protein folding in vivo really occurs from a completely synthesized and unstructured random coil peptide. Several authors have argued that it would be possible in principle to initiate folding in the N-terminal part of a protein because the polypeptide chain emerges in an ordered manner from the ribosome during protein translation [1, 32, 244, 245]. This process could give rise to an N-terminal folding nucleus, which can subsequently serve as a site of assembly for the remaining portions of the polypeptide chain. This phenomenon is referred to as cotranslational rather than posttranslational folding [244], and a number of molecular biology studies with expressed truncated proteins [246-249] have indicated that the folding of certain classes of proteins may well proceed in this fashion. If such an N-terminal folding nucleus were to be present within a group of proteins with identical folds, it should have a recognizable feature near the N-terminal end of the protein. One group of proteins that has a well-characterized overall fold, which can be readily recognized on the basis of their amino acid homology are the nucleotide binding proteins [250]. In particular the well characterized Rossmann fold, also known as the "classical dinucleotide binding chain" fold, is found in all NAD\(^+\), NADP\(^+\) and FAD binding proteins [250, 251]. In addition, a related, but structurally dissimilar "mononucleotide binding chain" fold is found in many proteins that bind nucleotides such as ATP and GTP [250, 252, 253]. That this class of proteins may be suitable for studying the formation of an N-terminal folding nucleus was suggested in a paper by Wierenga et al. [254]. They commented that the nucleotide binding \(\beta\alpha\beta\) unit
could be a possible nucleation site for protein folding: “... In the known structures these \( \beta\alpha\beta \)-structural entities do always occur near the N terminus ... This suggests that these \( \beta\alpha\beta \)-folds might function as a nucleation center for the folding of the complete domain ...” The same region was also indicated as a possible folding unit by an analysis of the ribosomal translation rate [255]. Finally, it is known that N-terminal proteolytic fragments of Lactate Dehydrogenase (LDH) retain the capacity to bind dinucleotides, indicating that they are properly folded [256]. The \( \beta\alpha\beta \) unit, which is an integral part of the Rossmann fold, is a common supersecondary structural element in proteins [257]. In all, at least 40 proteins are presently known or predicted to have this fold, and considering the “age” of the fold, which was estimated to be over 3 billion years [257] there will probably be more than these.

In this work we have used Molecular Dynamics (MD) calculations to obtain further information about the stability of peptides derived from the N-terminal portion of LDH, a protein with the dinucleotide binding fold. Over the last few years MD calculations, when used judiciously, have become a reliable means for studying the stability and flexibility of small proteins in water [43, 258]. With continuing improvements in force fields and computing capacity, it has become possible to study processes such as protein denaturation [37, 38, 41, 42, 258]. In addition, simulations of peptides in aqueous solution can now be performed routinely and such studies have been done to assess the stability of turns and helices in short linear peptides [36, 40, 202, 206, 213, 238, 259]. It is not possible to assess the thermodynamic stability of a peptide or protein in solution directly from MD simulations, due to limited sampling of unfolded as well as partially folded states. However, it is possible to investigate the kinetic stability i.e. the resistance against unfolding and compare this for various peptides or proteins in different environments or at different temperatures. This has recently been done in a number of cases, four regarding \( \alpha \)-helical peptides [44, 67, 225, 260] and in another case a \( \beta \)-hairpin was studied [147]. Finally, it has also been noted that MD simulations can assist in an analysis of protein folding pathways [261].

Tsou [244] has suggested that an N-terminal folding nucleus should have a structure that may not be identical, but should resemble the final structure in the completed protein. Therefore we decided to use the folded parts as determined in the crystal structure of the protein as starting structures for our MD calculations. In order to simulate the effect of the ribosomal protein synthesis, we studied the N-terminal region, and systematically increased the length of the simulated polypeptide chain. For comparison a number of other units derived from the same amino acid sequence were also simulated. While this approach does not assess folding directly, the kinetic stability of the various peptides can be compared in this manner, as was done previously in MD studies of \( \alpha \)-helices and \( \beta \)-turns [36, 40, 44, 67, 147, 202, 213, 225, 238].
6.2 Methods

Ten MD simulations of peptides from LDH in water, varying in length from 6 to 76 amino acid residues, were performed. One of these was a deprotonated peptide, in which the sidechain of a Lysine residue (Lys22) was made neutral, in order to assess the influence of a salt bridge on the stability. In our simulations we ignored the first 20 residues of the protein, because in the complete protein they form a motif which is sticking out into the solvent. This loop is necessary for oligomerization, where the loop stabilizes the tetramer [256]; therefore the loop presumably is less important for the structure of the monomeric folded protein.

Figure 6.1: Overview of the peptides names and sequences used.

Table 6.1: Sequence characteristics (H = α-helix, G = 3_10-helix, T = Hydrogen Bonded Turn, S = Bend, E = Extended strand participating in Beta Sheet, Space = No defined secondary structure).

All but four of the peptides start at residue Asn21. To distinguish the simulations we...
named them \( \beta_1, \alpha_1, \beta_\alpha, \beta_2, \beta_\alpha \beta_1, \alpha_2, \beta_\alpha \beta_2, \beta_\alpha \beta_\alpha \) and \( \beta_\alpha \beta_\alpha \beta_\beta \) respectively, where the names refer to the structural elements simulated. The simulation of the deprotonated peptide was named \( \beta_\alpha \beta^- \). An overview of the simulations is given in Fig. 6.1. Simulations \( \alpha_1, \beta_\alpha, \beta_\alpha \beta_1, \beta_\alpha \beta^- \) and \( \beta_\alpha \beta_2 \) were 2.0 ns, the other simulations were 1.0 ns long. The sequence for the longest \( \beta_\alpha \beta_\alpha \beta_\beta \) peptide along with some characteristics are given in Table 6.1. Plots of the \( \beta_\alpha \beta_1 \) and \( \beta_\alpha \beta_\alpha \beta_\beta \) peptides are given in Fig. 6.2 and Fig. 6.3 respectively.

The coordinates for the peptides were taken from the crystal structure of dogfish apo-lactate dehydrogenase which was determined by Abad-Zapatero [262]. The structure was retrieved from the Brookhaven pdb database, entry 6LDH. They were solvated in a rectangular box of SPC water [164] using a shell of at least 0.8 nm on every side of the peptide. Although the use of counterions would make the total system electrostatically neutral, we feel that this would not improve the results from the simulation without using a proper long-range electrostatic method such as Ewald summation [53, 197] or particle-particle, particle-mesh (PPPM) [236, 237]. Although these methods are well-established there are some serious drawbacks to using them: 1) a crystalline environment is imposed on a system that is meant to be a peptide in infinite dilution, 2) a correct lattice sum method requires the presence of counterions which, when not sampled properly will induce artifacts, 3) the CPU time is much longer due to different scaling, \( O(N^2 \log N) \) for Ewald vs. \( O(N) \) for simple cut-offs, although the required CPU time for the PPPM method is comparable to a simple cut-off [237]. Reaction field methods assume a dielectric continuum outside a specified cut-off range around a given particle [263], which implies that the whole peptide should be within cut-off distance; this is impractical for peptides with a typical length of 2 nm, twice the cut-off distance. A further complication towards using e.g. the PPPM method is, to our knowledge, that there is no implementation of this method on message passing parallel computers which we have used for our simulations.

Some promising methods to deal with long-range electrostatic interactions are currently under development [233, 264]; these will be implemented on our parallel computers in the near future.

All simulations were performed using periodic boundary conditions. The solvated peptides were subsequently energy minimized with the steepest descent method for 100 steps.
Then the system was run with position restraining on the peptide for 50 ps to allow for further relaxation of the solvent molecules. During these restrained MD runs the temperature was controlled using weak coupling to a bath of constant temperature ($T_0 = 300$ K, coupling time $\tau_T = 0.1$ ps) and the pressure was controlled using weak coupling to a bath of constant pressure ($P_0 = 1$ bar, coupling time $\tau_P = 0.5$ ps) [168]. The production runs were done with the same pressure- and temperature-coupling constants as the restrained runs. Protein and solvent were coupled to the temperature bath separately in restrained as well as free MD. The center of mass motion of the entire simulation system was removed every step. The Gromos-87 forcefield [11] was used with modifications as suggested in [159]. The timestep was 2.0 fs, SHAKE [169] was used for all covalent bonds, and the cut-off for nonbonded interactions was set to 1.0 nm. Neighbourlists were used, and updated every 20 fs.

To each peptide we added a $\text{CH}_3\text{C}=\text{O}$ group at the N-terminus and an $\text{NH}_2$ group at the C-terminus; these neutral groups were used in order to avoid disturbing the peptides by introducing charges on their N- and C-terminal ends. For comparison with future NMR and CD experiments, residue Cys35 was mutated to Ser to eliminate the possibility of disulphide bridge formation in the test tube; we think this substitution is unlikely to have a great impact on the structure of the peptides. The number of atoms and the number of water molecules for each peptide simulation are given in Table 6.2. The simulations were carried out using the GROMACS software package [265] on custom-built parallel computers [51, 52] with Intel $i800$ CPU's and on a Silicon Graphics Power Challenge. The total computer time for all simulations was approximately 2000 h.

![Figure 6.3: Plot of the $\beta\alpha\beta\alpha\beta$ unit, was made using Molscript [170]](image-url)
Table 6.2: Overview of the peptides used in the simulations, the number of water molecules and the net charge on each peptide.

<table>
<thead>
<tr>
<th>Peptide</th>
<th>Residues</th>
<th># Residues</th>
<th># Waters</th>
<th># Atoms</th>
<th>Charge</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_1$</td>
<td>Asn21 - Gly29</td>
<td>9</td>
<td>1346</td>
<td>4120</td>
<td>+1</td>
</tr>
<tr>
<td>$\alpha_1$</td>
<td>Gly27 - Asp43</td>
<td>17</td>
<td>1612</td>
<td>4974</td>
<td>0</td>
</tr>
<tr>
<td>$\beta_0$</td>
<td>Asn21 - Asp43</td>
<td>23</td>
<td>1886</td>
<td>5854</td>
<td>+1</td>
</tr>
<tr>
<td>$\alpha_2$</td>
<td>Asp46 - Val51</td>
<td>6</td>
<td>1394</td>
<td>4238</td>
<td>-2</td>
</tr>
<tr>
<td>$\beta_0 \beta_1$</td>
<td>Asn21 - Asp52</td>
<td>32</td>
<td>1929</td>
<td>6057</td>
<td>-2</td>
</tr>
<tr>
<td>$\beta_0 \beta_2$</td>
<td>Asn21 - Asp52</td>
<td>32</td>
<td>1947</td>
<td>6110</td>
<td>-3</td>
</tr>
<tr>
<td>$\alpha_2$</td>
<td>Asp56 - Ser69</td>
<td>14</td>
<td>1627</td>
<td>5020</td>
<td>0</td>
</tr>
<tr>
<td>$\beta_0 \alpha$</td>
<td>Asn21 - Leu72</td>
<td>52</td>
<td>2340</td>
<td>7480</td>
<td>-3</td>
</tr>
<tr>
<td>$\beta_0 \beta_0 \beta_3$</td>
<td>Asn21 - Ala96</td>
<td>76</td>
<td>2510</td>
<td>8202</td>
<td>0</td>
</tr>
</tbody>
</table>

6.3 Results

When analysing long MD simulations of peptides or proteins it is useful to define properties that give a relative estimate of the stability. We have analysed the RMS deviation from the crystal structure, the radius of gyration and the solvent accessible hydrophobic surface area. These values can be plotted as a function of time; in a stable protein these values fluctuate around an equilibrium value, in a changing conformation there will be a drift in one or more of these observables. Since we do not expect all our peptides to be stable in the simulation it is also interesting to look at the dynamical processes taking place during the simulation. To illustrate this we have also looked at secondary structure analysis, distance matrices and salt bridges involving Lys22. These items will be addressed in subsequent sections.

6.3.1 RMS deviation from the crystal structure

The root mean square deviation (RMSD) of the backbone atoms (N, Ca, C) with respect to the crystal structure was calculated by least-squares fitting the backbone atom positions of each peptide to the crystal structure and subsequently calculating the RMSD (eqn. 6.1).

\[
RMSD(t) = \left[ \frac{1}{N_S} \sum_{i \in S} (r_i(t) - r_i(0))^2 \right]^{\frac{1}{2}}
\]

where \(r_i(t)\) is the position of atom \(i\) at time \(t\) and \(S\) is a given subset of all atoms with size \(N_S\). In Fig. 6.4 the RMSD averaged over all residues is plotted as a function of time for all the simulations.

The RMSD of the $\beta$-strands in Sim. $\beta_1$ and $\beta_2$ are very high; these peptides deviate very much from the crystal structure. $\beta_2$ deviates gradually, whereas $\beta_1$ is more fluctuating.
It should be mentioned that an RMS deviation of 0.7 nm for a 9 residue peptide \( (\beta) \) means that it is not very meaningful to fit the structures on top of each other. After an initial rise of the RMSD to 0.3 nm, the isolated \( \alpha_1 \)-helix is stable for 800 ps at an RMSD of 0.2 nm (from 350 until 1150 ps).

After this point in the simulation the RMSD gradually rises to more than 0.4 nm due to unfolding of the C-terminal part of the \( \alpha \)-helix. The \( \alpha_2 \) peptide deviates up to 0.4 nm from the crystal structure within 800 ps and remains at that level for the rest of the 2.0 ns simulation. In the \( \beta \alpha \) simulation the RMSD goes up after 600 ps due to unfolding of the helix starting from the N-terminal side, i.e. from the flexible loop. Around 1.0 ns the \( \beta \alpha \) peptide reaches an RMSD of 0.5 nm, after which point the RMSD decreases to 0.3 nm at the end of the simulation.

In the runs of the \( \beta \alpha \beta_1 \), the \( \beta \alpha \beta \) and the \( \beta \alpha \beta_2 \) peptides the RMS deviation slowly increases to 0.4 nm after 1000 ps, but after 1250 ps it decreases again until it stabilises at 0.3 nm. In all three \( \beta \alpha \beta \) units the RMSD is then virtually constant for over 700 ps until the end of the 2.0 ns simulation. In the two longest peptides the RMSD slowly rises to 0.2 \( (\beta \alpha \beta \beta \beta ) \) resp 0.3 \( (\beta \alpha \beta \alpha) \) nm. In both simulations there is still a small drift at the end of the simulation (1000 ps) indicating that the peptides are not fully equilibrated.

Figure 6.4: Root Mean Square Deviation of all backbone atoms from the X-ray structure. A running average over 25 ps is given to improve clarity.
The RMSD was also averaged over the backbone atoms per residue over the last 1000 ps of each simulation (i.e., over the entire trajectory for Sims. $\beta_1$, $\beta_2$, $\beta\alpha\beta\alpha$ and $\beta\alpha\beta\alpha\beta\beta$) (see Fig. 6.5). Here we can compare how parts of secondary structure behave in different peptides.

In most simulations that contain the $\alpha_1$-helix ($\beta\alpha$, $\beta\alpha\beta_1$, $\beta\alpha\beta^-\alpha$ and, to a lesser extent in $\beta\alpha\beta\alpha$) the C-terminal end of the $\alpha_1$-helix is flexible. The $\alpha_1$-helix is flexible at the ends only whereas the $\alpha_2$-helix is flexible over its total length, especially near Gly60. In all simulations containing residues 70-72 (Leu-Phe-Leu) ($\beta\alpha\beta_2$, $\beta\alpha\beta\alpha$ and $\beta\alpha\beta\alpha\beta\beta$) this section of the polypeptide chain is very flexible because three hydrophobic side chains are solvent exposed, and are trying to move away from the solvent. There is no suitable place to go for these side chains however, so they remain solvated. Another noticeable feature is the Gly-rich loop (residues 27-32) which shows a peak in $\beta_1$, $\beta\alpha$ and $\beta\alpha\beta\beta$ though varying in height.

![Figure 6.5: Root Mean Square Deviation from the X-ray structure per residue, averaged over the last 1000 ps of each simulation.](image)

6.3.2 Radius of gyration

For each peptide the radius of gyration $R_g$ was calculated according to eqn. 6.2

$$R_g = \left( \frac{\sum_i r_i^2 m_i}{\sum_i m_i} \right)^{\frac{1}{2}}$$

where $m_i$ is the mass of atom $i$ and $r_i$ the position of atom $i$ with respect to the center of mass of the peptide. The radius of gyration is a rough measure for the compactness.
of a structure. In all but the longest peptides ($\beta_1\beta_2$, $\beta\alpha\beta\alpha$ and $\beta\alpha\beta\alpha\beta\beta$) we find some change in $R_g$ (Fig. 6.6).

In Sim. $\beta_1$ the decrease at about 200 ps is due to a collapse of the peptide where a $\beta$-turn-like structure is formed, which later unfolds again. $\beta_2$ collapses somewhat and remains stable after 700 ps. In $\alpha_1$ $R_g$ decreases gradually because the loops at both ends of the $\alpha_1$-helix sometimes curl inward. In contrast the large fluctuations in $R_g$ in Sim. $\alpha_2$ are due to unfolding and refolding of parts of the $\alpha_2$-helix. In the first 300 ps of the $\beta\alpha$ simulation we see $R_g$ decreasing with 0.1 nm because the C-terminal loop at the end of the helix collapses inward, making a short lived hydrogen bond between the sidechain NH$_2$ of Asn21 and the carbonyl oxygen of residue Asp43. Movements in the Gly-rich loop (res 27-32) also contribute to the change in $R_g$.

After 1.0 ns $R_g$ decreases slowly to 0.75 nm. In Sim. $\beta\alpha\beta$ $R_g$ decreases somewhat after 800 ps due to movements in the second loop region; in $\beta\alpha\beta_1$ in contrast $R_g$ rises slightly after 800 ps although the final value (0.9 nm) is close to the initial one (0.87 nm).

### 6.3.3 Secondary structure analysis

A secondary structure analysis was performed with the DSSP program which determines the existence of hydrogen bonds as criteria for the presence of secondary structure [65]. In Fig. 6.7 we have plotted the secondary structure elements of each residue in the different peptides as a function of time. Colors were used to distinguish between secondary structure types, the most important colors being red for $\beta$-sheet and blue for $\alpha$-helix.
Figure 6.7: Secondary structure as a function of time for all simulations.
From Fig. 6.7 it can be seen that the isolated $\beta_1$ (Res. 21-27) and loop (Res. 28-29) form a hydrogen bonded turn centered around Gly27 sometimes, and a bend most of the time. In $\beta_2$, a hydrogen bonded turn appears to be the most stable form in the simulation, during formation of the turn structure a transient $3_{10}$ helix is found, at 320 ps and around 375 ps. Neither of the two $\beta$-strands remains even close to their starting structure. The $\alpha_1$-helix is rather stable, compared to the crystal structure only two residues on each side come off initially; three residues in the middle of the $\alpha_1$-helix also lose their $\alpha$-helicity temporarily, but they rapidly regain their original structure. After 1500 ps however, 4 residues on the C-terminal end lose their helicity although they remain hydrogen bonded in a turn. As can be seen in the $\alpha_1$ simulation (around 125 ps) the helix secondary structure type (blue) can be rapidly changed into the turn structure type (yellow) and back, indicating that these secondary structure types can interconvert rapidly within the definitions of the DSSP program. In the $\beta\alpha$ unit the hydrophobic $\beta$-strand and the Gly-rich loop apparently are “pulling” on the $\alpha_1$-helix thereby unfolding it from the N-terminal side; at 1000 ps $\alpha_1$ is completely unfolded in this peptide. The $\alpha_1$-helix rapidly reforms however, starting from the N-terminal end at Met33 which directly follows a Gly residue; an 9 residue $\alpha$-helix is formed in 200 ps. Simulations $\beta\alpha\beta_1$ and $\beta\alpha\beta^-$ can be compared in detail using this analysis too; both units are altogether stable, the number of hydrogen bonds in the $\beta$-sheet varies but is never less than two. In the $\alpha$-helical region of these peptides there is a difference between the two units, the $\alpha_1$-helix loses three residues on the C-terminal end in the $\beta\alpha\beta_1$ simulation whereas it loses seven $\alpha$-helical residues in the C-terminal end in the $\beta\alpha\beta^-$ simulation. The loss of helicity is due to rearrangements involving the hydrophobic residues in the second loop region and at the end of the $\alpha_1$-helix. After 2.0 ns 8 helical residues remain in simulation $\beta\alpha\beta_1$ and 4 in $\beta\alpha\beta^-$, and a stable $\beta$-sheet in both. Considerable structural changes occur in the $\alpha_2$ peptide simulation, with $\alpha$-helical residues transforming into $\beta$-helix or a hydrogen bonded turn, initially at the C-terminal end only, but after 600 ps the $\alpha_2$-helix unfolds from the N-terminal side and for several periods of up to 100 ps (1150 to 1250 ps), all $\alpha$-helix is gone. A central Gly residue (Gly60) probably is the reason for breaking. During the whole simulation we see $\alpha$-helical parts in either the N-terminus or the C-terminus of the peptide but rarely simultaneously (except around 1700 ps). In the $\beta\alpha\beta_2$ peptide we see that the $\alpha_2$-helix unfolds from the C-terminus. All $\alpha$-helical residues after Gly60 are gone after 850 ps. In contrast the $\beta$-sheet is very stable, it restricts the conformational space of the residues in between. It is also interesting to note that considerable non-native secondary structure is formed after the C-terminal part of the $\alpha_1$-helix has unrolled. When the $\alpha_2$-helix is connected to the $\beta\alpha\beta$ peptide to form the $\beta\alpha\beta\alpha$ peptide the $\alpha_2$-helix is more stable than in isolated form in solution (Fig. 6.7). In the $\beta\alpha\beta\alpha$ peptide the most flexible part (apart from the three C-terminal residues) is the Gly-rich loop (Res. 27-32), which unfolds the $\alpha_1$-helix from the N-terminal side, similar to what was observed in the $\beta\alpha$ simulation. As in the $\alpha_2$ simulation the $\alpha_2$-helix breaks at Gly60, but at either side of Gly60 the $\alpha_2$-helix is rather stable. The longest $\beta\alpha\beta\alpha\beta$ peptide is stable during the whole simulation, only the $\alpha_2$-helix (which breaks at Gly60 again) and the three solvent-exposed hydrophobic
residues (70-72) change their secondary structure, although almost always 10 residues remain in α-helical conformation in the α2-helix. In both the βαβαα and βαβαββ the β-sheet is stable; there are almost always two or more hydrogen bonds connecting the individual β-strands; nevertheless in the βαβαββ peptide the first and second β-strand are even more stable than in the βαβαα, due to the addition of the two extra β-strands on either side of the original β-sheet (β-strands are in order β1β2β3). Although this increased fixation of β-strands also gives rise to a more stable α1-helix the secondary structure pattern of the α2-helix in the βαβαββ peptide is similar to that in the βαβα peptide.

6.3.4 Distance matrices

Although the secondary structure analysis is very important for understanding the simulation, it does not tell all about the tertiary structure of the protein. We have therefore calculated distance matrices, in which a matrix of residue-residue distances is computed as a function of time. We defined the distance between two residues $A_i$ and $A_j$ as the smallest distance between any pair of atoms ($i \in A_i$, $j \in A_j$). The distance matrix is thus symmetric by definition. We have visualised this information using greyscales to represent distances. In Fig. 6.8 the distance matrices for five different time frames in simulations βαβ1, βαβ− and βαβ2 are plotted. The distances have been discretized in thirteen levels from 0 (black) to 1.2 nm (white) as is depicted in the figure legend. All distances greater than 1.2 nm are plotted in white. Although the figures appear to be very similar, a closer look reveals very detailed information in these figures. The α1-helix can be seen in the center of the figures, where a subdiagonal (and a superdiagonal) line close to the diagonal shows that residues at position $(n+4)$ in the sequence are close to residues at position $(n)$, where $(n)$ runs from residue 29 to 38. At 800 ps in both simulations part of the α1-helix has unfolded at the C-terminal end, but more so in the βαβ− simulation than in the βαβ1 simulation. Parallel β-strands show up as sub (super) diagonal lines further removed than four residues from the diagonal. It can also be seen that residues 21 through 27 are close to residues 47 through 52, which is exactly the β-sheet region in the peptide. The β-sheet remains intact during the entire simulation of these peptides. Finally, it can also be seen from these distance matrices that unfolding of the C-terminal part of the α1-helix coincides with closer contact of almost all residues in the loop region that connects the α-helix to the second β-strand. βαβ2 behaves rather similar to the other βαβ unit, the β-sheet is stable and the end of the α2-helix is somewhat disordered after 1000 ps. Also at $t=1000$ ps we see events take place around residue Gly60: some local close contacts between residues 59, 60 and 61 have formed. In Fig. 6.9 we have plotted the distance matrices of the larger βαβα and βαβαββ peptides at 0, 500 and 1000 ps. As in the smaller peptide the secondary structure elements such as the two α-helices are easily recognized in the distance matrix.
Figure 6.8: Residue-Residue distance matrices from five time frames of simulations \( \beta\alpha\beta_1 \) (left) and \( \beta\alpha\beta^- \) (center) and \( \beta\alpha\beta_2 \) (right). The distance between residues \( A_i \) and \( A_j \) is the smallest distance between any pair \((i,j)\) of atoms \((i \in A_i, j \in A_j)\).

In Sim. \( \beta\alpha\beta\beta \) around residue 85 the \( 3_{10} \)-helix is visible as a dark region and it is also possible to deduce the order in which the four \( \beta \)-strands comprise the \( \beta \)-sheet from this plot \((\beta_1\beta_2\beta_3)\). From the distance matrices at 500 ps and 1000 ps it can be seen that the \( \alpha \)-helices have moved somewhat closer, giving rise to more \( \alpha-\alpha \) contacts (roughly residues 30-40 and 55-65) in both \( \beta\alpha\beta \) and \( \beta\alpha\beta\beta \) and more close \((n,n+4)\) contacts at 1000 ps in \( \beta\alpha\beta\beta \) only. The most important conclusion from these plots is however that the tertiary structure from the crystal structure is still present at the end of each of
the simulations that have at least a $\beta\alpha\beta$ unit.

Figure 6.9: Residue-Residue distance matrices of three time frames from simulation $\beta\alpha\beta\alpha$ (left) and $\beta\alpha\beta\beta\beta$ (right). The distance between residues $A_i$ and $A_j$ is the smallest distance between any pair $(i, j)$ of atoms ($i \in A_i, j \in A_j$).
6.3.5 Solvent accessible surface area

To examine the effect of the solvent on hydrophobic residues, the solvent accessible surface area was calculated with the DSSP program [65]. The program calculates the surface area and plots it per residue. We have summed up the total area for hydrophobic side chains and plotted that as a function of time in Fig. 6.10.

In some of the peptides we see that the total hydrophobic surface area is reduced over the course of the simulation (β2, βα, βαβ, βαβ1, and βαβ2). In βαβ1 we note a clear decrease in the solvent accessible hydrophobic area around 900 ps, followed by a sharp increase. Eventually the area is similar to the value at t=0. Thus, in some peptides there is a clear hydrophobic effect which seems to drive the conformational changes that occur in our simulations. When we average the solvent accessible hydrophobic surface area per residue over the simulation (Fig. 6.11), we see that some residues are protected from solvent almost completely in the larger peptides whereas they are exposed in the smaller ones.

This applies especially to the sequence Ile23 through Gly28 comprising the first β-strand and Val48 through Val51 which largely makes up the second β-strand that are exposed to solvent in the βαβ1 unit, while they are protected in the larger βαβαβ unit peptide.

Figure 6.10: Solvent accessible hydrophobic surface area during the simulations. A running average over 25 ps is given to improve clarity.
6.3.6 Salt bridges

It is interesting that oppositely charged residues can be found at the start of the first (Lys22) and the second strand (Asp46, Glu47) of the parallel $\beta$-sheet in $\beta\alpha\beta_1$. In the crystal structure of LDH Glu47 forms a salt bridge with Lys22, while the side chain of Asp46 (O$_\beta$) forms a hydrogen bond with the backbone NH of Glu47.

From the beginning of simulation $\beta\alpha\beta_1$ we see salt bridges occurring intermittently between Lys22 and Asp46 or Glu47 (Fig. 6.12). Although the Lys22 to Glu47 salt bridge is present most of the time before 1.0 ns, Glu47 is replaced by Asp46 for some time around 300 ps; after 1.0 ns Asp46 and Glu47 both form salt bridges with Lys22 but neither of them is very persistent.

In simulation $\beta\alpha\beta^-$, where we deprotonated the Lys22 side chain from NH$_3^+$ to an NH$_2$ group, hydrogen bonds between Lys22 and Glu47 also occur, but less frequently. A
hydrogen bond between Asp46 and Lys22 is never seen (Fig. 6.12). The large peak in the distance between Lys22 and Glu47 of more than 1 nm coincides with a temporary reduction in the number of hydrogen bonds in the $\beta$-sheet. From this figure it can also be seen that the motion of the negative side chains is coupled in this simulation, they move as a couple all the time, whereas in simulation $\beta\alpha\beta_1$ there is no such coupling. This may be because Glu47 is very much immobilized by the salt bridge whereas the Asp46 side chain is still free to move. In the $\beta\alpha\beta\alpha$ and $\beta\alpha\beta\alpha\beta_3$ simulations the salt bridge between Lys22 and Glu47 is present almost continuously (Fig. 6.12), implying that the residues are increasingly fixed in their position due to surrounding residues. In sim. $\beta\alpha\beta\alpha$ the Asp46 to Lys22 is sometimes present simultaneously with the Glu47 to Lys22 saltbridge.

6.4 Discussion

It has often been suggested that protein folding is a spontaneous process that starts during protein synthesis on the ribosome [1]. There has been some controversy on the topic however, because there is currently not a lot of direct experimental evidence to support this notion. Nevertheless Tsou [244] has strongly argued that protein folding during biosynthesis is at least partially cotranslational, while the extent of folding may vary for different proteins. It has also been found that the fully unfolded form of a complete polypeptide does not exist within the living cell [32]. In recent years some evidence for cotranslational folding has been provided; for example Fedorov [245] demonstrated the correct folding of the N-terminal end of the $\beta$ subunit of tryptophane synthetase during its synthesis on the ribosome. In addition, it has been demonstrated that the ribosome itself can induce folding [260, 267]. Proteins with repeating $\alpha\beta$ structures with parts of their C-terminal ends truncated, such as ras-protein [246], phosphoglycerate kinase [247], glyceraldehyde-3-phosphate dehydrogenase [249] or phosphoribosyl anthranilate isomerase [248], all fold into a stable native like structure. Interestingly, in the case of the latter enzyme, the carboxy-terminal fragment was unstructured by itself, but rapidly acquired its native structure upon binding to the N terminal domain. Experimental work has shown that N-terminal fragments of porcine muscle lactate dehydrogenase can bind dinucleotides, indicating that the Rossmann fold is present in these fragments [256, 268]. Although also C-terminal fragments of LDH were found to grossly possess the native structure, the stability of these fragments is drastically reduced compared to the native protein [33]. Taken together these observations suggest that the N terminal region of these proteins might function as a folding nucleus upon which the remainder of the structure is assembled. Recent MD simulation work on amphiphilic $\alpha$-helical peptides in aqueous solution [36, 40, 44, 202, 213, 259] have shown that most isolated $\alpha$-helices have a tendency to unfold in less than 200 ps under these conditions. In another case, of myoglobin $\alpha$-helices, this was less clear; only one of these (helix F) unfolds within one nanosecond at 298 K, of the other $\alpha$-helices helix A, G and H seem to be most stable [67]. In yet another simulation study of the Helix I/Loop I fragment of Barnase, which is assumed to be an initiation site for folding of the enzyme, the authors found the $\alpha$-helical part to be
rather stable, mainly due to hydrophobic interactions [260]; from their simulation data the authors build a detailed folding pathway for the α-helix. Nevertheless, most peptide α-helices are unstable in aqueous solution; water penetration and insertion into the peptide backbone appears to be the major mechanism that gives rise to helix instability [238, 259].

We have found that the stability of the isolated α₁-helix in water is significantly higher than most α-helices in MD simulations; the hydrophobic surface of the α-helical peptide apparently shields the backbone from water insertion. In contrast, the amphiphilic α₂-helix of LDH simulated in water is relatively unstable and behaves very similar to other α-helical peptides in MD simulations. In Fig. 6.7 it can be seen that both ends of the peptide are sometimes α-helical, most notably the C-terminal end. In the center of this α-helix is a Gly residue which facilitates water insertion in the α-helix backbone. The stability in the α₁-helix is likely to be real, because of the hydrophobic character of the peptide. Although simulation α₁ is not equilibrated after 2.0 ns, the same α-helix in simulation βα fully unfolds and subsequently refolds again which strongly suggests that the α-helical structure is most favorable for this sequence. Dill et. al. [34] have suggested that such structures could form and be stabilized through hydrophobic contacts between residues in close proximity in a polypeptide chain. Indeed it seems that upon refolding the hydrophobic surface area of the βα peptide is reduced (see

Figure 6.12: Distance between charged groups of side chains of Lys22 and Asp46 and between Lys22 and Glu47 during the MD simulations. A running average over 25 ps is given to improve clarity.
Fig. 6.10). Recent MD simulations of a very hydrophobic membrane spanning peptide in water have revealed that such $\alpha$-helical peptides are of comparable stability to the LDH $\alpha_1$-helix [225]; the A helix of myoglobin which is very stable in a simulation also has a very hydrophobic sequence in the middle of the peptide spanning residues W7-A15 [67].

MD simulations of short peptides in water have demonstrated that turn-like structures can rapidly form, disappear and reform [213, 214]. This notion has been confirmed by NMR studies of short turn-forming peptides [212, 260, 270]. Indeed our simulations of the isolated $\beta_1$, $\beta_2$ and also $\alpha_2$-helix show similar behaviour. In the simulations of longer peptides, the most flexible regions are generally the turn forming regions and the ends of the peptides. All proteins with a nucleotide binding fold contain a Gly region between the $\beta_1$ and $\alpha_1$-helix that forms a characteristic turn in the final structure [251, 271]. In the larger structures, particularly in the $\beta\alpha\beta\beta\beta$ peptide the flexibility of this region is substantially reduced. We believe that the glycines in this region may form a flexible joint that facilitates the overall folding of the $\beta\alpha\beta_1$ unit. Folding studies of turn forming peptides are in agreement with this, as they show that a central Gly residue allows the two ends of the peptide to approach each other faster [272]. It is interesting to note that the first 45 residues of adenylate kinase, which contain its nucleotide binding fold and its Gly-rich loop, can fold into a structure of two helices and three $\beta$-strands, which resembles the folding in the intact protein [273]. Like the intact protein this peptide can bind MgATP which stabilizes the fold [274]. In our simulation of the $\beta\alpha$ unit the N-terminal strand is connected to the helix by the Gly-rich loop. Due to this loop we see much more flexibility in the $\alpha_1$-helix than in the isolated $\alpha_1$-helix simulation.

Most MD simulations in water performed to date, have focused on well characterized small globular proteins or isolated peptides with elements of secondary structure. Here we have simulated isolated elements of supersecondary structure derived from LDH. We note that in our simulations the secondary and tertiary structure of the $\beta\alpha\beta_1$ and $\beta\alpha\beta_2$ units is quite well preserved. In general, the structure in these characteristic units of the Rossmann fold seems less well preserved than in compactly folded proteins, but significantly better than in linear peptides. In none of our simulations does a break occur in the $\beta$-sheet; once formed, this unit seems extremely stable. Perhaps this stability is related to the need for cooperative and simultaneous disruption of three or more hydrogen bonds, as was also found from a model study [275]. In addition, hydrophobic contacts between the $\beta$-sheet and the $\alpha_1$-helix contribute significantly to the stability of this supersecondary structure (for contribution of electrostatics, see below). In the absence of MD simulations of isolated $\beta$-sheets it is difficult at this stage to interpret our data in great detail. However, it has been reported that the $\beta$-sheet of lysozyme for example is very resistant to denaturation as well [276]. In order to compare our data for the $\beta\alpha\beta_1$ unit we also simulated the $\beta\alpha\beta_2$ unit of LDH. This unit is also comparatively stable, although somewhat less so, because the $\alpha_2$-helix is not fully preserved unlike the $\alpha_1$-helix in the $\beta\alpha\beta_1$ unit. However, also in the $\beta\alpha\beta_2$ simulation the small $\beta$-sheet is stable and does not come apart. Again hydrophobic contacts between $\alpha$-helix and $\beta$-sheet stabilize this unit.
A minimal requirement for a folding nucleus is that it can stabilize structure in polypeptide segments that are subsequently added to it. Indeed the $\beta\alpha\beta_1$ unit of LDH fulfills this requirement (see Results and below). Since the $\beta\alpha\beta_1$ unit is the first N-terminal supersecondary structure element that could fold into a stable structure, it can in principle fulfill the role of an N-terminal folding nucleus in LDH. Even when later parts of the sequence would be more stable thermodynamically in an isolated peptide, during synthesis on the ribosome they are formed later; when our $\beta\alpha\beta_1$ unit does indeed form during translation, this will be the nucleus upon which the remainder of the protein folds. Clearly, studies with synthetic peptides are required to verify this behaviour experimentally. Provided that solubility of these peptides is sufficient, such studies should demonstrate that a "nascent" $\beta\alpha\beta_1$ unit can form in aqueous solution, analogous to the well-known "nascent" alpha-helix formation [66, 212].

A simulation of the $\beta\alpha\beta^-$ unit was set up with the purpose to investigate the role of the salt bridge between Lys22 and Glu47. Removal of this electrostatic interaction gives rise to a destabilization of the $\alpha_1$-helix in the $\beta\alpha\beta_1$ unit (see Fig. 6.7), underscoring the importance of this interaction for the overall structural stability of the $\beta\alpha\beta_1$ unit. It is possible that the long-range electrostatic interaction between Lys22 and the combination Asp46 and Glu47 can help to drive the formation of the $\beta$-sheet of $\beta\alpha\beta_1$. Studies with mutated proteins and synthetic peptides have indeed shown that attractive electrostatic interactions can place secondary structure elements in their correct orientation; moreover they can even accelerate their folding [277–279].

We have also simulated regions that extend beyond the first $\beta\alpha\beta$ unit. We note that the $\alpha_2$-helix is notoriously unstable on its own, however as part of the $\beta\alpha\beta$ structure it gains considerably in structural stability. When further elements of the structure are added to form the $\beta\alpha\beta_3$ structure also the extra $\beta$-strands are stable in the simulation. The reduction in solvent accessible surface area of the $\beta$-sheet and $\alpha$-helix regions in the $\beta\alpha_3$ unit also contributes to a remarkable stabilization of these units. This is related to the more efficient internal packing of hydrophobic residues that can occur in larger units.

The inaccurate treatment of long-range electrostatic interactions requires further consideration. It has been well established that for uncorrelated dipoles (such as in water) the use of a cut-off of 1.0 nm is acceptable, but if full charges are present artefacts may occur [197]. For distances below the cut-off the charge interactions may be somewhat exaggerated due to the neglect of screening effects and the omission of electronic polarizability. For distances above the cut-off the interactions are ignored. In aqueous solution screening is such that effective charge interactions are very small: this is clearly exemplified by the difference in pK values for the two acid groups in adipic acid which are at a distance of approx. 0.6 nm and differ by 0.02 pK unit, corresponding to a negligible 0.12 kJ mol$^{-1}$. While our MD simulations do not prove in themselves that folding of this part of the protein occurs in a cotranslational manner, our results are consistent with this suggestion, because we establish here that from the $\beta\alpha\beta$ unit onwards, the folded portion is relatively
stable. This is a necessary minimum requirement for cotranslational folding. The outcome of our studies hints at the possibility that the folding of most proteins containing a Rossmann fold, which all share amino-acid homology, can in principle proceed in a similar cotranslational fashion. This agrees with other work which indicates that the first 50 amino acids of many dehydrogenases contain a folding unit [255]. Whether this process can play a role in the folding of other classes of proteins remains to be determined. However, a statistical analysis of the compactness, number of neighbouring amino acids, and stability of secondary structure in a large number of proteins suggests that it may be widespread [280].

6.5 Conclusions

Our MD simulations show that the $\alpha_1$-helix of LDH is rather stable, when compared to similar simulations of a range of $\alpha$-helices. The hydrophobic character of this helix makes it an ideal candidate for spontaneous formation through hydrophobic collapse [34]; its hydrophobic surface can subsequently interact with other hydrophobic residues of the $\beta_1$ and $\beta_2$ strands as noted throughout our simulations. The $\beta_3\beta_1$ unit is the first supersecondary structure element in the sequence of LDH; this unit is stable in our simulations, in particular its $\beta$-sheet never falls apart and the $\alpha_1$-helix is not destabilized. Its formation may be facilitated by the presence of the Gly-rich loop; indeed proper refolding of the $\alpha_1$-helix in the $\beta_3$ unit could be observed in our simulations. The $\beta_3\beta_1$ unit is further stabilized by the formation of the $\beta$-sheet and the cross-strand ion pair. If it is stable, it can function as a nucleus upon which the remainder of the structure can be assembled [254]; this is demonstrated by the markedly increased structural stability of the $\alpha_2$-helix in the $\beta_3\beta_\alpha$ unit compared to the simulation of the isolated $\alpha_2$-helix. Also the strands $\beta_3$ and $\beta_4$ are extremely stable when simulated as part of the $\beta_3\beta_\alpha\beta\beta$ unit; as isolated units, they would probably behave like the structurally flexible $\beta_2$ and $\beta_3$ strands. These MD results suggest a potential pathway for cotranslational folding of the nucleotide binding fold. Future high resolution NMR and CD experiments with N-terminal peptides of LDH with increasing length should test the validity of this proposal.

Acknowledgements

The authors would like to thank Dr. A.R. van Buuren and Dr. R.M. Scheek for useful comments. DvdS acknowledges support from the Netherlands Foundation for Chemical Research (SON) with financial aid from the Netherlands Organization for Scientific Research (NWO). HJV acknowledges support from the Medical Research Council of Canada and from the Netherlands Organization for Scientific Research (NWO) for a visiting professorship.