Bending of the Calmodulin Central Helix: A Theoretical Study

David van der Spoel, Bert L. de Groot, Steven Hayward, Herman J.C. Berendsen and
Hans J. Vogel


The crystal structure of calcium-Calmodulin reveals a protein with a typical dumbbell structure. Various spectroscopic studies have suggested that the central linker region of Calmodulin, which is α-helical in the crystal structure, is flexible in solution. In particular Nuclear Magnetic Resonance (NMR) studies have indicated the presence of a flexible backbone between residues Lys77 and Asp80. This flexibility is directly related to the function of the protein as it enables the N- and C-terminal domains of the protein to move towards each other and bind to the Calmodulin-binding domain of a target protein. We have investigated the flexibility of the CaM central helix by a variety of computational techniques: Molecular Dynamics (MD) simulations, Normal Mode (NM) analysis and Essential Dynamics (ED) analysis. Our MD results reproduce the experimentally determined location of the bend in a simulation of only the CaM central helix, indicating that the bending point is an intrinsic property of the α-helix, for which the remainder of the protein is not important. Interestingly, the modes found by the ED analysis of the MD trajectory are very similar to the lowest frequency modes from the NM analysis and to modes found by an ED analysis of different structures in a set of NMR structures. Electrostatic interactions involving residues Arg74 and Asp80 seem to be important for these bending motions and unfolding, which is in line with pH-dependent NMR and CD studies.
Structure and Dynamics of Peptides

7.1 Introduction

Calcium is an important secondary messenger in all eukaryotic cells. Specialized ATP-driven pumps secrete calcium to the extracellular environment, or into the intracellular storage organelles, creating a large calcium gradient over the membrane. During stimulation the calcium can rapidly flow down the concentration gradient into the cytoplasmic space of a cell. A host of calcium-binding proteins is responsible for translating the sudden rise in the intracellular calcium concentration into a carefully orchestrated cellular response. A unique feature of these regulatory calcium-binding proteins is that they all comprise characteristic helix-loop-helix calcium binding sites \[281, 282\]. Calmodulin (CaM), a 148-residue acidic protein, appears to be the most versatile member of this family of proteins. It is present in all eukaryotic cells; in addition it has been implicated in the activation of at least 30 different target proteins and enzymes \[282, 283\]. Considering that most other calcium-regulatory proteins generally only activate one specific target, this promiscuity is one of the most interesting features of Calmodulin. Consequently, research has focused on the molecular features that allow this protein to act efficiently on the Calmodulin-binding domains of target proteins, which do not share significant amino-acid homology.

The first distinct feature of Calmodulin is its high Met content. The preponderance of Met in its target peptide binding sites is thought to create two highly pliable, yet sticky surfaces, which can accommodate the binding of peptides with different amino acid sequences \[284, 285\]. Secondly, Calmodulin has a rather unconventional dumbbell structure, and the intrinsic flexibility of its central linker region allows a reorientation of the position of the two domains of CaM such that it can optimally interact with its target peptide. In the crystal structure of the \(\text{Ca}^{2+}\)-form of the protein, this region appears as an extended \(\alpha\)-helix. However, in solution, studies of the apo- and \(\text{Ca}^{2+}\)-form of the protein have shown that the central part of this region is flexible \[62, 63, 286–291\]. In complexes of CaM with target peptides, the \(\alpha\)-helicity of the central helix decreases further \[287, 292–294\].

The capacity of this region of CaM to adjust its secondary structure in response to the physiological conditions is of considerable interest, and has prompted us to undertake molecular dynamics (MD) simulations of this part of the protein. Two earlier MD simulations have been reported for CaM; these were done over relatively short time periods (150 ps \[295\], 600 ps \[296\]) and in the absence of full solvent; in one study artificial bends were introduced in the central linker region \[296\]. In order to extend on these earlier studies we have performed our simulations in a box containing explicit water molecules. We have done a 500 ps simulation of the entire protein (Sim. Pro) as well as a 3 ns simulation of the 28-residue isolated central helix (Sim. Pep). In this way we can compare the dynamics of the CaM central helix in the protein with the isolated peptide to investigate whether the flexibility of the backbone of the CaM central helix is a property of the \(\alpha\)-helix alone.

In general, peptide simulations do not reproduce equilibrium conditions. However, it is possible to study kinetic processes, like the unfolding of peptides \[39, 43, 44, 47, 148, 297\]. The majority of MD studies reported for isolated \(\alpha\)-helices to date involved model pep-
tides where no experimental benchmark is available, some notable exceptions being MD
simulations of the Ribonuclease S-Peptide analog [36] and the Myoglobin H-helix [40, 67].
Although no published structural data for the excised central helix of CaM is available, we
have assumed that the we can compare our simulation results to the data for the protein
crystals, at least at the beginning of the simulation; this approach allows us to test the
stability of the central helix in aqueous solution. Finally, we will analyze the motions in
the isolated CaM central helix, to test whether intrinsic motions of the α-helical peptide
are important for flexibility or unfolding of the peptide.

7.2 Methods

7.2.1 Starting structures

The starting structure for our simulation was taken from the crystal structure, pdb-entry
3CLN [3] refined at 2.2Å. Residues Asp64 through Asp93 were cut out of the protein.
Of these, Asp64 was mutated into an acetyl group by removing the NH group and the
side chain and Asp93 was mutated into an NH2 group by removing all atoms beyond
the backbone N. This left us with a peptide corresponding to residues Phe65 through
Phe92 with neutral caps on the ends. Sequence characteristics for the peptide are given
in Table 7.1.

Table 7.1: Sequence characteristics for the CaM central helix. Secondary structure was deter-
nined by DSSP[65]. (H = Helix, Space = No defined secondary structure).

<table>
<thead>
<tr>
<th>Number</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>5678901234567890123456789012</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Residue</td>
<td>FPEFLTMARKMKDTDEEEIEREAFPVF</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| Structure (3CLN) | HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH

7.2.2 Molecular dynamics

The helix was solvated in a cubic box with an edge of approx. 5.0 nm filled with 4012
SPC (Simple Point Charge) water molecules [164] for a total of 12353 atoms. Although
the shape of an α-helix would fit in a rectangular box, which would save a large number
of water molecules and therefore computer time, our simulations are so long that an α-
helix can easily rotate over 90°, thus allowing the peptide to interact with an image of
itself. Therefore it is necessary to use cubic boxes when simulating α-helical peptides.
An energy minimization of the solvated peptide was performed using the steepest descent
algorithm for 100 steps. Energy minimization and all simulations were performed using
periodic boundary conditions. Then a restrained MD simulation of 20 ps was performed, where the peptide atoms were harmonically restrained to their crystal positions with a force constant of 1000 kJ mol\(^{-1}\) nm\(^{-2}\) to allow for further relaxation of the solvent molecules. During the restrained MD run the temperature was controlled using weak coupling [168] to a bath of constant temperature \((T_0 = 300 \text{ K}, \text{ coupling time } \tau_T = 0.1 \text{ ps})\) and the pressure was controlled using weak coupling to a bath of constant pressure \((P_0 = 1 \text{ bar}, \text{ coupling time } \tau_P = 0.5 \text{ ps})\). The production run was done with the same pressure and temperature-coupling constants as the restrained run. 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 to keep the effective simulation temperature at 300 K. The Gromos-87 forcefield [11] was used with modifications as suggested in [159] and explicit hydrogen atoms in aromatic rings [162], see chapter 3 of this thesis. The time step was 2.0 fs. SHAKE [169] was used for all covalent bonds, and a twin-range cut-off criterion for non-bonded interactions was used. The short range forces (Lennard-Jones and Coulomb) were cut off at 1.0 nm and calculated every step, whereas the long range forces (Coulomb) were cut off at 1.8 nm and were updated during generation of the neighbor list, which was done every 20 fs.

For the protein simulation we used the starting structure as for the peptide simulation (3CLN [3]). All simulation details were the same as for the isolated α-helix, except as noted below. A rectangular box was used of 5.8 x 7.8 x 5.5 nm with 6768 water molecules, including the 69 crystal waters. The 4 calcium ions from the crystal structure were used, as well as 14 sodium counter ions to make up for the large net charge of -22 on the protein. A time step of 1 fs was used and SHAKE [169] was used for hydrogen atoms only. The length of the simulation was 500 ps. To distinguish both simulations we will subsequently refer to them as “Sim. Pep” and “Sim. Pro”. Both simulations were carried out using the GROMOS software package [13] on our custom-built parallel computer [52] with 32 Intel i860 CPU’s. The production run for the CaM central helix took 17.3 days on this computer, the protein simulation took 9.7 days.

### 7.2.3 Normal mode analysis and essential dynamics

A normal mode analysis (NMA) [298–300] of the CaM central helix in Cartesian coordinate space was performed using the implementation in the GROMOS simulation package [11]. We used the GROMOS vacuum force field [11], rather than the standard force field for solvated biomolecules. In this force field, the charged side chain groups are replaced by groups with only a dipole. Before the NMA, the structure was energy minimized using the same force field. In a NMA, eigenvectors are determined by diagonalization of the Hessian matrix. Both during energy minimization and the construction of the Hessian, no cut-off for non-bonded interactions was used. The root mean square deviation after minimization with respect to the crystal structure was 0.094 nm.

Essential dynamics analyses were done as described before [301] using the implementation in the WHATIF package [302].
7.3 Results

The primary focus of this work is on the CaM central helix (Sim. Pep), therefore we will not present the simulation for the intact protein (Sim. Pro) and this will only be mentioned where it is useful to compare the results of the two separate simulations.

7.3.1 Secondary structure

A very important analysis on structure of peptides and proteins in our opinion is a secondary structure analysis. We have used the DSSP program [65], which computes the secondary structure of each residue in a sequence from the atomic coordinates. Using a color code it is possible to follow the secondary structure for each residue as a function of time. This is shown for Sim. Pep in (Fig. 7.1).

![Figure 7.1: Secondary structure of the CaM central helix as a function of simulation time.](image)

From the trajectory we have extracted a snapshot every 20 ps and these coordinates were used to compute the secondary structure. Little happens to the α-helix initially, but after 700 ps residues Met76 through Thr79 lose their α-helical structure. After 1200 ps, only Lys77 seems to be non-helical, but after 1900 ps a larger portion of the α-helix starting from Met71 turns into a β-helix. This structure is then rather stable until the end of the simulation, except around 2500 ps when there is an 11 residue β-helix. The C-terminal end of the helix (starting at Asp80) is stable throughout, only small fluctuations at the very end of the α-helix occur. We have plotted the time-average of the α-helicity per residue for both Sim. Pep and Sim. Pro in Fig. 7.2, using the criterion of Hirst and Brooks [67] which is based on φ/ψ angles.

It is clear from this figure that the central residues are less α-helical in Sim. Pep. The CaM central helix in Sim. Pro is stable, it does not unfold at all under these conditions.

What happens to the peptide is illustrated by some snapshots from the Sim. Pep trajectory (Fig. 7.3). The α-helix bends in the center, but it remains rather helical on both ends. It can be seen though, that the N-terminus of the α-helix (left side in the figure) is somewhat swollen at 3000 ps, corresponding to the larger radius associated with
the \(\pi\)-helix structure type found by DSSP (Fig. 7.1).

The angle between the N-terminal part of the \(\alpha\)-helix and the C-terminal part was calculated from the average angle of three vectors in the N-terminal part and three vectors in the C-terminus. These vectors were defined by \(C_\alpha-C_\alpha\) vectors of \((n,n+7)\) residues. The 7-residue spacing was taken because it corresponds to a rotation over 700\(^\circ\) in a perfect \(\alpha\)-helix; in this way, our vectors are almost parallel to the helical axis. This procedure gives nine angles at each time step, which were averaged and plotted in Fig. 7.4. The angle is almost 180\(^\circ\), corresponding to a straight \(\alpha\)-helix, at the start of the simulation, but after 600 ps it drops quite rapidly to an average value of 90\(^\circ\), with minima below 70\(^\circ\).

### 7.3.2 RMS deviation

The RMS deviation from the crystal structure is a measure of unfolding in peptide simulations. We have calculated the RMS deviation of the backbone atoms with respect to two different crystal structures (Fig. 7.5): the starting structure (3CLN [3]) as well as the structure of CaM complexed with a peptide from myosin light chain kinase (MLCK) (1CDL [292]). In the latter crystal structure the CaM central helix is unfolded in the center (residues Arg74 through Lys77) and bent around the MLCK peptide. After a sharp rise to more than 0.3 nm (at 250 ps) the RMSD with respect to the starting structure (3CLN) decreases again to 0.08 nm at 510 ps. Then the RMSD rises steadily until 1400 ps and a plateau value of 0.7 nm is reached. The RMSD with respect to the complexed structure (1CDL) starts at 0.73 nm, it displays a small decrease to 0.63 nm around 250 ps followed by an increase to 0.7 nm around 500 ps. Then it drops slowly until a minimum of slightly less than 0.3 nm is reached around 1680 ps. Finally, the RMSD rises to a plateau value of 0.4 nm. The RMSD of the CaM central helix in Sim. Pro with respect to the starting structure (straight) fluctuates between 0.1 and 0.2 nm (data not shown).

We have also plotted the RMS deviation of the \(C_\alpha\) atoms of each residue with respect to
the starting structure, averaged over the first 500 ps in Sim. Pep and Sim. Pro (Fig. 7.6).

The rationale behind this is, that it allows us to see which residues are most flexible in the part of the simulation where the α-helix is still stable. If we disregard the first and last residues, the most flexible region is that from residue Arg74 through Asp80 in Sim. Pep. In Sim. Pro there is only a slight bump in the center of the α-helix. In the same figure we have added RMS data from a collection of 21 NMR structures of CaM complexed with the CaM binding domain of MLCK [293]. The peak in the RMSD is almost at the same position in the sequence.

![Snapshots from the MD trajectory of Sim. Pep.](image)

**Figure 7.3: Snapshots from the MD trajectory of Sim. Pep.**

Plot was made using MOLESCRIPT [170].

### 7.3.3 Side chain interactions

Interactions between side chains in the CaM central helix were studied in detail because the large number of charged side chains may be important for the stability of the peptide. We have found that there are helix-stabilizing interactions as well as destabilizing interactions; some of these interactions are present in the crystal structure, others are not. In Fig. 7.7 we have plotted the distance and angle between Phe65 and Phe68. It can be seen that the aromatic groups are quite close to each other during the entire simulation (0.59 nm on average) the angle has a two-peaked distribution, around 30° and around 150°. These angles are equivalent because of the symmetry in the Phe residues. The other aromatic pair (Phe89/Phe92) apparently does not interact, the mean distance between the aromatic planes is 0.95 nm (data not shown). Stabilizing salt bridges and hydrogen bonds are present throughout the CaM central helix, but especially in the C-terminal part.
Especially important are the interactions between Glu82 and Arg86, which is also present in the crystal structure [3] and between Glu87 and Arg90.

A number of repulsive side chain combinations are present in the sequence of the CaM central helix at \((n,n+3)\) and \((n,n+4)\) positions. We have plotted the distances between these in Fig. 7.9. There are fluctuations in the distance between repulsive side chains of about 0.2 nm, but no larger changes, except for the distance between Arg74 and Lys77. The latter changes drastically after 500 ps from 0.6 nm to 1.1 nm. After 1000 ps it slowly decreases and after 1500 ps it fluctuates around 0.6 nm. The distance between Arg86 and Arg90 is very stable until 2000 ps when a rather large fluctuation drives the two Arg residues apart. In Fig. 7.10 a number of side chain interactions that can stabilize the bent state have been plotted. There are salt bridges between Arg74 and Asp80, Lys75 and Asp80 and Arg74 and Glu84. Furthermore, a hydrogen bond exists between the Thr79 alcohol group and the Asp80 side chain from 780 ps through 2460 ps (data not shown).

7.3.4 Essential dynamics and normal mode analysis

The snapshots of the trajectory of the CaM central helix (Fig. 7.3) in Sim. Pep, make it clear that the motion of the peptide is governed by large collective displacements of the individual atoms. This feature makes it interesting to study the peptide by methods which define such collective displacements, such as normal mode analysis (NMA) [298–300] and essential dynamics (ED) [301] or principal component analysis (PCA) [68, 303]. Ideally, a large fraction of the total fluctuations in the peptide can be described as a combination of displacements along a small number of vectors, called the essential modes (for ED), or normal modes (for NMA). The techniques for this type of analysis have been described in the literature [68, 304] and we will not discuss them in detail here. An essential dynamics analysis was performed on the first 500 ps of Sim. Pep and on the integral Sim. Pro trajectories, using the essential dynamics implementation in the WHAT IF package [302]. Only the backbone atoms (N, Cα, C) were used for the analysis because our prime interest is in the backbone motion. Since there are 84 backbone atoms, there are 252 degrees of freedom. In the case of Sim. Pro the CaM central helix was cut out of the protein trajectory for this analysis, such that the peptides were same. For comparison,
we also performed an ED analysis on the last 500 ps of Sim. Pep, but except for a direct comparison of this with the previously mentioned ED analysis of Sim. Pep we will not discuss this analysis.

The fluctuations along each eigenvector $e_i$ and the cumulative fluctuations from the eigenvectors $(e_1, e_2, \ldots, e_m)$ are plotted in Fig. 7.11. A small number of eigenvectors is sufficient to describe almost all motion in the CaM central helix: the first 10 eigenvectors describe 90% of the total fluctuations in Sim. Pep, which is consistent with essential dynamics analysis work on proteins [305–307]. This means that we can define an essential subspace spanned by the first 10 eigenvectors, that covers 90% of the fluctuations of the CaM central helix in Sim. Pep. It can be seen that the first two modes describe a larger part of the fluctuations in Sim. Pep than in Sim. Pro; from $e_3$ onward the fluctuations are comparable in both simulations. As a result of this, it takes more eigenvectors to describe an essential subspace that describes 90% of the fluctuations in Sim. Pro, about 15.

To characterize the most important eigenvectors of Sim. Pep, we have plotted the structures corresponding to the minimum and maximum displacement along the first three eigenvectors plus the average structure, where the minimum and maximum displacement were taken from the first 500 ps of MD trajectory (Fig. 7.12). Both $e_1$ and $e_2$ correspond to a bending mode, which can be seen from the vibrating string like deviation from the average structure. In both cases the wavelength corresponding to the motion seems to be slightly longer than the length of the $\alpha$-helix. The motion in $e_3$ is hard to determine from this figure. Another way to characterize the eigenvectors is a component plot (Fig. 7.13). Each triplet of components of an eigenvector corresponds to the motion of an atom and we have calculated and plotted the length of this vector for each atom. The vibrating string like deviation from 0 in $e_1$ can be seen in the component plot, i.e. the atoms around residue 72 and 84 hardly move in this eigenvector, whereas there is a maximum in the amplitude around residue 80, in the center of the $\alpha$-helix. The motion in $e_2$ also has the...
bending property, but superimposed on that is a motion that has a 4 residue periodicity along the chain. If we disregard the bending motion in e2, we see that all residues on one side of the α-helix have similar Cartesian displacements, which indicates that a twisting motion is superimposed on the bending motion.

In e3, we see a standing wave with three nodes, at residues 69, 79 and 89. To compare the eigenvectors of Sim. Pep with other eigenvectors we have computed inner products of eigenvector sets. Since the eigenvectors are normalized, an inner product of one means that eigenvectors are identical. In Table 7.2 we have printed the inner products for the first eigenvectors (e1, e4) of Sim. Pep with the first eigenvectors of the ED analysis of the last 500 ps of Sim. Pep (e1end, e4end), the first eigenvectors of Sim. Pro, the first four normal modes of the CaM central helix (as described further on) and the first four eigenvectors from an ED analysis of an experimental data set.

The first bending mode (e1) is still present at the end of the simulation; it corresponds to e1end. To a large extent, e4end is contained in the first four eigenvectors of Sim. Pep, in particular e2 and e4. The correlation between other pairs of vectors is not as pronounced. Between Sim. Pro and Sim. Pep the first two modes are basically interchanged; when we sum the squares of the inner products for the first two eigenvectors we find that the similarity is 80% while the third and fourth eigenvectors of each simulation are very similar as well.

A normal mode analysis [298–300] of the excised CaM central helix was performed as described in the Materials and Methods section. The ten lowest frequencies are given in Table 7.3. To compare the normal modes to the eigenvectors we first have to extract the backbone components from the normal modes, and then renormalize the vectors n_i to a new set of vectors n_i'. This new set is not orthogonal, but projection of a vector n_i onto the eigenvectors (e1, e2, ..., e_m) still renders the complete vector back. It can be seen that
n' \textsubscript{1} corresponds almost completely to e\textsubscript{2}, while n' \textsubscript{2} is very similar to e\textsubscript{1}. n' \textsubscript{3} is contained in a combination of e\textsubscript{1} and e\textsubscript{2} while n' \textsubscript{4} is a combination of e\textsubscript{3} and e\textsubscript{4}. It is also possible to construct a set of eigenvectors from experimental structures. For this purpose we used the 21 structures of CaM complexed with the CaM-binding domain of MLCK [293] (pdb entry 2BBN). From these structures a covariance matrix was built, and the eigenvectors were determined; these eigenvectors then define the difference vectors between the structures. Here, we see that there is a clear correspondence between e\textsubscript{1} from Sim. Pep and e\textsubscript{nmr} from NMR data, but there seems to be no correlation between other pairs of vectors.

7.4 Discussion

It is generally thought that Calmodulin can adapt its conformation to enhance binding of different compounds between the two domains, which are connected by the long \alpha-helix. Indeed, it is known that CaM can bind many different proteins or peptides, often inducing \alpha-helical structure in the peptide or protein ligand [282, 285]. From the crystal structure of CaM complexed with a peptide from MLCK [292], as well as the NMR structure of a similar complex [293], it is immediately clear that the central helix (residues Phe\textsubscript{65} through Phe\textsubscript{92}) is flexible in the center, at least from Arg\textsubscript{74} through Glu\textsubscript{83}. It should be noted that there are some differences between NMR and X-Ray structure which have been attributed to the different composition of the peptide used [292, 293].

The uncomplexed CaM in aqueous solution, like the complexed CaM, has a flexible region in the central helix. Hydrogen exchange measurements [287], Nuclear Overhauser Enhancements (NOEs) [63] and order parameters [288] as determined by NMR experiments, have shown that the region from Met\textsubscript{76} through Asp\textsubscript{80} is flexible. This region is a subset of the disordered regions found in the complexed structures. It is not hard
to imagine that this inherent flexibility facilitates ligand-binding; upon complexation
the CaM central helix unrolls as much as necessary to surround the ligand; with its
hydrophobic N-terminus and hydrophilic C-terminus, amphipatic peptides can be optimally
accommodated. The fact that the CaM central helix is a rather regular α-helix in all the
plain crystal structures of Ca²⁺-CaM is not too disturbing since the crystals were grown
using organic co-solvents, and it has been shown using CD spectroscopy, that the α-helix
content of CaM increases in solvents that favor the crystallization conditions [308].

In this work, we demonstrate that the special properties of the CaM central
helix that facilitate complexation with ligands, can be reproduced in am MD
simulation of the isolated α-helix. This suggests, that the flexibility is a property of the CaM central he-
lix only which is not dependent on, or triggered by the N- and C-terminal domains. In the time
span that the α-helix is still stable (the first 500 ps of the simulation) we see that the flexibility is highest in the central region of residues Arg74 through Asp80 (Fig. 7.6). After
this period, we find that the peptide unfolds, starting with a breaking of the α-helix in the region from Met76 through Thr79 around 700 ps. The C-terminal part of the α-helix remains α-helical for the rest of the simulation whereas the N-terminal part fluctuates between α-helix and π-helix, with at one instant as many as 11 residues
involved in a π-helix (Fig. 7.1). Averaged over the whole simulation time, the α-helicity is lowest for residues Met76 through Thr79 (Fig. 7.2) which is consistent with NMR ex-
periments [287, 288]. The dynamic process of bending takes places on a somewhat shorter
time scale (500 ps) than the bending motion as determined by NMR relaxation exper-
iments of holo-CaM [288] and apo-CaM (3 ns) [290]. This can probably be explained by
the large N- and C-terminal domains of the protein which will slow down bending. Indeed,
in Sim. Pro we find that the amplitude of the bending motions is reduced considerably
as compared to Sim. Pep.

Figure 7.8: Distance between attractive side chains, defined as the shortest distance between any atom of both side chains, stabilizing the CaM central helix. A running average over 25 ps was plotted to improve clarity.
The bending angle of the CaM central helix in Sim. Pep is quite large, it bends from 180° to 70° (Fig. 7.4). Based on neutron scattering experiments, a “bent” model for Calmodulin in solution was made, in which the angle is about 125° [62], which is in between the crystal structure and our data from Sim. Pep. The small angle of 90° which seems to be the equilibrium angle for the peptide in our simulation, is not possible for the protein because the two domains would overlap. The trajectory of Sim. Pro is not long enough to see a bending motion with a large amplitude, although the essential dynamics analysis proves that the bending motion is the most important motion in the simulation.

The region of residue Met76 through Thr79 is also the most flexible region as determined by the average RMSD in the first 500 ps of the simulation, as well as experimental RMS data [293] (Fig. 7.6). The RMS deviation from the starting crystal structure (3CLN [3]) levels off after 1.3 ns, but at these high values of RMS deviation the least squares fitting procedure is not trivial and RMS is not a very useful criterion for equilibration or stability. Our side chain analyses (Fig. 7.7, Fig. 7.8, Fig. 7.9 and Fig. 7.10) show that there are distinctive side chain interactions stabilizing the N-terminal and the C-terminal end of the CaM central helix in solution. Among these interactions is a salt bridge Glu82-Arg86 that is present in the crystal structure and maintained during the entire simulation (Fig. 7.8). Another stabilizing interaction is the aromatic interaction between Phe65 and Phe68 (Fig. 7.7). Because this is an (n,n+3) interaction rather than an (n,n+4) interaction, the optimal T stacking, well known for aromatic side chains [6], is not possible. However, the conformation of the two side chains in our simulation does contribute to the stability of the α-helix. It should be noted that this interaction is not present in the crystal structure; there, the N-terminal end of the α-helix is embedded in the remainder of the protein. In the uncomplexed crystal structures (both in 1CLL [300] and 3CLN [3]) Arg74 participates in a hydrogen bonding network with residues in the

Figure 7.9: Distance between repulsive negative side chains (left) and positive side chains (right), defined as the shortest distance between any atom of both side chains. A running average over 25 ps was plotted to improve clarity.
N-terminal domain of the protein; this is also found in computer simulations [310]. In Sim. Pro we find that Arg74 interacts with Thr70 most of the time and only occasionally with Glu54, the interaction that prevails in the crystal structure. Since we have a fully solvated protein, it is not surprising that the Arg74 side chain makes other contacts as well as the crystal contact.

In Sim. Pep interactions with the N-terminal domain are not possible, instead a salt bridge between Arg74 and Asp78 forms initially, but it breaks upon bending of the helix. After 1.3 ns a very stable salt bridge is formed between Arg74 and Glu84, which prohibits restoration of the α-helix. A number of electrostatic interactions seem to be actively destabilizing the central region of the α-helix (Arg74-Lys77, Asp78-Glu82, Asp80-Glu84). There are also some side chains which are not in close proximity initially, that can further stabilize the bended form: Lys75-Asp80 and Arg74-Asp80 form close contact around 600 ps, after the α-helix is broken Asp80 forms a hydrogen bond with Thr79. A fluctuation in backbone conformation, possibly initiated by the repulsive interactions mentioned above (Fig. 7.9), makes these contacts possible.

It seems that residue Asp80 is also important for unfolding, it is engaged in repulsive interactions that destabilize the α-helix as well as in attractive interaction that can stabilize the bended form. Furthermore, it is known that aspartate residues are bad α-helix formers in peptides [311]. NMR and CD studies have shown that Calmodulin is more α-helical, and has a larger rotation correlation time, indicating a less globular structure, when the pH is decreased [312]. Since a decrease in pH to 4.5 may protonate one or more of the Glu residues in the CaM central helix, this confirms our observation that electrostatic interactions destabilize the CaM central helix.

The essential dynamics analysis of the CaM central helix in Sim. Pep shows that the large correlated motions in the CaM central helix involve bending in e1, a combined
twisting and bending in $e_2$ and a standing wave in $e_3$ reminiscent of a “first overtone” of $e_1$ (Fig. 7.13). Thus it seems, that the CaM central helix behaves somewhat like a string when the $\alpha$-helix is still formed.

When we consider the inner products (Table 7.2A), it is clear that at the end of the simulation $e_1$ is still important. An extrapolation of $e_1$, corresponding to a large displacement along $e_1$, leads to unfolding of the peptide. This motion may be induced by the electrostatic interactions described above. The good correspondence of eigenvectors between Sim. Pep and Sim. Pro (Table 7.2B) indicates that the simulations are very comparable; they basically span the same essential subspace and hence we can conclude that the simulation of the excised CaM central helix (Sim. Pep) is a good model system for the entire protein, although the amplitudes of the motion are larger in Sim. Pep. We also find that the vectors found by normal mode analysis are very similar to the essential modes (Table 7.2C). Again, we note that normal modes and essential modes span a similar subspace. On the one hand this may seem surprising, because the deviations from the average structure in 500 ps simulation are considerable. On the other hand, the most important motions in Sim. Pep resemble vibrating string-like motions, which can be described well by harmonic analysis [298-300]. The frequencies and mode of bending found by the normal mode analysis of the CaM central helix are comparable to those found by Kitao et al. for melittin [303]. Melittin is a tetramer of identical polypeptide chains consisting of a single $\alpha$-helix of 26 residues. This $\alpha$-helix has a Gly residue at position 12 and a Pro residue at position 14, and the most important motion is bending around this region of the peptide; it should be noted that melittin is bent in the crystal structure as well [313].

![Figure 7.11: Eigenvalues (bottom) and cumulative fluctuations (top) as a function of eigenvector index for Sim. Pep and Sim. Pro. Note that only the first 20 of 252 eigenvectors are plotted.](image-url)
Figure 7.12: Cα trace of the CaM central helix for the structures corresponding to the minimum and maximum displacement along the first three eigenvectors of Sim. Pep and the average structure (in black). Plot was made using MOLSCRIPT [170].

While one expects α-helix destabilization when Pro or Gly residues are present in the center of an α-helix, the CaM central helix does not have either of these residues, yet the motions are very similar to those in melittin. Surprisingly, e1 is also similar to e1nmr; apparently the difference between the 21 NMR structures [293] coincides to some extent with displacement in the direction of e1, i.e. the main difference between these structures is their bending angle around the flexible region of the CaM central helix.

It can be concluded that the simulation of the excised CaM central helix (Sim. Pep) is a good model for the behavior of the α-helix in the intact protein. The location of the bend in Sim. Pep corresponds very well to experimental data from NMR measurements [287, 288] and the dynamics of the isolated peptide is very similar to that in the intact protein. We would like to emphasize the fact that different theoretical methods (MD/ED and NMA) give consistent results in a relatively complicated system. Recently, Balsera et al. have argued that an ED analysis does not render a useful essential subspace [314]. However, the test system employed by these authors is G-actin, a 375 residue protein consisting of 4 sub domains, which is far larger than any of the proteins studied by ED so far (e.g. [301, 303–307]), while their simulation is short (500 ps) compared to their system size.
In contrast, we are very confident that our methodology is sound since our results agree very well with experimental results; this confidence is strengthened further by the consistent results from different theoretical methods.

From our results, we can conclude that the unfolding pathway of the CaM central helix is described by an extrapolation of the bending motions in the intact α-helix (e1 and e2 from our ED analysis). Although this way of unfolding may be a particular feature of the CaM central helix, the notion that an extrapolation of intrinsic motions may lead to unfolding of α-helices underlies the well-known model for unfolding through a 3_{10}-helix as well [36,315].

Figure 7.13: Atomic components of the first three eigenvectors in Sim. Pep, plotted as the length of the vector defined by the three eigenvector components corresponding to each atom.

Table 7.2: Inner products of eigenvectors between first four essential modes (EM) from Sim. Pep (e_i) and A: EM from last 500 ps of Sim. Pep (e_{i\text{end}}). B: EM from Sim. Pro (e_{i\text{pro}}). C: normal modes (n_i) and D: EM from 21 NMR structures (e_{i\text{nmr}}) [293]. The most important similarities are marked in bold.

<table>
<thead>
<tr>
<th></th>
<th>e_{i\text{end}}</th>
<th>e_{i\text{end}}</th>
<th>e_{i\text{end}}</th>
<th>e_{i\text{end}}</th>
</tr>
</thead>
<tbody>
<tr>
<td>e_1</td>
<td>0.785</td>
<td>-0.118</td>
<td>0.063</td>
<td>0.202</td>
</tr>
<tr>
<td>e_2</td>
<td>-0.052</td>
<td>0.220</td>
<td>0.072</td>
<td>-0.602</td>
</tr>
<tr>
<td>e_3</td>
<td>0.023</td>
<td>-0.365</td>
<td>0.015</td>
<td>-0.260</td>
</tr>
<tr>
<td>e_4</td>
<td>-0.167</td>
<td>0.055</td>
<td>0.399</td>
<td>0.414</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>e_{i\text{pro}}</th>
<th>e_{i\text{pro}}</th>
<th>e_{i\text{pro}}</th>
<th>e_{i\text{pro}}</th>
</tr>
</thead>
<tbody>
<tr>
<td>e_1</td>
<td>0.410</td>
<td>0.848</td>
<td>-0.010</td>
<td>0.028</td>
</tr>
<tr>
<td>e_2</td>
<td>0.852</td>
<td>-0.405</td>
<td>-0.038</td>
<td>-0.044</td>
</tr>
<tr>
<td>e_3</td>
<td>-0.074</td>
<td>0.078</td>
<td>0.714</td>
<td>-0.428</td>
</tr>
<tr>
<td>e_4</td>
<td>-0.002</td>
<td>-0.112</td>
<td>0.215</td>
<td>0.569</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>n_1</th>
<th>n_2</th>
<th>n_3</th>
<th>n_4</th>
</tr>
</thead>
<tbody>
<tr>
<td>e_1</td>
<td>-0.192</td>
<td>0.859</td>
<td>0.628</td>
<td>-0.092</td>
</tr>
<tr>
<td>e_2</td>
<td>0.902</td>
<td>0.321</td>
<td>-0.508</td>
<td>-0.054</td>
</tr>
<tr>
<td>e_3</td>
<td>-0.026</td>
<td>-0.001</td>
<td>-0.123</td>
<td>-0.657</td>
</tr>
<tr>
<td>e_4</td>
<td>0.062</td>
<td>-0.035</td>
<td>0.067</td>
<td>-0.586</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>e_{i\text{nmr}}</th>
<th>e_{i\text{nmr}}</th>
<th>e_{i\text{nmr}}</th>
<th>e_{i\text{nmr}}</th>
</tr>
</thead>
<tbody>
<tr>
<td>e_1</td>
<td>0.778</td>
<td>0.062</td>
<td>-0.119</td>
<td>0.033</td>
</tr>
<tr>
<td>e_2</td>
<td>-0.180</td>
<td>0.156</td>
<td>-0.282</td>
<td>-0.168</td>
</tr>
<tr>
<td>e_3</td>
<td>-0.018</td>
<td>-0.070</td>
<td>-0.052</td>
<td>0.077</td>
</tr>
<tr>
<td>e_4</td>
<td>-0.126</td>
<td>0.240</td>
<td>-0.024</td>
<td>-0.023</td>
</tr>
</tbody>
</table>
Table 7.3: Frequencies (in cm⁻¹) corresponding to normal modes of the CaM central helix.

<table>
<thead>
<tr>
<th>Mode</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Freq.</td>
<td>3.96</td>
<td>4.21</td>
<td>4.41</td>
<td>8.19</td>
<td>8.44</td>
<td>10.0</td>
<td>10.3</td>
<td>11.7</td>
<td>12.4</td>
<td>12.7</td>
</tr>
</tbody>
</table>

Acknowledgments

The authors would like to thank Dr. Andrea Amadei for many stimulating discussions. 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 (MRC) of Canada and the North Atlantic Treaty Organization (NATO).