Chapter 6

Hydrogen bonding pattern of cytochrome c

6.1 Introduction

Experimental determination of the hydrogen-deuterium exchange rates by means of NMR spectroscopy is a powerful tool to provide information on protein structure, stability and the onset of protein unfolding. The exchange of tacked molecular hydrogen with hydrogen of the solvent, which for small solute molecules is a fast and simple chemical exchange reaction, may exhibit complex dynamic features if studied with proteins, where hydrogen exchange may be slowed down exhibiting non-exponential behavior or even be blocked. The hydrogen exchange kinetics can be affected by the protein environment due to burial effects or properties of secondary structure where the lifetime of backbone hydrogen bonds may depend on flexibility and stability of the underlying structural motives. In some cases hydrogen exchange can occur only after local or even global unfolding events involving several amino acids, in other cases also small local fluctuations may be sufficient to allow for hydrogen exchange reactions [147, 148, 149, 150, 151, 152]. Determination of exchange kinetics of backbone amide hydrogens (NH) in a protein under native conditions can give insight into stability and fluctuations of the folded polypeptide chain with local resolution, defining more and less stable regions, according to their hydrogen exchange kinetics [150, 153, 154, 155, 156]. Under denaturing conditions one can monitor structural rearrangements as well as local and global unfolding events [157, 158]. Molecular dynamics (MD) can usefully complement backbone hydrogen exchange experiments, since it provides a detailed description of hydrogen bonds under various conditions. Fluctuations in the hydrogen bonding pattern during an MD simulation can be easily analyzed and successfully compared to experiments, yielding information on the tendency of helix and sheet structures to open up, which relates to their stability [159]. A critical aspect of such attempts is the fact that exchange cannot be seen by MD and the tendency to exchange hydrogens is inferred from motions of solvent and protein atoms which occur on a

much shorter time scale. However, the comparison of MD data to exchange experiments has successfully been applied and discussed in several studies [160, 159].
In the present study, we analyzed a small bacterial cytochrome c protein (Bcytc), for which NMR results on backbone hydrogen exchange were recently reported for the oxidized state [154]. This protein is the soluble domain of bacterial cytochrome c553 from *Bacillus pasteurii*. It contains one heme covalently bound to a pair of cysteine residues and axially coordinated to a Met/His pair. It consists of 71 residues only and is thus considered to be a minimal model for cytochrome c proteins. In spite of low sequence homology, the Bcytc fold shares structural features with the more complex mitochondrial cytochrome c [161]. The latter has been investigated by native-state hydrogen exchange providing an exhaustive picture on stability and folding/unfolding under native conditions [152, 153, 162]. We simulated the dynamics of oxidized Bcytc under native conditions to describe the properties of the hydrogen bonding pattern and its fluctuations and to interpret them in terms of stability of the different parts of the protein. We compared results from these MD simulations with current data from NMR spectroscopy of backbone hydrogen exchange in Bcytc [154] providing a consistent picture on fluctuation and stability. The aim of our work is therefore to investigate the fluctuations of bacterial cytochrome c in its native state and confirm that fluctuations on the ns timescale can be correlated to data on exchange events.
Fluctuation and disruption of backbone hydrogen bonds may be considered as onset of the protein unfolding process. Therefore, we also conducted MD simulations to unfold the protein with a bias potential similarly as we did in a previous study for a cold shock protein [163]. We tried to correlate these MD simulation data with hydrogen bond lifetimes and stability as obtained from NMR spectroscopy. Surprisingly, these MD simulations of protein unfolding could not reasonably be correlated with the experimental data and the present MD simulation results.

6.2 Materials and methods

6.2.1 Structures

The starting conformation for our MD simulations on bacterial cytochrome c553 from *Bacillus pasteurii* (Bcytc) was the average NMR structure of the oxidized protein [164] (PDB code 1K3H) that was energy minimized before usage. We also considered the crystal structure of the oxidized Bcytc (PDB code 1C75) [165] to compare fluctuations of atomic coordinates calculated from MD simulation data with the corresponding values derived from experimental B factors. In the determination of hydrogen bond length fluctuations we also considered the family of 30 conformers that were determined by NMR spectroscopy [164] (PDB code 1K3G). The experimentally determined molecular structures constitute the soluble portion of the cytochrome c553 that involves 71 amino acids with residue numbers running from 22 to 92, since the N-terminal tail, responsible for binding at the cytoplasmatic membrane [166], is lacking.

6.2.2 Molecular modeling and dynamics

The MD simulations were performed with CHARMM [29]. The heme in its oxidized state was modeled according to the CHARMM22 force field [33]. The covalently bound Cys32 and Cys35 and axially ligated Met 71 and His 36 were modeled by appropriate patches with harmonic restoring forces. The atomic partial charges for the oxidized heme were introduced by modifying the standard charge set of heme given in the CHARMM22 force field, which refers to the reduced state. The differences of atomic partial charges between oxidized and reduced heme was obtained from the charge sets calculated ab initio for both heme redox states and used for electrostatic energy computations [167]. These atom specific charge differences were added to the corresponding atom charges of the reduced heme charge model given in the CHARMM22 force field, while maintaining symmetry and total net charge of the oxidized heme state, which is -1 including the two ionized propionates. The same procedure was applied to the covalently bound Cys 32 and Cys 35 and the axial Met 71 and His 36 ligands. Also here, we referred to the charge sets used for electrostatic computations [167]. All atomic charges involving the oxidized heme are given in the supplementary material of the publication [168]. The charges from the electrostatic energy computations were not used directly, since they refer to a solute model with dielectric constant $\varepsilon = 4$ and to a specific conformation resulting in a non-symmetrical charge model.

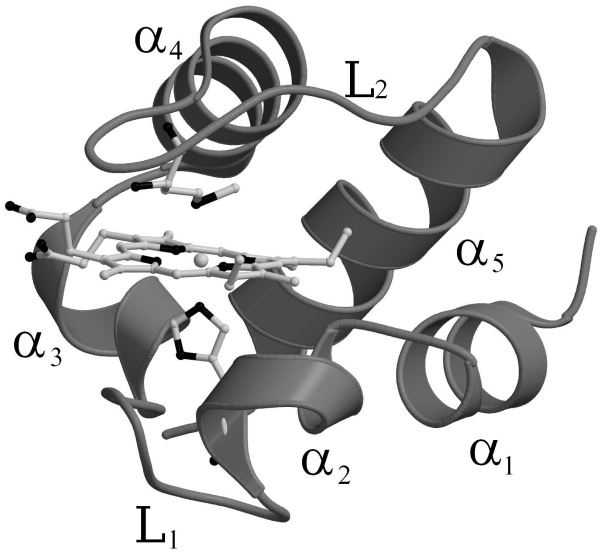

Figure 6.1: NMR structure of cytochrome c from *B. pasteurii* used as starting conformation for MD simulations (PDB code 1K3H) [164]. Secondary structure elements (helices and loops) are labeled. The heme with its two axial ligands, His36 and Met71, are displayed by balls and sticks. The picture was generated using MOLSCRIPT [133].

The solvent was modeled using explicit water molecules defined by the TIP3P water model

[44]. The protein was placed in the center of a cubic water box of size 55 Å using periodic boundary conditions. Excess water molecules with water oxygen to protein non-hydrogen atom distances smaller than 2.8 Å were removed. The resulting molecular system contained 5306 water molecules and with the protein a total of 16976 atoms. MD trajectories of 3 ns time range were generated with a time step of 2 fs using the Verlet time propagation algorithm combined with the shake algorithm to fix the hydrogen atom bond lengths. Electrostatic interactions were truncated using a switch function at a cutoff value of 13 Å. Temperature was kept constant at 300 K using the Nosé-Hoover thermostat [47, 48]. A number of test simulations were also performed with other time propagation algorithms (Verlet at constant energy or Langevin dynamics) and at temperatures of 250 K and 350 K. The MD trajectories were started after several cycles of energy minimization using the SD and ABNR modules of CHARMM to relax the molecular system. First 1000 SD minimization steps were performed for the water molecules only, while keeping the protein atom coordinates fixed. Next 500 SD steps followed by 1000 ABNR steps were applied to the protein atoms only and finally 1000 ABNR steps were applied to the whole molecular system. All MD simulations were computed in parallel using two CPUs on a Linux PC (AMD Opteron). In this way, a 3ns MD simulation applied to the molecular system with 16976 atoms needed about 430 hours real time. Root mean square (RMS) fluctuations of atomic coordinates ($\vec{r}_n$: position vector of atom n) or specific atom pair distances ($d_{mn} = ||\vec{r}_n - \vec{r}_m||$) were evaluated as standard deviations with respect to time averaged coordinates:

$$\langle \vec{r}_n \rangle = \frac{1}{N_t} \sum_{i=1}^{N_t} \vec{r}(t_0 + (i-1)\Delta t) \tag{6.1}$$

or distances:

$$\langle d_{mn} \rangle = \frac{1}{N_t} \sum_{i=1}^{N_t} d_{mn}(t_0 + (i-1)\Delta t) \tag{6.2}$$

yielding for instance the RMS fluctuations of the position of atom n:

$$\langle \Delta r_n \rangle = \langle (\vec{r}_n - \langle \vec{r}_n \rangle)^2 \rangle^{\frac{1}{2}} \tag{6.3}$$

To monitor structural changes during an MD simulation we also calculated the time evolution of the RMS deviation, $\Delta^{\text{ref}} r_n(t)$, of the whole protein, with respect to a reference set of coordinates $\{\vec{r}_n^{\text{ref}}\}$ instead of the average coordinates $\{\langle \vec{r}_n \rangle\}$:

$$\Delta^{\text{ref}} r_n(t) = \left(\frac{1}{N_{\text{atoms}}} \sum_{n=1}^{N_{\text{atoms}}} ||\vec{r}_n(t) - \vec{r}_n^{\text{ref}}||^2 \right)^{\frac{1}{2}} \tag{6.4}$$

The sum in eq. 6.4 runs over all N_{atoms} backbone atoms. In the present applications we used the starting point of the MD simulation, i.e. the energy minimized protein structure, as reference coordinate set $\{\vec{r}_n^{\text{ref}}\} = \{\vec{r}_n\}_{(t=0)}$. The RMS deviation is calculated with CHARMM using the ORIENT tool to remove translation and rotation of the whole molecule.

6.3 Results

We present here results of MD simulations for Bcytc, focusing on the data obtained in a long term trajectory of 3 ns at T = 300 K generated as described in the method section. Varying temperature in the range 250 K to 350 K did not provide significant changes in the results, thus, we do not present those in detail. The time evolution of backbone atom RMS deviation from the energy minimized starting structure at t = 0 (see eq. 6.4) is plotted in fig. 6.2. Overall the protein structure remains stable, although one can observe a constant increase of the RMS deviation in the first 1.5 ns, followed by a oscillation due to several structural transitions, like the one indicated by an arrow in fig. 6.2, left. The most relevant contributions to an increase of RMS deviation are due to loop regions L_1 and L_2 (fig. 6.1), which are participating in the structural transitions, as one can observe in fig. 6.2 right part, where the RMS deviations for the loop regions are shown individually. For both loops the total RMS deviation is larger than the average value for the whole protein, thus, pointing to a notable instability. Also helix α_1 shows a remarkable motion on the N-terminal region (see the following data on fluctuations). On the other hand, the remaining parts of the protein are more rigid, with RMS deviation around 1 Å or below for helices α_4 and α_5 with maxima below 1.5 Å (data not shown). Loop L_2 exhibits peaks in the RMS deviation, which suggest an oscillation between two states, while loop L_1 starts fluctuating significantly after about 1 ns of MD simulation. Further protein regions contributing to an increase of RMS deviation of Bcytc are the chain termini and the short helix α_3, next to loop L_1. fig. 6.3 shows the superposition of the initial Bcytc structure with a snapshot taken at 2.8 ns, which is at a maximum of RMS deviation close to the end of the trajectory. The most fluctuating segments L_1, L_2 and α_3 in fig. 6.3 are marked by arrows.

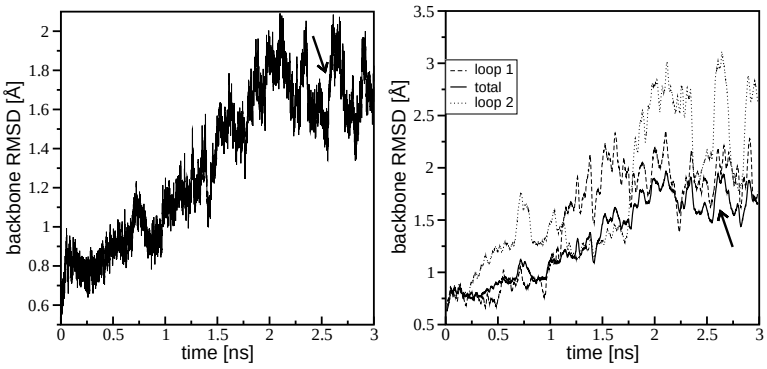

Figure 6.2: Left: Time evolution of RMS deviation according to eq. 6.4 averaged over all backbone atoms of Bcytc referring to the initial set of atomic coordinates evaluated from the 3 ns MD simulation at T = 300 K. Structural changes that occur involving loops L_1 and L_2 are marked by an arrow. Right: Time evolution of the RMS deviation, revealing the instability of loops.

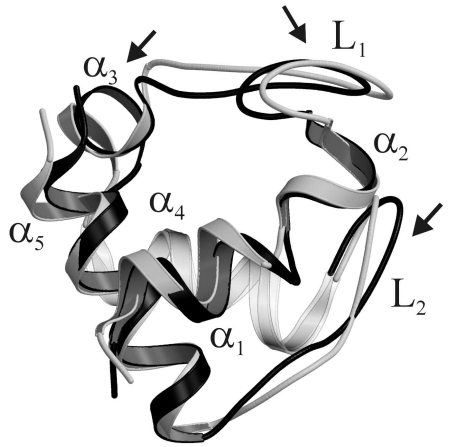

Figure 6.3: Superposition of two snapshots of the Bcytc structure, taken at 0 ns (light gray) and at 2.8 ns (dark gray). This figure generated by MOLSCRIPT [133] demonstrates the motion of loops L_1 and L_2 and of helix α_3 indicated by arrows.

6.3.1 RMS fluctuations of the Bcytc backbone

The RMS fluctuations of the Bcytc backbone atoms were evaluated from the 3 ns MD trajectory for each atom individually using eq. 6.3. By averaging these RMS fluctuations over all backbone atoms belonging to a given residue, we got a measure of the flexibility of each residue. These residue specific averaged fluctuations were compared to the experimental fluctuations given by the B factors from the crystallographic data of Bcytc [165] (fig. 6.4). The latter are obtained by applying the relation: $\langle \Delta r_n \rangle_{\mathrm{exp}} = \sqrt{\frac{3B_n}{8\pi^2}}$. In fig. 6.3, the lower open circles represent the RMS fluctuations for the 3 ns trajectory at 300 K. The closed circles are the corresponding fluctuations derived from the experimental B factors.

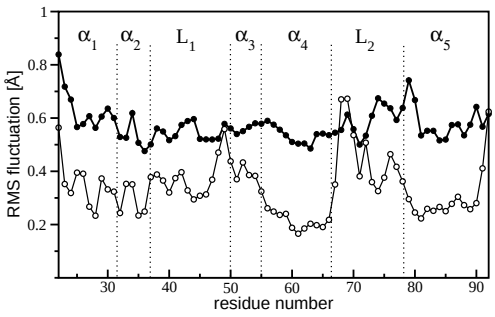

Figure 6.4: RMS fluctuations of atom positions averaged for each residue according to eq. 6.3. Only backbone atoms are considered. Closed circles report fluctuations as obtained from crystallographic B factors [165]; open circles refer to MD simulation data at T = 300 K.

The RMS fluctuations are minimal in the C-terminal part of helix α_4 (residues 60-65) and in the central region of helix α_5 (residues 80-90). The N-terminal helices, α_1 and α_2, and also the short helix α_3 adjacent to loop L_1 are on the average more fluctuating than the C-terminal helices α_4 and α_5. In the following the results on MD simulation are discussed in detail starting from the N-terminal helix. Helix α_1 shows a minimum in fluctuations at Val24 and Val28, which together with Val27 and Ala26 constitute the hydrophobic internal part of this amphipathic helix. Within helix α_2 minima of fluctuation are found at Cys32, Cys35, which are covalently bound to heme, and His36, the heme coordinating histidine that invoke additional rigidity in this regime. The RMS fluctuations become larger in the first loop region L_1 (residues 38-50) showing nevertheless minima at Leu40 and at Gly44, Ala45 and Pro46. At the negatively charged Asp49, whose side chain is oriented towards the solvent and close to the C-terminus, a significant maximum in fluctuation is located. Helix α_3 that follows shows a minimum at Ala51 and helix α_4 is the most rigid segment in the protein according to our computations, reaching the global minimum at Leu61. In the second loop region L_2 (residues 66-78) the largest fluctuations are at Gln68 and Glu69, which are mainly involved in the structural oscillation of loop L_2 (see fig. 6.2). A minimum of fluctuation is also observed at Met71, which is axially ligated to heme, and also at Gly74. The C-terminal helix α_5 shows again a remarkable rigidity in its central part. Finally, the fluctuations increase at the C-terminus as expected. The experimental values of atomic fluctuations derived from the crystallographic B factors show overall a very similar pattern, consisting of small fluctuations for helices α_4 and α_5, alternating minima and maxima in the N-terminal helix α_1 and higher peaks at the first and second loop, which nevertheless fluctuate more heavily in our calculation than in the crystal, if compared to the other protein segments. However, most local minima in fluctuation are at the same position in both sets. The experimental amplitudes of fluctuation are in general larger by a factor of 2 than the calculated values, if loop regions are excluded. This discrepancy is due to the fact that atomic fluctuations derived from crystal structures have not only a contribution from a time average as it is the case for the MD simulation data but also from an ensemble average. The time average underlying the crystallographic B factor covers a much longer time span of typically minutes instead of nanoseconds. The ensemble average adds contributions from static molecular and crystal disorder [169].

6.3.2 Hydrogen bond formation and disruption

We observed the time dependence of the hydrogen bonds formed by backbone amide and carbonyl groups during MD simulation. Hydrogen bonds were considered to be formed or disrupted at a cutoff distance of 4 Å between hydrogen bond donor nitrogen and acceptor oxygen atoms and a NH-O cut-off bond angle of 120°. This hydrogen bond criterion is generous and includes also weak hydrogen bonds, where a backbone oxygen may simultaneously be involved in more than one backbone hydrogen bond. Nevertheless, for most of the backbone CO groups a single NH donor group is the dominant hydrogen bonding partner in spite of the generous cut-

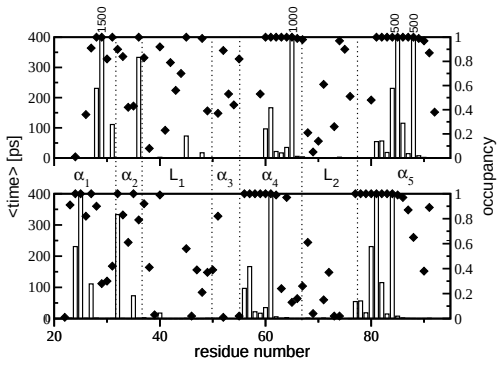

Figure 6.5: Backbone hydrogen bond stability and dynamics derived from the 3 ns MD trajectory at 300 K. Closed diamonds: fractional lifetime (occupancy) of backbone hydrogen bonds (scale on right side). Open bars: average lifetime of these backbone hydrogen bonds (scale on left side). Top: plotted versus residue numbers of the hydrogen donor backbone NH groups (NH reference). Numbers at the top of bars provide lifetimes of long-living hydrogen bonds in ps. Bottom: plotted versus residue numbers of the hydrogen acceptor backbone CO groups (CO reference). A hydrogen bond is considered to be formed, when the acceptor oxygen atom lies within 4 Å from the donor nitrogen atom in a cone of 120°. For a more detailed description see text.

off conditions. In our analysis, the most frequent hydrogen bonding partner is considered. For each donor atom (backbone nitrogen) we counted how often and how long one acceptor atom (backbone oxygen) lies inside the cone defined by the cut-off conditions. Thus, we monitored for each backbone amide group (NH reference) the hydrogen bond occupancy (fraction of time that the hydrogen bond is formed) with the most frequent backbone carbonyl group binding partner. Since some of the backbone carbonyl groups identified in this way may the majority of time be engaged in hydrogen bonds with another backbone amide group, we also monitored for each backbone carbonyl group (CO reference) the hydrogen bond occupancy with the most persistent amide group binding partner. In fig. 6.5 we plotted the occupancy of the dominant hydrogen bonds (diamonds with scale on right) formed by each NH group with a backbone CO group (NH reference, upper part) and vice-versa (CO reference, lower part) as a function of the residue number. Hydrogen bond occupancy is defined as ratio of times where the hydrogen bond is present relative to the total time length of the considered trajectory. Hence, hydrogen bonds that are never disrupted correspond to unit occupancy. Most of the hydrogen bonds that are disrupted a number of times in our trajectory possess nevertheless an occupancy nearly equal to unity. The hydrogen bonds with occupancy equal to unity are mainly α helical (i,i-4) hydrogen bonds located within the C-terminal region of helix α_4 (residues 60 to 65) and in the central region of helix α_5 (residues 81 to 88). A smaller amount of these persistent hydrogen bonds are also present in the N-terminal helix α_1 (Val27-NH with Asp23-CO, Val28-NH with Ala24-CO and Gln29-NH with Glu25-CO have an occupancy higher than 0.9) and in the

remaining helices. From the MD simulation data we found for instance that the backbone hydrogen bond in helix α_2 between the groups His36-NH and Cys32-CO has an occupancy of unity, and in helix α_3 the hydrogen bond between the groups Tyr55-NH and Ala51-CO has high occupancy (0.82). Highly populated backbone hydrogen bonds involving also non-helical residues are the bonds Leu40-NH with Gly37-CO, Ala45-NH with Cys35-CO, Ile48-NH with Leu40-CO, which are all located in loop L_1, with an occupancy of unity. In loop L_2 a highly occupied hydrogen bond is formed by Gly74-NH with Ile64-CO. In order to discriminate between hydrogen bonds that remain intact for most of the trajectory and those which are formed and disrupted more frequently, we calculated their average lifetime. A single unbinding event can reduce the lifetime by as much as a factor of two. We consider hydrogen bonds, which are disrupted less than 10 times during the 3 ns trajectory as extremely stable, or persistent, and all hydrogen bonds with a lifetime of more than 50 ps as stable. The bars in fig. 6.4 provide the hydrogen bond lifetimes of Bcytc backbone groups. In the top part the hydrogen bonds are numbered according to the residue number of the NH donor groups participating in the hydrogen bond. In the bottom part the same is done with respect to the acceptor CO groups. Five backbone hydrogen bonds persist very long during the 3 ns of the trajectory. In fig. 6.5, upper part, their lifetime is given in picoseconds at the corresponding residue position. These are hydrogen bonds in helix α_1 (Gln29-NH with Glu25-CO), in helix α_2 (His36-NH with Cys32-CO), in helix α_4 (Leu65-NH with Leu61-CO), and in helix α_5 (Val84-NH with Glu80-CO, Leu88-NH with Val84-CO). In general, most persistent hydrogen bonds involve residues in helices α_4 and α_5. In this respect, the N-terminal helix α_1 appears to be less rigid than the C-terminal helices α_4 and α_5. The backbone hydrogen bonds in the non-helical portions of the protein are generally disrupted and reformed several times during the 3 ns of MD simulation, in spite of a high occupancy. Hence, they have short lifetimes. These results may depend on the time resolution chosen for monitoring binding and unbinding events [170]. The present analysis was carried out using a high time resolution with a 50 fs time frame to record atomic coordinates. Increasing the time frame of recording the atomic coordinates to 100 fs and 0.2 ps increased all lifetimes by about a factor of 2 and 4, respectively, but did not significantly affect the differences among them (data not shown).

6.3.3 Hydrogen bond distance fluctuations

Now we like to check whether the results on hydrogen bond stability found in the preceding analysis are consistent with hydrogen bond length fluctuations. For that purpose, we monitored for the 3 ns of MD simulation data the distance fluctuations of all hydrogen bond donor-acceptor atom pairs from the backbone. These are pairs of backbone CO and NH groups with arbitrary sequence distance, which may form hydrogen bonds. Note that the sequence distance four corresponds to intra-helical hydrogen bonds. These distance fluctuations were correlated with (i) the average lifetime of the backbone hydrogen bonds, evaluated from the same trajectory as described above and shown in fig. 6.5; (ii) the corresponding hydrogen bond distance

fluctuations as obtained from the analysis of the 30 Bcytc NMR structures (PDB code 1K3G [164]) by calculating the standard deviation of the corresponding O-N atom pair distances applying Eq. 1 to the ensemble of 30 NMR conformers. The result of this analysis is shown in fig. 6.6. The lifetimes of the most stable backbone hydrogen bonds as derived from the MD trajectory at 300 K are displayed as function of the NH group residue number (bars, same data as in fig. 6.5). The solid line shows the corresponding hydrogen bond length fluctuation derived from the MD simulation data. Diamonds show the hydrogen bond length fluctuations as evaluated from the set of 30 NMR structures. Hydrogen bonds with long lifetimes exhibit small fluctuations of bond lengths, as expected. More interesting is the correlation between computed and measured fluctuations of backbone hydrogen bond lengths. We observed in both data sets small fluctuations in the helices α_4 and α_5 and large fluctuations in the N terminal part and in loop L_2. Moreover, long lifetimes (bars), which correspond to minima in the calculated distance fluctuations, often correlate with minima in the fluctuations obtained from the NMR data set. Nevertheless, there are some deviations mainly in the highly flexible regions (loops), which may, however, also be due to an inherently poor statistics, since for the analysis of the experimental data only 30 different NMR structures were available.

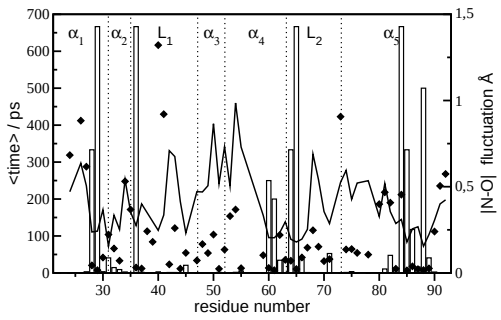

Figure 6.6: Comparison of backbone hydrogen bond lifetimes with bond length fluctuations monitored as function of the NH group residue number considering the dominant hydrogen bond partner. Open bars: hydrogen bond lifetimes derived from the 3 ns MD simulation (same as in fig. 6.4, scale on the left). Solid line: N-O distance fluctuations of hydrogen bonds formed by backbone NH with CO groups, scale on the right. Closed diamonds: N-O distance fluctuations for the same hydrogen bonds as derived from the 30 NMR conformations (PDB code 1K3G [164]), scale on the right.

6.4 Discussion

6.4.1 Fluctuations and stability of secondary structure

The fluctuations of the backbone atoms together with the hydrogen bonding pattern provide a consistent description of the structural flexibility of the protein. According to our results, it is

evident that loop regions are more flexible than helices, as one would expect. In particular the first loop L_1 strongly fluctuates on the C-terminal side, near to helix α_3, showing a considerable instability (see figs. 6.2-6.4). The second loop L_2 oscillates between two conformations (figs. 6.2, 6.3). This result suggests that both loop regions are considerably mobile and might even unfold easily under near native conditions, independently from the more rigid helices α_1, α_4 and α_5. It was observed in hydrogen exchange experiments on mitochondrial cytochrome c [153, 171, 162] that the loops account for a possible initial step in unfolding under near native conditions. Hence, our MD simulation data on bacterial cytochrome c are in this respect in agreement with these experimental results, in that they confirm some degree of similarity between the two systems. The analysis of backbone atom fluctuations for each individual residue (see fig. 6.4) allows one to distinguish local variations in protein flexibility. In loop L_1, where the first part extending from residue 37 to 43 is flexible, there is a significant minimum at Ala45 that forms a strong hydrogen bond with Cys35, which in turn is covalently bound to heme. This minimum in fluctuation may become even more pronounced by the intrinsically more rigid Pro46 whose backbone CO group forms a stable hydrogen bond with the imidazole NH group of His36, which is axially ligated to heme. The latter hydrogen bond that is also present in the starting NMR structure, is conserved during the 3 ns MD simulation at 300 K and represents a general structural feature of cytochrome c molecules, being present also in mitochondrial cytochrome c [159]. A similar variation in flexibility is visible in loop L_2, where large fluctuations were found in the neighborhood of Gln68 and Glu69, involved in the structural change. A local minimum in RMS fluctuations occurs at Met71, which is kept at its position by the axial coordination with heme. All helical regions show interesting variations in flexibility, and there are differences between the N-terminal and the C-terminal segments of Bcytc. The N-terminal part of the polypeptide chain consisting in helices α_1, α_2, loop L_1 and helix α_3 is on the average considerably more flexible than the C-terminal part where two rather stable regions are located, namely the C-terminal part of helix α_4 (residues 60-65) and the core region of helix α_5 (residues 80-90). Variation in flexibility is observed also within the helices. For instance in helix α_1 one can distinguish minimal fluctuations corresponding to buried residues (Ala24, Val28, Cys32) and large fluctuations at solvent-exposed residues (Ala26, Gln30) (fig. 6.3). In helix α_5, the internal residues Glu80, Ala81, Val84, Ala85 and Leu88, which form the most stable intra-helical hydrogen bonds, show minimal fluctuations, whereas solvent-exposed amino acids like Ala79 and Glu82 correspond to local maxima of fluctuations. The MD simulation data on fluctuations shown in fig. 6.4 are corroborated by the dynamics of intra-protein hydrogen bonds (fig. 6.5), which in turn are confirmed by the analysis of hydrogen bond length fluctuations presented in fig. 6.6. The larger fluctuations of helices α_1 and α_2 correspond to a smaller number of stable (i.e. high occupancy and long lifetime) hydrogen bonds as well as to larger fluctuations in hydrogen bond lengths.

Mithocondrial cytochrome c has been studied by means of molecular dynamics [160] to determine a correspondence with data on hydrogen exchange [153, 171, 162]. The description

in terms of stable N and C terminal helices and easily unfolding loops was corroborated by that study. It was suggested [160] that the analysis of the geometry of intramolecular hydrogen bonds may not be suited for the comparison with data on exchange since, according to the MD data, the disruption of an intramolecular hydrogen bond is not sufficient to guarantee the formation of a hydrogen bond with water, what is necessary for exchange to occur. Instead, a correlation between exchange data and degree of exposure to solvent, i.e. number of water molecules in the vicinity of the amide group was found. Nevertheless in our study the analysis of lifetime of backbone hydrogen bonds, correlated with the calculated backbone atomic fluctuations, still contributes to globally describe the flexibility of the protein in a way that is coherent with exchange experiments, as discussed in the last section. A difference from mithocondrial cytochrome c arises concerning the rigidity of the N-terminal helix, which is apparently more fluctuating in the bacterial protein.

6.4.2 Heme neighborhood

The residues involved in interactions with heme have a tendency to form stable hydrogen bonds, and correspondingly their structure is more rigid, i.e. the backbone exhibits smaller fluctuations. As we showed in the previous section, this is the case for the two heme coordinating residues, His36 and Met71 and for the covalently bound cysteines Cys32 and Cys35. Cys32-NH is forming a persistent hydrogen bond with Val27-CO. The amide proton of His36 is forming a stable, long-living hydrogen bond with Cys32-CO, whereas Met71-NH is bonded to the highly fluctuating Gln68-CO. The side chain of His36 is also involved in a strong hydrogen bond with Pro46-CO. At neutral pH both propionates (Prop5 and Prop7, see fig. 6.7) of the heme are deprotonated [164]. In the NMR structure of oxidized Bcytc the hydroxyl group of the Tyr55 side chain forms a strong hydrogen bond with the negatively charged acidic group of Prop7, with an initial O-O distance of 2.66 Å [164]. Although the hydrogen bond between Tyr55 and Prop7 is formed and disrupted several times during the MD simulation, due to oscillating Prop7 acidic oxygens, it is present during most of the 3 ns of the MD simulation, being broken for about 100 ps only during heavy structural fluctuations involving loop L_1 between 1 and 2 ns. Prop5 is not forming any significant hydrogen bond in the initial NMR structure, but during the dynamics it forms a hydrogen bond with the NH group of Gly70, which is present roughly during 80% of the simulation time. These hydrogen bonds formed by the propionates are shown in fig. 6.7. In our MD simulation Tyr55 showed an interesting behavior. Not only the side chain of Tyr55, but also its backbone amide group forms a hydrogen bond, namely with the backbone carbonyl group of Ala51, belonging like Tyr55 to helix α_3, with 2.79 Å for the initial N-O distance in the NMR structure. This hydrogen bond is highly populated and very stable in the first nanosecond (average Tyr55-NH-OC-Ala51 distance is 1.95 ± 0.05 Å) and is disrupted just before loop L_1 starts fluctuating. This confirms that the destabilization of the C-terminal end of the loop L_1 is accompanied by a structural change in helix α_3. The presence of the Tyr55-NH-OC-Ala51 hydrogen bond in the native Bcytc is corroborated by the

hydrogen bond length fluctuations derived from the 30 NMR structures. This hydrogen bond is present in roughly one half of the 30 structures from set of NMR conformations (PDB code 1K3G). The interaction between Tyr55 and Prop7 is pH dependent. At pH 5.1 Prop7 is protonated in the oxidized form of Bcytc, as one can infer from the crystal structure [165], where it forms a hydrogen bond with the carbonyl oxygen of Ala47. While the aromatic ring of Tyr55 is nearly parallel to the heme plane at neutral pH, it is rotated away from Prop7 at lower pH in the crystal structure and forms a hydrogen bond with a crystal water [165]. The hydrogen bond of the Tyr55-OH group with Prop7 present in the NMR structure of oxidized Bcytc, appears more pronounced in the NMR structure of the reduced form [172] (PDB code 1N9C). A hint for a possible functional relevance of Tyr55 comes from sequence alignment. The tyrosine at this position is a conserved residue in the Gram-positive bacterial cytochrome c family, as one can observe for instance in the sequence alignment in ref. [164]. Only in one of nine cases, tyrosine is replaced by a lysine, which, however, is a positively charged hydrogen donor group that may even form a salt bridge with Prop7. This is not the case in cytochrome c molecules from eukaryotic organisms, where tyrosine is more often replaced by a hydrophobic phenylalanine or isoleucine (see the alignment in Ref. [165]). According to current experimental data and calculations on bacterial c553 [173] and mitochondrial cytochrome c [174] this tyrosine is not directly involved in the electron transfer process. But, it may play a role in stabilizing the local hydrogen bond network and in tuning the heme redox potential.

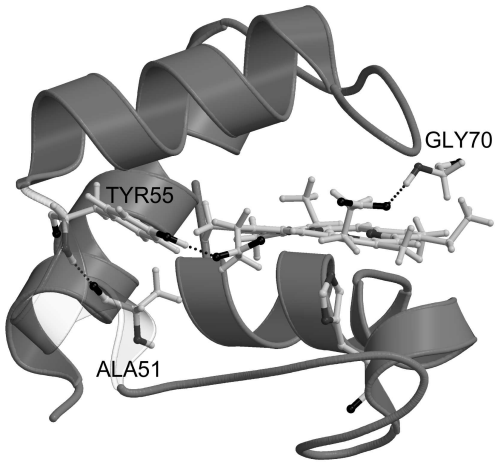

Figure 6.7: Snapshot from the 3 ns MD trajectory at 300 K showing hydrogen bonds formed by the two heme propionates. Prop7 (on the left side) is hydrogen bonded to the side chain of Tyr55, whose backbone NH group is interacting with Ala51-CO. Prop5 (on the right side) is forming a hydrogen bond with Gly70-NH. The picture was generated using MOLSCRIPT [133].

6.4.3 Comparison with backbone NH exchange experiments

In this section we discuss our results in comparison with the NMR experiments of hydrogen exchange published by Bartalesi and coworkers [154], which offer a general and exhaustive description of hydrogen bond stability and equilibrium unfolding of Bcytc under near native conditions. Our analysis is intended to be purely qualitative, because opening events related to hydrogen-deuterium exchange occur in reality on time scales which are much longer than 3ns, so that we cannot observe them in our simulation. However if exchange events correlate with local fluctuations we should be able to detect such correspondence. Based on their data, Bartalesi et al. [154] sort residues to the same opening unit, if they exhibit a similar value of free energy for hydrogen exchange. This suggests a common unfolding process for residues belonging to the same opening unit. Residues belonging to an opening unit with large free energy of hydrogen exchange will open only upon global unfolding, while for low values of free energy residues may exchange the backbone hydrogen through local fluctuations. If in the latter case the low value of free energy is for several nearby residues essentially the same, the hydrogen exchange process may also be connected with sub-global structural fluctuation as for instance a partial helix unwinding. Low free energies of exchange are experimentally found for three subsets of residues, which in decreasing order of opening ability are:

(A) the residues 26-32 comprising the second half of helix α_1 and the initial part of helix α_2,

(B) the central part of loop L_2 (residues 71-76),

(C) the residues 67-68 and 81-83 connecting loop L_2 to the C-terminal part of helix α_4 and to the N-terminal part of helix α_5, respectively.

These three opening units (A, B, C) are all characterized by a relatively low value of free energy of opening and are supposed to change via local fluctuations. At a higher value of free energy of hydrogen exchange, but still compatible with local fluctuations, there is an opening unit (D) comprising the C-terminal end of helix α_4, namely residue 66. There are also two opening units with a large free energy of hydrogen exchange close to the global free energy of protein unfolding, suggesting that these regions exchange hydrogens only upon global unfolding. These are:

(E) the core of the C-terminal helix α_5 (residues 84-89),

(F) the residues 60-65 belonging to the central portion of helix α_4.

To compare our results with the NMR experiments on hydrogen exchange in Bcytc, we calculated RMS fluctuations averaged over the residues belonging to the six different opening units as defined above. We also calculated the average occupancies for backbone hydrogen bonds formed within the same opening units, by using the MD simulation data on hydrogen bond statistics presented in fig. 6.5. Both data sets are shown as histograms in fig. 6.8. In the

top part, average occupancies for the hydrogen bonds are classified according to the opening units from A to F. In the bottom part, average RMS fluctuations are shown for the same protein segments. Large values of average occupancy of hydrogen bonds are in general associated with small average fluctuations, as discussed before. The classification from A to F in terms of increasing stability, found in experiments, is to a large extent reflected by our MD simulation results, especially by referring to average occupancies. The increasing free energy of opening observed experimentally goes along with increasing average occupancy of hydrogen bonds, with the exception of unit A, which in our results shows a higher average stability of hydrogen bonds than unit B and C, comprising loop L_2. The analysis in terms of average RMS fluctuations (lower part in fig. 6.8) exhibits a further difference in ranking, since unit D shows smaller fluctuations than unit E. Nevertheless, both data sets agree in that they strongly distinguish between two stability regimes, namely one set characterized by low occupancy and high fluctuations comprising units A, B and C and one "high stability" set containing units D E and F, with lower fluctuations and hydrogen bonds with larger occupancy.

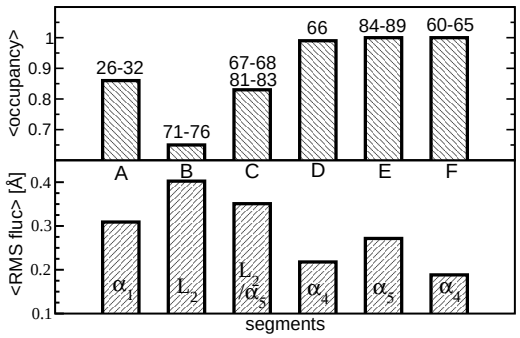

Figure 6.8: Occupancy and fluctuations of hydrogen bonds grouped in opening units A to F of increasing stability. Top histogram: average occupancy of backbone hydrogen bonds formed by amide hydrogen atoms within selected sequence segments of Bcytc as indicated at the top of the bars evaluated from the 3 ns MD trajectory at 300 K. Bottom histogram: RMS fluctuations of backbone atoms of the same segments of Bcytc. Corresponding secondary structure elements are also indicated. The data are classified as opening units from A to F according to the ease of hydrogen exchange monitored by NMR spectroscopy [154] From left to right these sequence segments have a decreasing tendency to belong to structural elements, which open up with ease. For more details see text.

The N-terminal helix α_1 of Bcytc is fluctuating at amplitudes that are larger than the ones present in the C-terminal part of the protein, namely in helix α_5, thus, featuring hydrogen bonds that are less stable. This leads to the conclusion that the amide hydrogens in helix α_1 may exchange more promptly with the bulk solvent, what is actually observed. The C-terminal helix α_5 possesses a number of stable hydrogen bonds with small amplitudes of fluctuation, which slows down hydrogen exchange. In summary, the analysis in terms of RMS fluctuations yields overall qualitative agreement with the experiments.

6.5 Conclusions

The structural stability is investigated by MD simulations for a small bacterial cytochrome c molecule. A well defined pattern correlating stability and flexibility of secondary structure elements under native conditions was found. We identified differences in stability and dynamics among the helical units of the protein, distinguishing a rather flexible N-terminal half involving the helices α_1, α_2 and α_3 from the two more rigid helices α_4 and α_5 at the C terminus. We observed large fluctuations involving the first loop region L_1 and a structural oscillation between two states in the second loop region L_2. These loop regions were experimentally recognized to act as initial unfolding units in mitochondrial cytochrome c [153]. The comparison with results from backbone hydrogen exchange experiments provides a consistent picture, which allows to combine our observations on fluctuations in hydrogen bonding pattern with the experimental values of free energies for hydrogen exchange. Interestingly, the difference in stability between the N and C-terminal helices, which we found in our MD simulation study of Bcytc, was also observed in cytochrome c from very different organisms [153, 175]. Moreover, we analyzed the hydrogen bond network around the heme group, highlighting the conserved residue Tyr55, which is hydrogen bonded to one propionate group of the heme. In recent computations it was demonstrated how the screening of the heme propionate charges is tuning the heme redox potential [167]. The present study suggests that Tyr55 may play such a role.