The effects of urea and of pH on protein structure studies by molecular dynamics simulation
Mueller, Daniela Sung-Mi
2.1 Introduction

Cytochrome c is a phylogenetically “ancient” enzyme that is found throughout the eukaryotic and prokaryotic kingdoms. It is an integral component of the electron transport chain in mitochondrial and bacterial membrane systems and therein essential for the generation of energy in the cell. Like many redox-active proteins c-type cytochromes carry a prosthetic group, an iron ion-coordinating heme group, which is covalently bound to two cysteine residues and presents the catalytic centre of the protein (Fig. 2.1). Cytochrome c-550 (cytc) was isolated from the rhodobacterium Paracoccus versutus (the genus is also known under the name *Thiobacillus*) [Lommen et al. 1990]; the suffix “c-550” derives from the heme type C and the visual absorption peak at a wavelength of 550 nm. Cytc has been studied extensively, and was classified as a class I cytochrome c like the archetypal mitochondrial cytochromes c and many bacterial cytochromes [Diederix et al. 2001].

In addition to their role in electron transport, c-type cytochromes also show peroxidase activity in the presence of hydrogen peroxide $\text{H}_2\text{O}_2$ when the heme-iron is in the ferric state $\text{Fe}^{3+}$ [George 1953; Harbury & Loach 1960; Ubbink et al. 1992]; hence they are also referred to as peroxidase mimics. Peroxidases are enzymes that contain a heme group and efficiently catalyse the oxidation of substrates using the environmentally innocuous oxidant hydrogen peroxide [Diederix et al. 2001]. Whereas essential\(^1\) peroxidases are generally very instable, C-type cytochromes are stable and remain highly soluble even under conditions of

\(^1\) An essential peroxidase is an enzyme with the primary activity of a peroxidase.
2.1 Introduction

extreme heat, acidity and basicity. Furthermore the covalent linkage of the heme prosthetic group to the protein matrix prevents dissociation of the catalytic moiety from the protein. [Diederix et al. 2001] As a consequence the potential to use c-type cytochromes as peroxidases in commercial applications has attracted much interest. Wild-type ferricytochrome c-550 has a second order rate constant for the reaction of $H_2O_2$ of $k = 22.8 \pm 0.4 \text{M}^{-1}\text{s}^{-1}$ for hydrogen peroxide concentrations $[H_2O_2] < 100\text{mM}$ [Worrall et al. 2005a], which is $\approx 1000$-fold lower than the activity of microperoxidases [Diederix et al. 2001; Harbury & Loach 1960]. Canters and co-workers have related this low activity to the low accessibility of the heme pocket for potential peroxidic substrates which are too large to access the catalytic centre within the protein [Diederix et al. 2001, 2002a,b]. They found the peroxidase activity of cytC unfolded in guanidinium hydrochloride to be increased $\approx 1200$-fold, and that this increase correlated well with unfolding, indicating better accessibility of the active site to the substrate [Diederix et al. 2002a]. From this they concluded that in partially unfolded cytC the conformation of the protein matrix around the heme pocket was expanded, allowing peroxide substrates to access the active site more easily and to be catalysed more readily [Diederix et al. 2004, 2002a; Worrall et al. 2005a].

The structure of cytC from P. versutus was solved by X-ray crystallography (Fig. 2.1) [Worrall et al. 2005b]. As is common to all c-type cytochromes the heme group is bound to cytC by the Cys-Xaa-Xaa-Cys-His (CXXCH) heme binding sequence motif, where the side chain thiol groups of Cys15 and Cys18 form thioether linkages with the two vinyl groups of the heme [Ambler 1963; Worrall et al. 2005b]. The Fe(III) ion is hexa-coordinated by the four pyrroline N-atoms and two protein ligands, the $N^\tau$-atom of the His19 side chain and the

![Figure 2.1 Cytochrome c-550 (cytC) from Paracoccus versutus. Cross-eyed stereo image of the X-ray crystallographic structure of cytC. The heme group, Cys15 and Cys18 are depicted in stick representation. The secondary structure shown was assigned using the STRIDE software [Frischman & Argos 1995; Humphrey et al. 1996]. Colour legend: purple – α-helix, blue – 3_10-helix, yellow – β-sheet, cyan – turn, white – loop, orange – heme-iron.](image-url)
Unfolding of cytochrome c-550 in urea

Sδ-atom of Met100 [Dickerson et al. 1971; Gadsby et al. 1987][IUPAC 1975]. The alkaline unfolding of cytochromes c has been studied extensively, and multiple structural species have been found to occur at high pH \( \geq 9.5 \) [Brautigan et al. 1977; Hong & Dixon 1989; Theorell & Åkesson 1941]. Various studies have also found evidence that the dissociation of the axial heme ligand Met100 precedes unfolding under alkaline conditions [Diederix et al. 2002b; Hoang et al. 2003; Worrall et al. 2005b], and that the methionine ligand is exchanged with a lysine [Brautigan et al. 1977; Gadsby et al. 1987]. Also increased peroxidase activity was linked to the loss of the heme-iron ligands, which would otherwise cover the heme-iron and block the access to solvent [Worrall et al. 2005a]. After removal of the Met-ligand the iron would be free to bind a peroxide anion at the vacated coordination site [Dumortier et al. 1999]. In addition to the coordination by a lysine residue a large structural rearrangement of the ligand loop was observed during the unfolding process [Rosell et al. 1998; Winkler 2004; Worrall et al. 2005b]. Therefore specific rearrangements of the ligand loop could result in the enlargement of the heme cavity.

If cytc could be stabilised in an expanded conformation, it could be used as an industrial peroxidase [Diederix et al. 2001]. If furthermore the conformation of cytc could be easily and reversibly manipulated, the peroxidase activity could be regenerated multiple times and cytc applied as a recyclable industrial peroxidase. In the reference experiments of this study, guanidinium hydrochloride (Gdn·HCl, C(NH\(_2\))\(_3\)·Cl\(^–\)) was used to unfold cytochrome c-550 [Diederix et al. 2002a,b; Worrall et al. 2005a,b]. In hydrochloric acid guanidinium forms the cationic hydrochloric salt C(NH\(_2\))\(_3^+\) with a p\(K_\alpha\) of 12.5, which is thought to denature proteins by longer-range electrostatic interactions [Camilloni et al. 2008]. However, in common MD simulation setups, which use cut-off truncation of coulombic interactions to limit computational costs, electrostatic interactions take a very long time to converge [Baker et al. 1999; Bergdorf et al. 2003]. Ionic solvents are therefore less suitable for simulations of intrinsically slow processes like protein unfolding. For this reason neutral urea ((H\(_2\)N\(_2\)CO) was used for the in silico solvation of cytc.

Urea is widely-used as a chemical denaturant for the unfolding of proteins, and has been used for the denaturation of cytochrome c [Ahmad et al. 1996; Bhuyan 2002; Brunori et al. 2003; Creighton 1979; Elöve et al. 1994; Fedurco et al. 2004; Khoshtariya et al. 2006; Latypov et al. 2006]. While protein denaturation by urea has been studied and documented extensively [Bolen & Rose 2008; Ibarra-Molero et al. 2001; McCarney et al. 2005], the exact mechanism by which urea induces unfolding is still not known. Direct interactions with the protein as well as indirect effects via interactions with the co-solvent water have been proposed [Makhatadze & Privalov 1992; Timasheff 1993]. One experimental study directly compared urea with guanidinium with respect to interactions with the protein, where urea was found to form hydrogen bonds directly with the peptide group but guanidinium chloride not [Lim et al. 2009]. Different urea models have been parametrised for use with different
MD force fields [Caballero-Herrera & Nilsson 2006; Chitra & Smith 2002; Smith et al. 2004a] and urea has been the subject of numerous MD simulation studies [Åstrand et al. 1994; Idrissi et al. 2000; Stumpe & Grubmüller 2007a,b; Trzesniak et al. 2004; Wallqvist et al. 1998], including being used as a denaturant for proteins [Bennion & Daggett 2003; Caballero-Herrera et al. 2005; Schiffer et al. 1995; Smith et al. 2004b; Smith 2004; Stumpe & Grubmüller 2008; Tirado-Rives et al. 1997; Tobi et al. 2003; Weerasinghe & Smith 2003]. For reviews of the experimental and theoretical studies of urea and protein denaturation I refer the reader to Beck et al. [2007]; Dobson et al. [1998]; Graziano [2002]; Idrissi [2005]; McCarney et al. [2005]; Tran & Pappu [2006].

In an experimental study it was found that in 6 M urea cytochrome c unfolded to a molten-globule state in which the global conformation was strongly altered throughout; at the same time a rather compact and native-like tertiary structure was retained with a “chemical-like” equilibrium between the native-like and non-native local folding of the metal coordination [Khoshtariya et al. 2006]. Different types of denaturing conditions have been used to study the unfolding of c-type cytochromes, e.g. thermal, alkaline, acidic or chaotropic denaturation [Bhuyan & Kumar 2002; Kimura et al. 2007; Kumar et al. 2006; Víglaský et al. 2000; Wain et al. 2001; Winkler 2004]. Easily applicable and reversible conditions for the unfolding of P. versutus cytc were tested in the form of basic solvents and detergents which are commonly used to unfold proteins [Diederix et al. 2004, 2002b]. The reason for employing MD simulation to study cytc in urea was to observe possible structural changes of cytc in atomistic detail, in particular in the region of the heme cavity. MD simulation has been used to study c-type cytochromes in the past [Arteca et al. 2001; Cai et al. 1992; Collins et al. 1992; Kieseritzky et al. 2006; La Penna et al. 2007; Mao et al. 1999a, 2001], also to investigate thermal unfolding [Mao et al. 1999b; Prabhakaran et al. 2004; Roccatano et al. 2003].

This chapter presents molecular dynamics (MD) simulations of cytc that were performed to determine the structure and stability of cytc during unfolding by the chaotropic denaturant urea. The simulations were used to evaluate the conformational stability of the protein at three different urea concentrations, 0, 6 M and 10 M. The questions asked were:
i) Is it possible to simulate the urea-induced unfolding of cytochrome c-550? and

ii) Can urea stabilise the enzyme in a partially unfolded state?

Positive answers would suggest that urea, and possibly other denaturants, can be used to reversibly manipulate the catalytic activity of cytochrome c-550 peroxidase and recycle the enzyme for application as an environmentally friendly, “green” reduction catalyst.

### 2.2 Methods

#### Systems simulated

The starting structure for the simulations was taken from the X-ray crystal structure 2BGV [Worrall et al. 2005b] of ferricytochrome c-550 from Paracoccus versutus. In this structure residues Glu2–Pro120, the backbone nitrogen atom of Asp121 and the heme group were resolved. To complete the structure residues Asp121 and Ala122 were modelled. The heme group was linked to the protein by covalent bonds to the sulphur atoms of Cys15 and Cys18. This was done to allow for the loss of the iron-sulphur interaction observed experimentally during unfolding under alkaline conditions or associated with high concentrations of denaturant [Hoang et al. 2003; Worrall et al. 2005b]. Coordination of the heme-iron by the nitrogen atoms of the pyrroline groups and the His19.Nτ-atom was modelled by covalent bonds. As the objective of this study was to probe the unfolding of cytC, coordination of the ferric heme-iron Fe(III) with Met100 was modelled using a combination of Lennard-Jones and coulombic interactions. His19 and His118 were both singly-protonated on the Nπ-atom.\(^2\)

The side chains of Arg and Lys residues were protonated and Asp and Glu side chains deprotonated. Water molecules present in the crystal structure were removed. The structure was placed in a triclinic, periodic box with a distance of at least 1.1 nm between the protein and the sides of the box. The protein was then solvated with SPC-water [Berendsen et al. 1981] in combination with a flexible neutral urea model [Duffy et al. 1993; Smith et al. 2004a]. Four systems containing different proportions of water and urea molecules corresponding to 0, 6 M and 10 M urea were simulated. Five sodium ions Na\(^+\) were added to neutralise the net charge of the system. Table 2.1 lists the compositions and simulation times of all the systems simulated.

\(^2\) The nitrogen atoms of the imidazole ring of histidine are denoted by pros (“near”), abbreviated π, and tele (“far”), abbreviated τ, to show their position relative to that of the side chain [IUPAC 1975].
2.2 Methods

<table>
<thead>
<tr>
<th>[urea] /M $^a$</th>
<th>$N_{\text{molecules}}$ $^b$</th>
<th>$N_{\text{atoms}}$ $^c$</th>
<th>A% $^d$</th>
<th>t /[ns] $^e$</th>
</tr>
</thead>
<tbody>
<tr>
<td>water MD1</td>
<td>4883</td>
<td>15854</td>
<td>0</td>
<td>230</td>
</tr>
<tr>
<td>water MD2</td>
<td>4883</td>
<td>15854</td>
<td>0</td>
<td>150</td>
</tr>
<tr>
<td>water MD3</td>
<td>4883</td>
<td>15854</td>
<td>0</td>
<td>230</td>
</tr>
<tr>
<td>6 M urea</td>
<td>3367</td>
<td>15306</td>
<td>33.1</td>
<td>230</td>
</tr>
<tr>
<td>10 M urea MD1</td>
<td>2508</td>
<td>15057</td>
<td>51.3</td>
<td>170</td>
</tr>
<tr>
<td>10 M urea MD2</td>
<td>4064</td>
<td>23653</td>
<td>51.3</td>
<td>70</td>
</tr>
</tbody>
</table>

$^a$ molar concentration of urea, g·mol$^{-1}$ [Weast et al. 1986]

$^b$ number of molecules

$^c$ total number of atoms in the system

$^d$ anhydrous solute weight per cent of urea, g solute/100 g solution [Weast et al. 1986]

$^e$ simulation time

Simulation parameters

All simulations and energy minimisations were performed using the GROMACS package version 3.1.4 [van der Spoel et al. 2005] in conjunction with the GROMOS96 force field version 43A1 [van Gunsteren et al. 1996]. The temperature ($T = 300$ K) and pressure ($p = 1$ bar) were held constant by weak coupling to an external bath [Berendsen et al. 1984]. A twin-range cutoff of 0.8 nm and 1.4 nm was used in conjunction with a reaction field ($\varepsilon_{RF} = 54$) to correct for the truncation of electrostatic interactions beyond the long-range cutoff [Tironi et al. 1995]. Interactions within the short-range cutoff were updated at every time step. Longer-range interactions, together with the pair list, were updated every 5 steps. Bonds were constrained by applying the LINCS algorithm [Hess et al. 1997]. The integration time step was 2 fs. Initial atom velocities were assigned from a Maxwell distribution at $T = 300$ K. Each system was first energy minimised, then simulated for 100 ps to equilibrate the solvent, where the positions of all heavy atoms in the protein were harmonically restrained for 50 ps, and the heavy atoms of only the backbone for another 50 ps, using a force constant of 500 kJ mol$^{-1}$nm$^{-1}$.

Three MD simulations were run of the system containing only water and the protein, starting from different sets of initial velocities at 0 ps. One simulation each of the 6 M urea system and the two 10 M urea systems were run.
2 Unfolding of cytochrome c-550 in urea

Analysis

Root mean square deviation (RMSD) The RMSD of the positions of the backbone atoms was determined after performing a least squares fit of each configuration in the trajectory to (a subset of) the backbone atoms of the starting crystal structure. Short term fluctuations within 1 ns were smoothed by performing a 500 point-running average.

Radius of gyration The mass-weighted radius of gyration $R_g$ of a structure is related to the distances of all atoms to the centre of mass (COM) of the structure and was calculated according to

$$R_g = \left( \frac{\sum_i ||r_i||^2 m_i}{\sum_i m_i} \right)^{1/2}$$  (2.1)

where $r_i$ is the position of atom $i$ with respect to the COM of the structure, and $m_i$ the mass of atom $i$.

Secondary structure The secondary structure was determined based on the DSSP definitions of secondary structure [Kabsch & Sander 1983].

Molecular contacts Atoms the centres of which are separated by less than or equal to 0.6 nm were considered to be in contact. The number of contacts between the solvent molecules, the heme group and the protein were determined for configurations sampled every 200 ps.

Solvent accessible surface (SAS) SAS areas of each protein residue were estimated using the algorithm of Eisenhaber et al. [1995], with the diameter of the probe sphere being 0.14 nm. The SAS areas per residue were determined by summing the SAS areas per atom.

Radial distribution function (RDF) The cumulative number (CN) of atoms (e.g. solvent) with respect to a reference point or group of atoms (e.g. the protein) was calculated by integrating the number of atoms (particles) $N$ as a function of the distance $r$ to the reference group. The density of particles $\rho(r)$ at a given distance was obtained by normalising $CN(r)$ against the probe volume. For a one-particle distribution as described here, the radial distribution function (RDF) is defined as the ratio of the density $\rho(r)$ at a given distance with
2.3 Results

respect to the average density $\rho$ of the bulk

$$g(r) := \frac{\rho(r)}{\rho} \quad (2.2)$$

The RDF $g(r)$ is then related to the CN according to

$$\text{CN}(R) := \int_0^R \rho g(r) 4\pi r^2 \, dr = N_R \quad . \quad (2.3)$$

From equation (2.3) the RDF $g(r)$ of a (sub)set of $N$ particles from a total set $S$ can be numerically approximated as

$$g(r) \approx \frac{\Delta N_r V_S}{\frac{4}{3}\pi (r^3 - (r')^3) N_S} , \quad r = r' + \Delta r \quad (2.4)$$

with $N_S$ the total number of particles in $S$, $V_S$ the volume taken up by the particles in $S$, and $\Delta r$ the thickness of the spherical shells used for determining $\Delta N_r = \text{CN}(r) - \text{CN}(r')$.

In this study, equation (2.4) was used for calculating a) the RDF of the atoms of cytc with respect to the heme-iron, and b) the RDF of water and urea atoms around cytc. In a) the set $S$ contained all heavy atoms of the protein, in b) $S$ comprised all urea and water atoms combined.

2.3 Results

Changes in secondary and tertiary structure

In order to investigate the effect of urea on the structural stability of cytc the positional RMSD was analysed with respect to the initial structure. Figure 2.2 shows the RMSD of the backbone atoms of the protein as a function of simulation time for the simulations in water, 6 M urea and 10 M urea. As can be seen in Figure 2.2A there was an initial rapid increase in the backbone RMSD of cytc during all of the simulations, irrespective of the urea concentration. Among the multiple simulations in water (Fig. 2.2A left panel), the RMSD values that are reached vary between 0.3 and 0.5 nm. The simulations in 6 M and 10 M urea reached RMSD values of up to 0.55 nm after simulation times of 200 ns.

To determine whether the structural deviations observed were due to the overall unfolding of the structure, or whether they were dominated by less structured parts of the protein, e.g. the loop regions, the backbone RMSD of only the helical regions was determined while maintaining the same fit to all backbone atoms as above. The helix RMSD, plotted in Fig-
2 Unfolding of cytochrome c-550 in urea

Figure 2.2 Root mean square deviation (RMSD) of the backbone atoms of cytc during MD simulation in water and urea. The RMSD was calculated with respect to the initial structure of the simulations (X-ray crystal structure 2BGV), of A) the entire backbone, and B) the backbone of the helices. Left panels: RMSD of cytc simulations in water, simulation 1 (black line), 2 (dark grey), and 3 (light grey); middle: in 6 M urea; right: in 10 M urea, simulation 1 (black), and 2 (dark grey).

ure 2.2B, had significantly lower values in all simulations, indicating that deviations of the extra-helical regions dominate the RMSD. Nevertheless the RMSD values were high with the helix RMSD in simulation MD3 in water reaching values comparable to those obtained in 6 M and 10 M urea. This raised the question of whether cytc was unstable in water using the GROMOS96 force field 43A1. Visual inspection of the trajectories however indicated that the protein maintained a high degree of structural integrity even in urea, suggesting that the high RMSD reflected changes in the relative positions of secondary structure elements as opposed to loss of structure per se.

The trajectories were analysed in terms of secondary structure to elucidate the extent to which the initial structure was maintained. Secondary structure assignments were made based on the DSSP definitions. For the DSSP analysis sets of five configurations from each trajectory were sampled between 142 to 150 ns of simulation. The predominant secondary
Table 2.2 Secondary structure of cytC in water and urea. The secondary structure of each residue in the X-ray crystal structure of cytC (2BGV [Worrell et al. 2005b]) that was used as the initial configuration, and in later configurations from the MD simulations in water and urea was determined according to the DSSP definitions. The values listed represent a consensus over five configurations taken every 2 ns from the trajectory in the interval 142–150 ns, except for the shorter simulation “10 M urea 2’ (* ) every 1 ns in the interval 63–67 ns. Letter code: ~ - random coil, E - β-sheet, B - β-bridge, S - bend, T - turn, H - α-helix, G - 3_10-helix; helix and turn structures "H", "G" and "T" are highlighted in dark grey, beta structures "E" and "B" in light grey.

<table>
<thead>
<tr>
<th>residue no.</th>
<th>2BGV</th>
<th>water 1</th>
<th>water 2</th>
<th>water 3</th>
<th>6 M urea</th>
<th>10 M urea 1</th>
<th>10 M urea 2 *</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>T T</td>
<td>B T T</td>
<td>T T</td>
<td>E E</td>
<td>T T</td>
<td>E E</td>
<td>B S T T</td>
</tr>
<tr>
<td>11</td>
<td>H H</td>
<td>H H</td>
<td>H H</td>
<td>H H</td>
<td>H H</td>
<td>H H</td>
<td>H H</td>
</tr>
<tr>
<td>21</td>
<td>H H</td>
<td>H H</td>
<td>H H</td>
<td>H H</td>
<td>H H</td>
<td>H H</td>
<td>H H</td>
</tr>
<tr>
<td>31</td>
<td>H H</td>
<td>H H</td>
<td>H H</td>
<td>H H</td>
<td>H H</td>
<td>H H</td>
<td>H H</td>
</tr>
<tr>
<td>41</td>
<td>H H</td>
<td>H H</td>
<td>H H</td>
<td>H H</td>
<td>H H</td>
<td>H H</td>
<td>H H</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>residue no.</th>
<th>2BGV</th>
<th>water 1</th>
<th>water 2</th>
<th>water 3</th>
<th>6 M urea</th>
<th>10 M urea 1</th>
<th>10 M urea 2 *</th>
</tr>
</thead>
<tbody>
<tr>
<td>51</td>
<td>H H</td>
<td>H H</td>
<td>H H</td>
<td>H H</td>
<td>H H</td>
<td>H H</td>
<td>H H</td>
</tr>
<tr>
<td>61</td>
<td>T T</td>
<td>T T</td>
<td>T T</td>
<td>E E</td>
<td>T T</td>
<td>E E</td>
<td>B S T T</td>
</tr>
<tr>
<td>71</td>
<td>H H</td>
<td>H H</td>
<td>H H</td>
<td>H H</td>
<td>H H</td>
<td>H H</td>
<td>H H</td>
</tr>
<tr>
<td>81</td>
<td>H H</td>
<td>H H</td>
<td>H H</td>
<td>H H</td>
<td>H H</td>
<td>H H</td>
<td>H H</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>residue no.</th>
<th>2BGV</th>
<th>water 1</th>
<th>water 2</th>
<th>water 3</th>
<th>6 M urea</th>
<th>10 M urea 1</th>
<th>10 M urea 2 *</th>
</tr>
</thead>
<tbody>
<tr>
<td>91</td>
<td>H H</td>
<td>H H</td>
<td>H H</td>
<td>H H</td>
<td>H H</td>
<td>H H</td>
<td>H H</td>
</tr>
<tr>
<td>101</td>
<td>T T</td>
<td>T T</td>
<td>T T</td>
<td>E E</td>
<td>T T</td>
<td>E E</td>
<td>B S T T</td>
</tr>
<tr>
<td>111</td>
<td>H H</td>
<td>H H</td>
<td>H H</td>
<td>H H</td>
<td>H H</td>
<td>H H</td>
<td>H H</td>
</tr>
<tr>
<td>121</td>
<td>H H</td>
<td>H H</td>
<td>H H</td>
<td>H H</td>
<td>H H</td>
<td>H H</td>
<td>H H</td>
</tr>
</tbody>
</table>
2 Unfolding of cytochrome c-550 in urea

structure assignments for each residue in the configurations sampled are given in Table 2.2 along with the secondary structure of the X-ray crystal structure 2BGV. It can be seen that in water all five alpha-helices remained stable after 150 ns of simulation, whereas at both urea concentrations alpha-helix 1 and alpha-helix 5 were partially unfolded. Most of the β-sheet and β-strand of the initial cytc conformation remained intact in all the simulations, with some slight shifts along the primary sequence.
2.3 Results

**Heme accessibility**

The packing of the protein matrix around the heme was examined as a means of quantifying structural changes within the protein related to the accessibility of the heme to the solvent and substrates. To measure the accessibility of the heme group the distribution of distances of the heavy atoms of the protein from the heme-iron was analysed. This was done by calculating the cumulative numbers (CN) of the radial distribution function (RDF) of protein atoms with respect to the heme-iron. The CN values for each simulation were averaged over configurations sampled every 0.2 ns from 130 to 150 ns and are plotted in Figure 2.3A. As can be seen all curves have a similar sigmoidal shape with the curves shifted towards larger distances \( r \) with increasing urea concentration. The shifts of the curves show that in urea the protein atoms on average shifted further away from the heme group than in water. The slope of the curve also decreased with increasing urea concentration and the distribution of protein atoms around the heme iron in urea was less dense than in water. The three CN curves of the simulations in water were averaged to serve as a control. The protein packing in each simulation was compared to the control by calculating the difference \( \Delta CN \) between the average of the three CN curves in water (control) and the CN in 6 M and in 10 M urea. Figure 2.3B clearly shows that the difference in the number of protein atoms between the CN and the control was approximately twice as large in 10 M urea as in 6 M urea. The difference plots furthermore show that in comparison with the average density in water the decrease in protein density in urea was most pronounced at a radius of approximately 1.7 nm around the heme group. The radial distribution functions (RDF) calculated from the CNs are plotted in Figure 1.2C and D. These show that in urea the density of protein atoms within \( \approx 1.7 \) nm of the heme-iron was lower than in water, and that this was compensated by higher densities at distances \( \geq 1.7 \) nm.

The radius of gyration \( R_g \) is another structural property that gives an estimate of the size of a molecule. The \( R_g \) calculated for the configurations extracted at 150 ns of simulation were similar to the initial configuration, varying by less than 1 % (data not shown).

Another measure of the accessibility of the heme group within the protein matrix is the number of contacts with solvent molecules and with the protein environment. Atoms that lay within 0.6 nm of the reference group were considered to be in contact. Solvent contacts with the heme group give a measure of (solvent) accessibility, protein contacts a measure of occlusion (from the solvent). In Figure 2.4 the numbers of solvent and protein contacts with the heme group are plotted as a function of simulation time. From panels A and B it can be seen that the number of contacts between the heme group and solvent increased in a urea concentration-dependent manner while the number of contacts with the protein decreased. In line with the increasing number of solvent contacts an increase in solvent accessible surface area (SAS) of the heme group was also observed (data not shown). A more detailed analysis
of solvent interactions with the protein are presented in the following section.

The largest changes in the number of protein and solvent contacts with the heme group were observed for simulation 2 in 10 M urea (Fig. 2.4). The number of solvent contacts increased by approximately twice as much after 45 ns compared to the value reached in the other 10 M urea simulation 1 after 170 ns. Configurations of cytc at these time points in the 10 M urea simulations are shown in Figures 2.5 and 2.6. As can be seen in Figure 2.5 A, in simulation 1 the protein matrix remained tightly packed around the heme group, whereas in simulation 2, shown in panel B, the protein matrix opened, exposing the heme to the solvent. In panels B and C it can also be seen that where in simulation 2 (pink) a cavity is formed, the configuration from simulation 1 (blue) and the initial structure (green tube) both enclose the heme group to both sides of the heme plane. In panel D the protruding cartoon traces of the configurations in 10 M urea illustrate that in particular helix 1 and several loops deviate from the initial configuration (green surface).

Figure 2.6 shows the backbone configurations from the simulations in 10 M urea superimposed upon the initial structure. It can be seen that in simulation 1 the tertiary structure overall deviated more from the initial structure than in simulation 2 (Fig. 2.6 A and B respectively), in particular helix 1, and the loop regions adjacent to helix 1 and helix 2. At the same time the middle section of this loop and the loop between helix 5 and helix 6 remained close to either side of the heme in simulation 1. In simulation 2 on the other hand, these loop...
2.3 Results

Figure 2.5 Legend see next page.
sections and also the domain around helix 1 shifted away from the heme. These shifts are also clearly visible in Figure 2.5D.

Protein-solvent interactions

Interactions between the different solvents and cytc were analysed with the aim of finding indications in regard to how urea affects the protein structure. This question was approached by comparing the distributions of water and urea. To this end the radial distribution functions (RDF) of solvent molecules around the protein were determined and are plotted in Figure 2.7. Comparing the RDF of water with the RDF of urea in the simulations in urea, it can be seen

Figure 2.6 Conformation of cytc in 10 M urea. Superposition of configurations from the simulations in 10 M urea upon the initial structure: A) after 170 ns of simulation 1 (blue), and B) after 42 ns of simulation 2 (pink) in 10 M urea, the heme group is coloured blue. Protein structures are shown in the cartoon representation of VMD [Humphrey et al. 1996], the heme group in stick representation. Configurations were fitted onto the backbone structure of the helices in the initial configuration. In the initial structure the heme group is coloured by atom type and the cartoon trace according to the secondary structure: purple – α-helix, yellow – β-sheet, cyan – turn, white – loop.
2.3 Results

![Graphs](image)

**Figure 2.7** Cumulative number (CN) of the radial distribution function (RDF), and RDF of solvent species with respect to cytc. A) The CN(r) of the atom density of water and of urea with respect to cytc was determined by summing the number of solvent atoms that were located within a distance r of the protein, in configurations at 150 ns. Legend: line colours indicate the solvent composition: black – water simulations 1–3, dark grey – 6 M urea, light grey – 10 M urea simulations 1 & 2; line styles indicate the solvent species analysed: continuous lines – water, dashed lines – urea. B–C) The RDF(r) of the atom density derived from the CN given in panel A, on two different length scales; legend as in A.
that a shell of urea molecules coated the protein, displacing the water molecules away from
the protein. Interactions between the sodium ions and the protein decreased with increasing
urea concentration (data not shown).

**Heme ligands**

Experimentally it has been shown that specific interactions between the protein and the heme
group correlate with either the native or (partially) unfolded cyt c [Brautigan et al. 1977;
Diederix et al. 2002b; Dumortier et al. 1999; Gadsby et al. 1987; Hoang et al. 2003; Worrall
et al. 2005b]. Here specific interactions were used as indicators of whether partially unfolded
states of cyt c occurred in the simulations.

![Figure 2.8](image-url)

**Figure 2.8** Hydrogen bonding between Lys99 and the heme group. Occurrence plot for the
hydrogen bond between the Lys99 Nε-atom and one of the carboxyl substituents CγOδ of the
heme group. The occurrence of the hydrogen bond was determined for configurations sampled
every 0.02 ns and is plotted as circles against the simulation time. Black vertical lines indicate the
ending times of the shorter simulations. Graphs 1–6 from top to bottom refer to simulations: 1) 
water simulation 1, 2) water 2, 3) water 3, 4) 6 M urea, 5) 10 M urea 1, 6) 10 M urea 2.

One specific interaction is the coordination between Met100 and the heme-iron. Loss
of this interaction was previously shown to be associated with alkaline unfolding [Hoang
et al. 2003; Worrall et al. 2005b]. To determine whether structural rearrangements during
the simulation lead to the loss of the Met100 heme interaction, the distance between the
Met100.Sδ-atom and the heme-iron was monitored. In all simulations the initial distance of
0.29 nm was lost within the first picoseconds of MD after which the values fluctuated around an average distance of 0.42 ±0.04 nm (data not shown).

The unfolding of cytc has also been reported to lead to interactions between the heme-iron and one or more lysine residues [Rosell et al. 1998; Worrall et al. 2005b]. Therefore the distances between all lysine residues and the heme-iron in the simulations were monitored. Under all solvent conditions the lysines remained at least 0.66 nm away from the heme-iron and remained at similar distances throughout the simulations (data not shown). Furthermore a number of hydrogen bonds between other functional groups in the heme and lysine residues were detected. An analysis of the correlation between the hydrogen bonding propensity and the concentration of urea indicated that a hydrogen bond between the Lys99 side chain and one of the heme carboxyl groups CγδO2 was anti-correlated with the presence of urea. Figure 2.8 contains an existence plot for this hydrogen bond showing that the interaction was found in all simulations in water, but only at the beginning in simulation 2 in 10 M urea and at isolated time points in the other simulations with urea. For significant proportions of the simulations with urea the distances between these two groups were larger than 0.4 nm (data not shown) and therefore well above the hydrogen bonding distance.

2.4 Discussion

The analysis of the secondary and tertiary structure of bacterial cytochrome c-550 during the simulations in water and urea (Tab. 2.2) showed that at all urea concentrations the protein remained intact during 150 ns. The persistence of residual structure after denaturation in urea was already reported for other proteins [McCarney et al. 2005; Shortle & Ackerman 2001]. The high RMSD values that were observed nonetheless (Fig. 2.2) were due to unstructured regions such as the highly fluctuating termini and loop regions, and slight shifts of the helices relative to each other (data not shown). The partial unfolding of the N- and C-terminal helices in both 6 M and 10 M urea were likely to be due to the higher mobility and solvent exposure that are commonly observed for protein termini and which resulted in an increase in solvent interactions that propagated along the sequence and destabilised the adjacent helices.

The distribution of protein atoms around the heme-iron was used as an indicator for the packing of the protein (Fig. 2.3). The wider distributions of the protein atoms in urea indicated that the molecule occupied a larger volume. This could be regarded as a tertiary structural effect, where secondary structure elements remained intact but moved apart slightly. Less density in packing implied an increased solvation of the protein, with urea and/or water molecules invading the protein structure and driving secondary structure elements apart. This was confirmed by a decrease in protein-internal interactions in the simulations with urea.
In 6 M urea the number of contacts between the heme group and the solvent molecules (Fig. 2.4) took intermediate values between what was observed for water and for 10 M urea. The changes in the SAS area of the heme group (data not shown) and the RDF cumulative number (CN) of protein atoms around the heme-iron were also correlated with the urea concentration (Fig. 2.3). This suggests that the structural changes observed were urea-dependent.

With the secondary structural elements staying largely intact (Tab. 2.2) and a medium-high RMS deviation from the initial backbone conformation (Fig. 2.2), the partially unfolded protein remained stable during 150 ns in 10 M urea. However, solvent interactions with the protein and heme in 10 M urea were still increasing towards the end of the 150 ns trajectory (Fig. 2.4). This suggests that the protein structure and/or solvent distribution had yet to reach equilibrium, and even more extensive unfolding might be observed if the simulations were extended. From experiment it is known that high concentrations of urea fully denature and unfold proteins. In an MD simulation with 10 M urea, i.e. at approximately the highest concentration typically used in protein denaturing experiments, a protein would be expected to fully unfold. Thermal and chemical unfolding have been determined experimentally to take place on the microsecond time scale [Huang & Oas 1995; Mayor et al. 2000], e.g. the cold shock protein CspB, which is known to be an extremely fast-folding protein, was reported to unfold at a rate of about 100 s\(^{-1}\) in 8 M urea [Schindler & Schmid 1996]. Accordingly the unfolding of cytc in urea can be expected to require more time than was accessed in this study. The reason simulation 1 in 10 M urea was stopped after 150 ns was that as a result of the expansion of the unfolding protein, there were increasing interactions with its periodic image.

Despite the apparent stability of cytc in urea, structural changes were nevertheless evident from the MD simulations. In particular the protein environment surrounding the heme group and solvent interactions with the heme were affected (Fig. 2.4), suggesting a urea-dependent widening of the heme cavity, as proposed by Caners et al. from experimental observations [Diederix et al. 2002a; Worrall et al. 2005a]. Particularly simulation 2 in 10 M urea displayed a notable widening of the heme cavity (Fig. 2.5). In this simulation, and also to a lesser extent in the other simulations with urea, a decrease in the number of contacts between the protein and the heme group indicated a progressive separation of the protein matrix from the heme (Fig. 2.4). This allowed more urea molecules to access the cavity, as was indicated by the increasing interactions of urea molecules with the heme group (data not shown). The configurations from the simulations in 10 M urea in Figure 2.5 clearly show that in simulation 2 the heme cavity opened, exposing the heme group to the bulk solvent. In simulation 2 the displacement of the loop containing the heme ligand Met100 away from the heme iron (Fig. 2.6) corresponded to one of the structural rearrangements observed in alkaline denatured cytc experimentally [Rosell et al. 1998; Winkler 2004; Worrall et al. 2005b].
2.4 Discussion

Interestingly the backbone conformation of the structure with the open heme cavity in this simulation was quite similar to the initial structure. This preservation of structure incorporating an exposed catalytic centre is a very promising prospect for an enhanced peroxidase that can be readily regenerated. In contrast to what was observed for 10 M urea the number of contacts of the heme group remained constant in 6 M urea (Fig. 2.4), suggesting that at the lower urea concentration the system had reached equilibrium and that the structure of cytc was stable. Whether urea had a direct effect on the conformation of the protein as a solvent, resulting in an expansion of the cavity by mechanical refolding, or whether specifically the heme group interacted preferentially with urea, leading to a coating of the heme group with urea that displaced the protein matrix, cannot be determined from the present simulations.

The urea molecules preferentially associated with the protein, forming a shell around cytc, while the water molecules were displaced away from the protein (Fig. 2.7). This result is in line with recent experimental data [Lim et al. 2009]. There was a very small accumulation of water molecules close to the protein, which probably reflects the internal hydration of the protein. A comparison of the RDF of water molecules around cytc in water and in urea showed that in urea even the first shell of water molecules was displaced by urea molecules (Fig. 2.7B). In the first shell of solvation the clustering of urea molecules around cytc was only slightly weaker in 6 M urea than in 10 M urea, suggesting a strong interaction between the protein and urea. The enhanced local density of urea around the protein could result from hydrophobic interactions between the protein and urea, as proposed in the literature as a possible mechanism for protein denaturation by urea [Kamoun 1988; Nozaki & Tanford 1963; Smith et al. 2004b]. The degree of clustering observed in the two systems in 10 M urea was less pronounced in the smaller system in simulation 1 than in the larger system in simulation 2. This was most likely an artefact due to the size of the system and the greater number of urea molecules available to cluster around the protein in the larger system. The decreased interactions of the sodium ions with the protein in urea was presumably a consequence of the displacement of water and the higher solubility of sodium ions in water [Serjeant & Dempsey 1979].

The Met100.Sδ-atom was uncharged in the present simulation setup. Thus no interaction with, or coordination of, the heme-iron was expected. Under alkaline conditions lysine residues were proposed to act as a sixth ligand for the central iron atom and replace the methionine ligand during the unfolding of cytc [Gadsby et al. 1987; Russell et al. 2000; Ubbink et al. 1994]. In the simulations a large distance was maintained between the lysine residues and the heme-iron in all cases (data not shown). A significant conformational change would be necessary to bring a lysine close enough to the heme-iron to enable them to interact. Although a structural rearrangement of the ligand loop was observed in simulation 2 in 10 M urea, it did not bring Lys99 close to the heme-iron (Fig. 2.6). However, the native hydrogen bond between the Lys99 side chain and one of the heme carboxyl groups was disrupted in all
the simulations with urea (Fig. 2.8). Disruption of this hydrogen bond is required in order for the hinge region between \(\alpha\)-helices 4 and 5 (Ser94–Asn107 in Paracoccus versutus cytc) to shift and Lys99 to displace the heme-iron ligand Met100, a phylogenetically conserved motion which was shown experimentally in alkaline unfolded cytc [Dumortier et al. 1999]. Removal of the Lys99 side chain away from the heme results in increased solvent exposure of the heme [Worrall et al. 2005b]. In the simulations this hydrogen bond proved to be extremely urea-sensitive to an extent that it almost did not form at all in two out of three simulations with urea.

Considering i) the expansion of the heme cavity, ii) the coating of the protein by urea, and iii) the displacement of the protein matrix away from the heme group, urea appeared to have a loosening effect on the tertiary structure of the protein which subsequently lead to a widening of the heme pocket.

### 2.5 Conclusion

Bacterial cytochrome c-550 was shown in MD simulations to partially unfold in solution with urea. Partial unfolding occurred on a time scale of 100 ns, and the extent of unfolding depended on the urea concentration. The partial unfolding was associated with increased access of the solvent to the prosthetic heme group in the cavity of the protein matrix, and a decrease in interactions between the heme group and the protein matrix. In 10 M urea, a significant widening of the heme cavity was observed, exposing the heme group and letting the solvent molecules access the catalytic site more easily, as proposed by Diederix et al. [2002b].

In the systems simulated containing urea, a solvation shell of urea molecules formed around the cytc, displacing the water molecules away from the protein. The number of interactions between the solvent and the protein in urea clearly increased compared to the simulations in water. At the same time the hydrophobic interactions within the protein diminished, which has been previously proposed as an effect of urea [Nozaki & Tanford 1963]. Enhancement of the solubility of the protein is believed to be one of the main mechanisms suggested for protein denaturation by urea [Kamoun 1988], and the simulations here provide evidence for this mechanism.

Progressive unfolding was observed in 10 M urea, while in 6 M urea and in neat water the protein conformation remained stable and the interactions between the solvent, the protein and the heme group remained constant within 230 ns of simulation. The density of the protein decreased in a urea concentration-dependent manner while the tertiary structure
remained intact. Only isolated terminal regions of two helices unfolded, similarly following a urea concentration-dependent trend, which suggests that the extent of unfolding can be modulated by the urea concentration. Therefore solvation in urea promises to be a suitable means for the partial and reversible unfolding of cytochrome c-550 peroxidase, as a recyclable, switchable enzyme for the environmentally safe reduction of peroxides.