Universiteit
Leiden
The Netherlands

Modelling copper-containing proteins

Bosch, Marieke van den

Citation
Bosch, M. van den. (2006, January 18). *Modelling copper-containing proteins*. Retrieved from https://hdl.handle.net/1887/4361

Version:	Corrected Publisher's Version
License:	Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden
Downloaded from:	https://hdl.handle.net/1887/4361

Note: To cite this publication please use the final published version (if applicable).

Chapter 6

Simulation of the Substrate Cavity Dynamics of Quercetinase

Summary

Molecular dynamics (MD) simulations have been performed on quercetin 2,3 dioxygenase (2,3QD) to study the mobility and flexibility of the substrate cavity. 2,3QD is the only firmly established Cu-containing dioxygenase known so far. It catalyses the breakage of the O-heterocycle of flavonols. The substrates occupy a shallow and overall hydrophobic cavity proximal to the metal centre of the homo-dimeric enzyme. The linker connecting the C-terminal and N-terminal domains in the monomer is partly disordered in the crystal structure and part of it forms a flexible lid at the entrance of the substrate cavity. This loop has been tentatively assigned a role in the enzyme mechanism: it helps lock the substrate into place. The dynamics of this loop has been investigated by MD-simulation. The initial coordinates were taken from the crystal structure of 2,3QD in the presence of the substrate kaempferol (KMP). After equilibration and simulation over 7.2 ns the substrate was removed and another equilibration and simulation of 7.2 ns was performed. The results show that the structures of the free enzyme as well as of the enzyme–substrate complex are stable in MD-simulation. The linker shows strongly enhanced mobility in the loop region that is close to the entrance to the substrate cavity (residues 154–169). Movement of the loop takes place on a timescale of 5–10 ns. To confirm the conclusions about the loop dynamics drawn from the 7.2 ns simulation, the simulation was extended with another 8 ns. When substrate binds into the cavity the loop orders remarkably, although mobility is retained by residues 155–158. Some regions of the loop (residues 154–160 and 164–176) move over a considerable distance and approach the substrate closely, reinforcing the idea that they lock the substrate in the substrate cavity. The enthalpic component of the interaction of the loop with the protein and the KMP appears to favour the locking of the substrate. Two water molecules were found immobilised in the cavity, one of which exhibited rotation on the picosecond timescale. When the substrate is removed, the empty cavity fills up with water within 200 ps.

6.1 Introduction

Quercetin 2,3-dioxygenase (quercetinase or 2,3QD) catalyses the conversion of flavonol derivatives into their corresponding phenolic carboxylic acids (depsides). The 3D-structure of 2,3QD from Aspergillus japonicus was solved recently by X-ray diffraction (XRD) techniques (Fusetti *et al.*, 2002; Steiner *et al.*, 2002), which opened intriguing perspectives on the reaction mechanism of the enzyme: the deformation of the substrate observed when it binds in the substrate cavity is thought to lower the activation barrier for the conversion of the substrate. Moreover, on the basis of the crystallographic data, it was surmised that the loop connecting the two domains of the monomeric unit of the enzyme might play a role in stabilising the substrate when it is bound in the enzyme cavity.

The substrate parent compound, flavonol, is depicted in Figure 6.1. The substrate range of the enzyme is not restricted to flavonol, but encompasses derivatives that contain one or more hydroxy substituents at positions in the rings A and/or B. Kaempferol (KMP), the flavonol derivative with OH groups at positions 5, 7 and 4', is considered in the present study. The aerobic conversion of KMP into the corresponding depside under concomitant loss of carbon monoxide, is summarised in Figure 6.2.

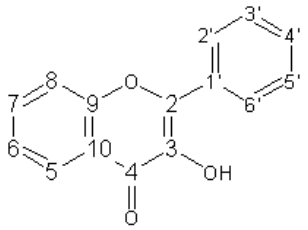


Figure 6.1. A representation of the flavonol skeleton. The flavonol ring and atom nomenclature is given in the Figure. Quercetin is obtained by inserting hydroxy (OH) substituents at positions 5, 7, 3' and 4'. Alternative 2,3 QD substrates are kaempferol where OH substituents are present at 5,7 and 4', galangin (OH at 5 and 7), fisetin (OH at 7, 3' and 4') and myricetin (OH at 5,7,3',4' and 5'). Compared to quercetin the relative degradation rate for the different substrates by *Aspergillus flavus* is 2.54, 0.30, 0.58 and 0.87 respectively (Oka *et al.*, 1972).

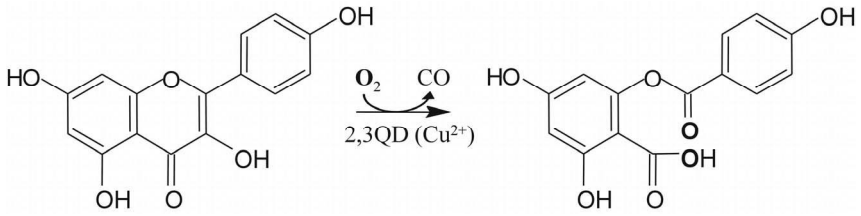
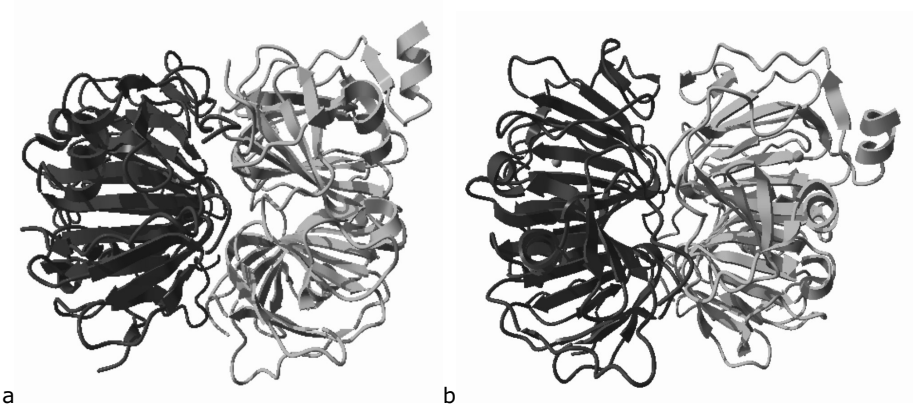


Figure 6.2. Scheme of the dioxygenation proces of kaempferol mediated by 2,3QD. The site of the oxygen insertion is shown in bold.

2,3QD from A. japonicus is a homo-dimer (see Figure 6.3) (Fusetti *et al.*, 2002). The 350 amino acid residuemonomer has a two-domain structure with a pseudo 2-fold rotational symmetry. On account of the extensive homology between the two domains, it has been surmised that the monomer is the result of a gene duplication. The N-terminal domain (residues 1–145) is connected by a 60 amino acid residue linker with the C-terminal domain (residues 206–350). The calculated molecular mass (M) of the monomer amounts to 37.9 kDa, but the actual value is much larger (about 50 kDa) because of extensive glycosylation. The crystal structure was determined of the enzyme from which the majority of the sugars had been removed enzymatically (Fusetti *et al.*, 2002).

The active site (Figure 6.4(a)) is located in the N-terminal domain and consists of a single Cu(II) ion immobilised in the structure by three histidine residues that bind the Cu ion through their Nε atoms. The Cu-site in the C-terminal domain was lost, apparently, in the course of evolution. In the crystal, the Cu atom exhibits two coordination geometries, depending on the conformation of the side-chain of Glu73. In one geometry (30% occupancy) the Glu coordinates to the Cu atom (Cu–O3 distance 0.22 nm) while a water molecule (W2 in Figure 6.4(a)) is found at a distance of 0.24 nm from the Cu atom, resulting in a trigonal bipyramidal coordination; in the other geometry (70% occupancy) the Glu side-chain is turned away from the Cu atom, and there is a single water molecule (W1 in Figure 6.4(a)) at a distance of 0.21 nm from the Cu atom, leading, together with the three histidine ligands, to a distorted tetrahedral ligand configuration.

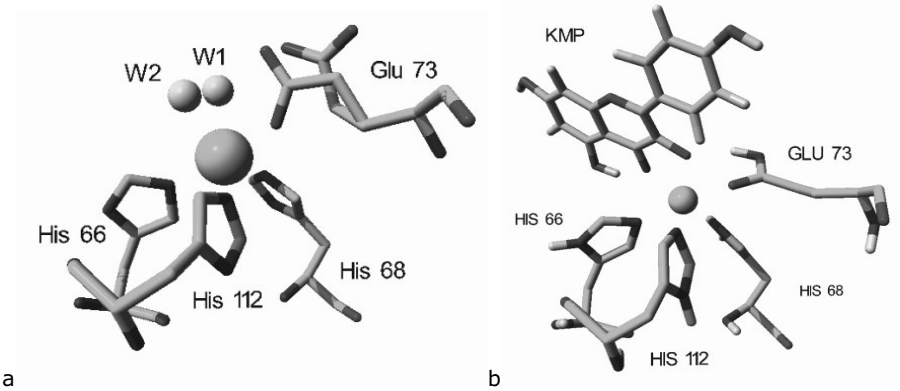


a b

Figure 6.3 (a) Ribbon display of the holo protein; (b) side view obtained by rotating the molecule over 45° around the horizontal, in-plane axis. The monomers are shown in black and grey. The copper atoms are shown as spheres.

The crystal structure of 2,3QD with the substrate KMP in place shows that KMP binds with its 3-hydroxyl oxygen atom to the Cu atom (Figure 6.4(b)) (Steiner *et al.*, 2002). Itis thought that Glu73 may play a catalytic role, in that it abstracts a proton from the 3-OH group. Binding of the KMP into the substrate cavity also leads to a movement of ring B out of the plane of rings A and C, leading to an sp3-like hybridisation and an induction of radical character at position C2. This makes the C2 position a suitable point of attack for the incoming oxygen atom. After attaching itself to the C2 atom, the oxygen atom may form a dioxetane ring by bridging to C4 followed by elimination of CO and oxygen insertion into the substrate (Steiner *et al.*, 2002).

An only partly resolved feature of the native 2,3QD structure is the conformation of the linker (residues 146–205) connecting the N and C-terminal domains, in particular the stretch from Thr146 to Asp176, which is located close to the substrate cavity entrance (see Figure 6.5(a) and (b)). In the native structure, the stretch from 155–169 is not defined because of enhanced mobility, although for one of the four molecules in the asymmetric unit the α -helical structure of residues 164–176 is visible (Figure 6.5(b)). In the enzyme–substrate complex (E.S) there is increased order in this region and only residues 155–158 remain invisible in the crystal (Figure 6.5(a)). The purported α-helical region (164–176) breaks up into two α -helices with slightly different orientations, only



a b

Figure 6.4. Cu-site of 2,3QD. (a) The Cu-site exhibits two possible co-ordination geometries in absence of substrate. In one geometry (30% occupancy) the Glu co-ordinates to the Cu (Cu-Oε distance 0.22 nm) with a water molecule at position W2; in the other geometry (70% occupancy) the Glu side chain is turned away from the Cu and the water molecule is at W1. (b) Cu-site in the enzyme-substrate complex of 2,3QD with KMP. The KMP is deprotonated at the 3OH position and Glu73 carries a proton at Oε2, while Oε1 is co-ordinated to the Cu-ion.

one of which is seen in the Figure. Moreover, there are large displacements (up to 0.8 nm) between the native and the E.S structure in this region (residues 164–176). It has been stated that the loop (residues 154–169) may play a role in the catalytic mechanism (Steiner *et al.*, 2002), i.e. the loop may close off the substrate cavity after substrate binding. In order to shed more light on this, a molecular dynamics (MD) simulation study has been initiated of the enzyme and of the E.S in aqueous solution with particular attention to the behaviour of the linker connecting the two domains of the monomer. It is found that this linker has enhanced mobility in the free enzyme, and that binding of substrate reduces this mobility considerably and leads to large movements of some linker segments. These findings are compatible with the XRD data. To determine which free energy contributions are responsible for the behaviour of the loop is beyond the scope of present-day MD-simulation, but it appears that enthalpic contributions stemming from loop/protein, loop/solvent and loop/KMP interactions favour the closure of the cavity once the substrate is in place. Thus, while the results of the MD-simulations are in agreement with the static picture provided by the XRD data, they extend these by giving dynamic information about the structure.

6.2 Methods

All simulations were performed using the GROMACS simulation package (van der Spoel *et al.*, 1999b) together with the GROMOS96 43A2 force-field (van Gunsteren *et al.*, 1996). The bonding parameters for kaempferol, Figure 6.5, were taken from the standard GROMOS96 force-field, see Tables 6.1-6.4 (Oostenbrink *et al.*, 2000; van Gunsteren *et al.*, 1996).

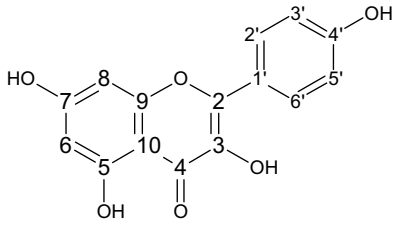


Figure 6.5. Schematic presentation of kaempferol.

Table 6.1: Constraints

Constrained bond distances (atom names)	Distance (nm)
O*-H*	0.100
C*-H*	0.109
C4-O4	0.123
C*-O* (* is not 4)	0.136
C*-C*	0.139

Table 6.2: Bond-angle bending parameters:

$$V_{angle}(\theta_{ijk}) = \frac{1}{2} k_\theta \left[\cos(\theta_{ijk}) - \cos(\theta_{ijk,0})\right]^2$$

Bond angle (atom names)	Ideal angle θ_0 (degree)	Force constant k_θ (kJ.mol^{-1})
C*-O*-H*	109.5	450
C*-C*-H*	120.0	505
C*-C*-C*	120.0	560
C*-C*-O*	120.0	560
C*-O*-C*	120.0	560

* any symbol. The O3 atom is deprotonated

Table 6.3: Harmonic dihedral angle parameters:

$V_{harm}(\xi_{ijkl}) = \frac{1}{2} k_{\xi} (\xi_{ijkl} - \xi_{ijkl,0})^2$

Harmonic dihedral angle (atom names)	ξ_0 (degree)	k_ξ (kJ.mol^{-1}.degree^{-2})
all 18 dihedral angles in the three rings: C6-C5-C10-C9 etc.	0	0.051
all 15 improper dihedral angles at sp2 carbon atoms in the rings: C6-C5-H6-C7, etc.	0	0.051

Table 6.4: Trigonometric dihedral angle parameters:

$V_{trig}(\varphi_{ijkl}) = k_{\varphi}(1 + \cos(\delta)\cos(m\varphi_{ijkl}))$

Trigonometric dihedral angle (atom names)	cos(δ)	m	k_φ (kJ.mol^{-1})
C6-C5-O5-H5	-1	2	7.11
C6-C7-O7-H7	-1	2	7.11
C3-C2-C1′-C6′	-1	2	7.11
C5′-C4′-O4′-H4′	-1	2	7.11

The charge densities for the Cu site in the enzyme were obtained from density functional theory calculations (Koch and Holthausen, 2001) with the ADF program (Baerends *et al.*, 2000; te Velde *et al.*, 2001). The atoms that were taken into account for the DFT study besides the Cu atom are the side-chains of the Cu-coordinating residues (His66, His68, Glu73 and His112) down to the Cα atom and, in addition, the kaempferol molecule for the E.S or a water molecule for the free enzyme. The coordinates of the active site were taken from the X-ray structures (entries 1H1M and 1JUH of the Protein Data Bank for the E.S and the free enzyme, respectively (Berman *et al.*, 2000)). The co-ordination chosen for the Cu site in the free enzyme is the five-coordinate trigonal bipyramidal configuration. Hydrogen atoms were added and were optimised using the ADF routine. The Becke–Perdew (Becke, 1988; Perdew, 1986a; Perdew, 1986b) exchange-correlation functional was used in a triple zeta valence plus polarization basis set of Slater-type orbitals (te Velde *et al.*, 2001). The recently developed multipole derived charge (MDC-q) analysis was used to convert the charge densities into atomic charges (Swart *et al.*, 2001). The charges of the aliphatic hydrogen atoms as obtained

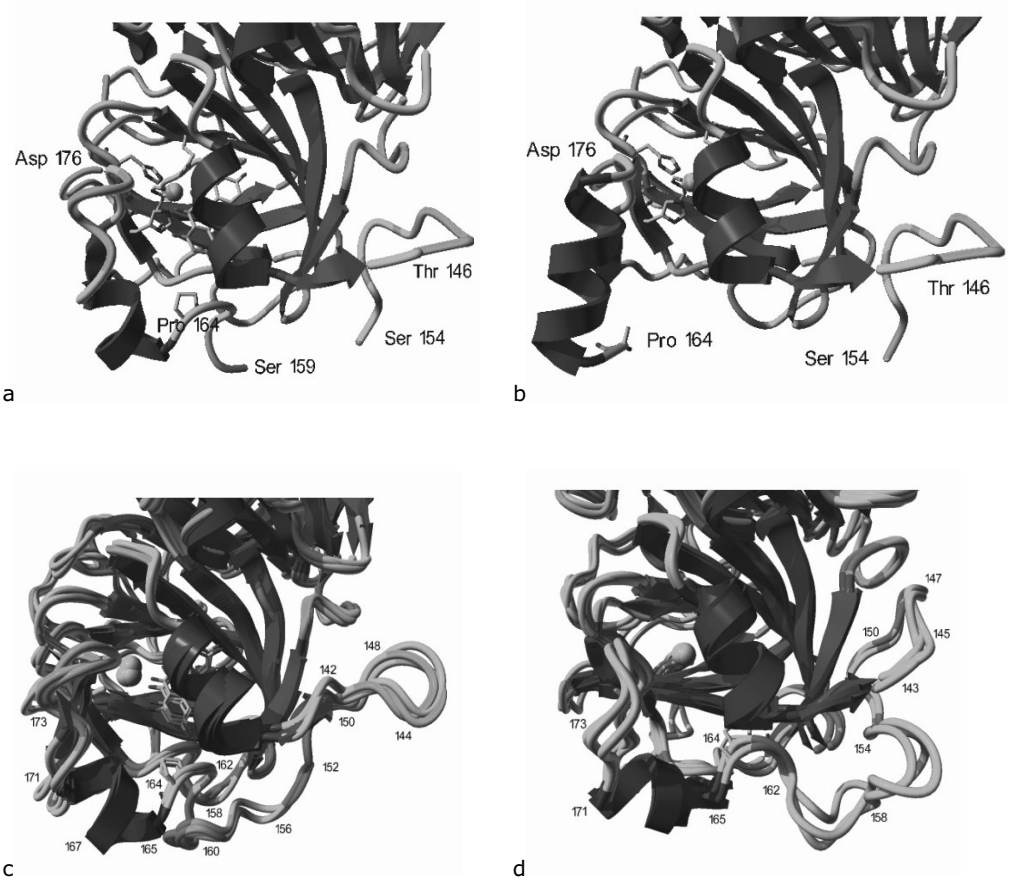


Figure 6.6. (a) and (b) View of the crystallographically determined active site (a) in the presence (molecule B of 1H1M) and (b) in the absence (molecule B of 1JUH) of the substrate KMP.

(c) and (d) Three superimposed snapshots (6.2, 6.7 and 7.2 ns) of the active site from the MD-simulations (c) in the presence and (d) in the absence of KMP. The Cu co-ordinating residues: His 66, 68, 112 and Glu 73 as well as Pro 164 and the KMP molecule are shown as sticks. All protein models are generated using the programs Yasara (http://www.yasara.org) and Pov-Ray (http://www.pov-ray.org).

from DFT were added to the corresponding carbon atoms. The resulting atomic charges are presented in Tables 6.5 and 6.6. Since the simple point charge (SPC) water model was used throughout the simulations for all water molecules, the charges of the water molecule as presented in Table 6.6 were replaced by those of the SPC water (0.41e on the hydrogen atoms, -0.82e on the oxygen atom) (Berendsen *et al.*, 1981). Consequently, the Cu centre carried a non-integer charge of +1.89e in the free enzyme. The force-field for the Cu atom was simplified by constraining the distances from Cu to the three histidine ligand atoms to their X-ray values: 0.208 nm, 0.209 nm and 0.205 nm for residues 66, 68 and 112, respectively. In the presence of kaempferol, two additional distances: Cu-$O3_{KMP}$ and Cu- $O\varepsilon1_{Glu73}$ were constrained at 0.198 nm and 0.205 nm, respectively. Forces between Cu and all other atoms were modelled through non-bonded interactions. For the free enzyme, this resulted in distances of the Cu to the $O\varepsilon1_{Glu73}$ and to the water molecule of 0.25 nm and 0.27 nm, respectively. Although these distances are larger than the crystallographic distances, mainly because of the fairly large Lennard–Jones C12 parameter used by GROMOS to model the Cu–O interaction, no

Table 6.5: Charge distribution (*e*) of the active site of 2,3 QD in presence of substrate KMP.

	His 66	His 68	His 112		Glu 73		KMP		
Cα	-0.009	0.023	0.014	Cα	0.034	O1	-0.408	C9	-0.037
Cβ	0.160	0.154	0.159	Cβ	-0.052	C2	0.203	C10	0.224
Cγ	0.102	0.076	0.085	Cγ	0.196	C3	0.224	C1′	-0.060
Cδ	0.077	0.031	0.072	Cδ	0.429	O3	-0.585	C2′	-0.163
Nδ	-0.598	-0.626	-0.598	Oε1	-0.465	C4	0.288	H2′	0.244
Hδ	0.454	0.459	0.460	Oε2	-0.481	O4	-0.435	C3′	-0.222
Cε	0.374	0.492	0.397	Hε2	0.440	C5	0.093	H3′	0.184
Nε	-0.363	-0.400	-0.357			O5	-0.417	C4′	0.060
						H5	0.436	O4′	-0.442
						C6	-0.302	H4′	0.462
						H6	0.258	C5′	-0.244
						C7	0.052	H5′	0.262
						O7	-0.418	C6′	-0.108
						H7	0.437	H6′	0.181
						C8	-0.316		
						H8	0.259	Cu	0.531
total:	0.197	0.209	0.232		0.101				0.241

further refinement of the Cu force-field was done, since these distances were stable during the simulations. The total charge on the free enzyme amounted to -15.12e. The shielding effect of the water and the use of a cut-off radius that entails the presence of a continuum background charge, make the addition of counter-ions in the simulation not necessary. There are no indications that the charge on the protein or the details of how the Cu-site was modelled, had an effect on the time-dependent behaviour of the loop (residues 154–169). For the MD-simulations, the starting coordinates for the E.S were taken from the X-ray structure (entry 1H1M of the Protein Data Bank (Berman *et al.*, 2000)). The coordinates for the amino acid residues that had no density in the electron density map (1–2 and 154–158) were generated by using the Swiss PDB Viewer (Guex and Peitsch, 1997). Polar and aromatic hydrogen atoms were added to the protein, which resulted in a total of 3488 atoms. Protonation states of the residues were chosen so as to correspond to a pH of 7. The histidine residues that coordinate the Cu atom with their Nε atom, were protonated at the Nδ position, as were the histidine residues His13 and His71, while His21, His148 and His201 were protonated at the Nε position and the Glu73 residue was protonated at the Oε2 atom. Kaempferol was deprotonated at O3.

The configuration of the loop region (residues 154–169) was optimised at the start of the MD-simulation by energy minimisation (EM) for 500 steps while freezing the rest of the protein. Another 500 steps of EM of the loop region were performed, constraining the bond lengths using the LINCS algorithm (Hess *et al.*, 1997). The protein was then centered in a rectangular periodic box with dimensions such that the distance between t

Table 6.6: Charge distribution (in *e*) of the active site of native 2,3QD.

	His 66	His 68	His112		Glu73		H_2O	Cu
Cα	0.057	0.091	0.086	Cα	0.121	O	-0.373	
Cβ	0.182	0.092	0.123	Cβ	-0.086	H	0.183	
Cγ	0.071	0.147	0.096	Cγ	0.208	H	0.307	
Cδ	0.109	0.099	0.122	Cδ	0.466			
Nδ	-0.565	-0.576	-0.570	Oε1	-0.429			
Hδ	0.478	0.486	0.489	Oε2	-0.551			
Cε	0.367	0.404	0.416	Hε2	0.493			
Nε	-0.359	-0.408	-0.498					0.725
total:	0.340	0.335	0.264		0.222		0.117	0.725

he protein and the walls of the box was more than 0.6 nm in every direction. The volume of the box was 6.41x6.81x7.18 nm^3. The box was filled with 8703 SPC water molecules (Berendsen *et al.*, 1981). The water configuration was relaxed by 50 steps of EM while freezing the protein. Then, all degrees of freedom of the system were relaxed with another 50 EM steps.

The initial atomic velocities were taken from a Maxwell–Boltzmann distribution at 50 K. The temperature was then increased to 400 K over 350 ps and reduced to 300 K over 100 ps. Further simulations were performed at constant temperature (T=300 K) and pressure (p=101,325 Pa) by weakly coupling the protein and the solvent separately to an external bath (t_T=0.1 ps and t_p=1.0 ps) (Berendsen *et al.*, 1984). A time-step of 2 fs was used. A twin-range cut-off method was used for non-bonded interactions. Lennard–Jones and Coulomb interactions within 0.8 nm were re-calculated every time-step, whereas non-bonding interactions between 0.8 and 1.4 nm were updated every five steps. After 1 ns of equilibration, the next 7.2 ns were used for analysis. Every 0.5 ps, the coordinates of the protein were saved for analysis.

The configuration at 7.2 ns of the simulation with substrate was used as starting configuration for the simulation of the native protein without substrate. The kaempferol molecule was removed, the charge of the Cu-site was changed (see Table 6.6) and the constraint between Cu and Glu73 removed. After an equilibration period of 1 ns, the next 7.2 ns were used for analysis. As analysis of the loop dynamics revealed motions on the timescale of a few nanoseconds, both simulations were extended by another 8 ns to check the validity of the conclusions. The present simulations were based on the crystal structure coordinates of the E.S. After (1+7.2) ns of simulation the E.S was converted to the free enzyme by taking out the KMP, after which the free enzyme was simulated. It would have been interesting to see if, conversely, a simulation starting from the crystal structure of the free enzyme into which at a later point in time the KMP would have been inserted, would have yielded the same results as presented here for the complex. However, the structures of the E.S and the free enzyme differ mainly in that the loop is less well defined in the free enzyme. An MD-simulation of the latter would therefore have to be preceded by extensive modeling of the loop for which the E.S coordinates would be the most logical starting point. This would effectively duplicate the MD-simulations reported above.

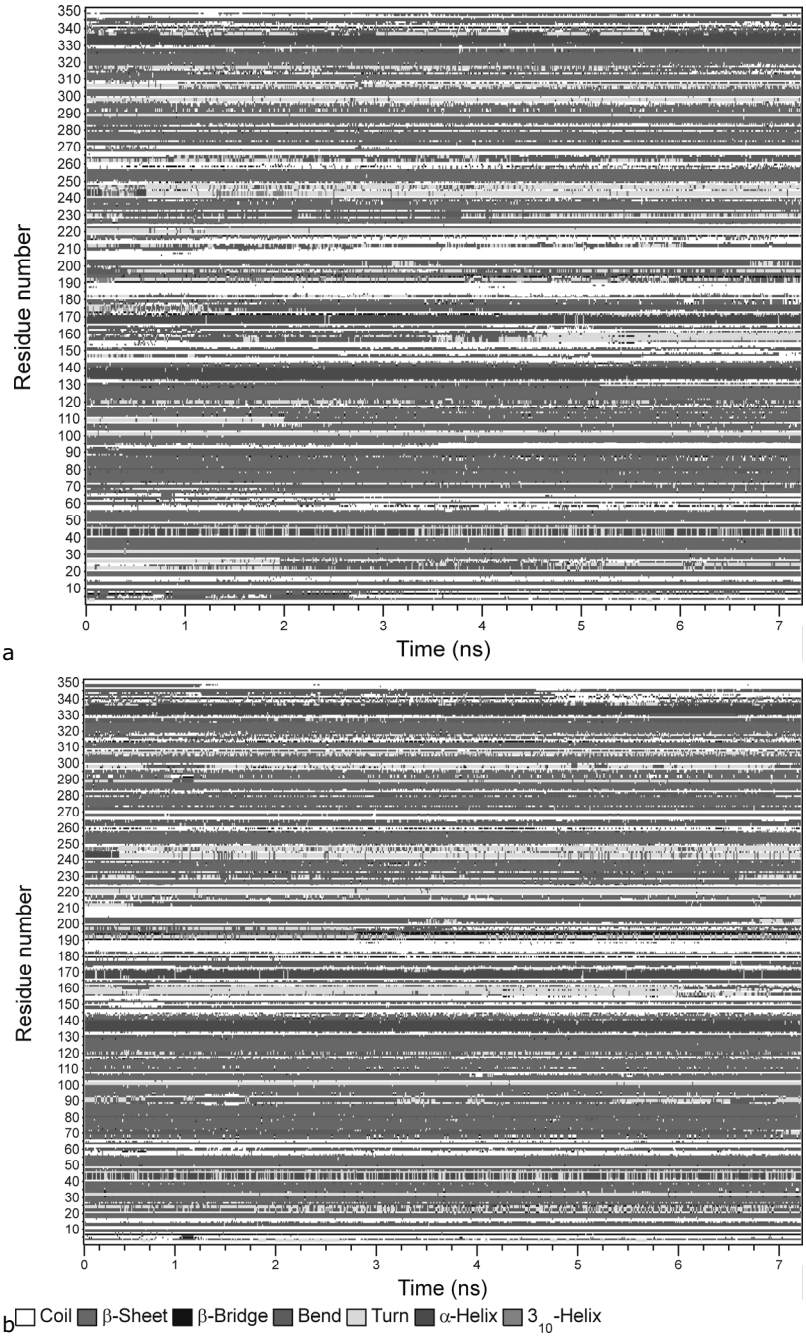


Figure 6.7. Secondary structure of 2,3 QD as obtained from MD-simulation in presence (a) and absence (b) of KMP.

6.3 Results and discussion

Overlays of a number of snapshots taken from the last ns of both simulations are shown in Figure 6.6(c) and (d). The analysis of both simulations is presented below. Since the characteristic time of the loop motion appeared to be of the order of a few nanoseconds, the simulations of the E.S and of the free enzyme were extended by another 8 ns. Most of the analysis will focus on the first 7.2 ns of simulation, but the whole 15.2 ns trajectories were considered for the root-mean-square fluctuation (RMSF) and root-mean-square difference (RMSD) analysis (vide infra).

6.3.1 Protein stability and fluctuation

Figure 6.7 shows the secondary structure (Kabsch and Sander, 1983) for the monomer in the presence (Figure 6.7(a)) and in the absence (Figure 6.7(b)) of KMP as a function of time. The secondary structure is remarkably constant both for the E.S and the free enzyme. Figure 6.8 shows the RMSF values of the backbone atoms (Cα, N and CO) averaged over the 4.0–7.2 ns period of the simulations as a function of residue number for the free enzyme (-KMP) and the E.S (+KMP). The crystallographic B-factors converted to RMSF values are shown as well. The regions where the B-factor curve is interrupted correspond with illdefined regions in the crystallographic electron density map. Although a one-to-one correspondence between crystallographic B-factors and simulated atomic positional RMSF values should not be expected, due to the difference between their definitions (Hunenberger *et al.*, 1995), the simulation results parallel the crystallographic data. In some regions, the RMSF is larger than the value derived from the crystallographic B-factor. They correspond with areas of contact between monomers in the crystal concluding that the side-chain and backbone motions in these areas are dampened in the crystal. This is the case for residues 145–150, which are in the proximity of residues 345–350 of the opposite monomer, and residues 43–47, which are near residues 172–175. Other residues with a high simulated RMSF value occur in loops connecting two β-strands (for instance, residues 69–72, 88–94, 228–232, 241–247 and 286–292).

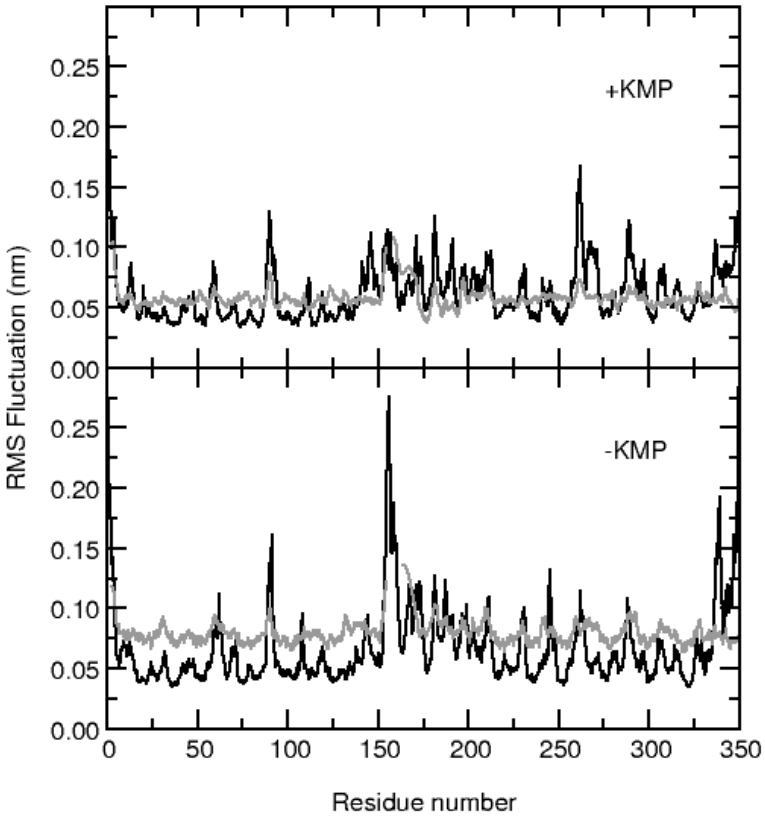


Figure 6.8. Root Mean Square Fluctuation (RMSF) of the protein backbone in presence (top) and absence (bottom) of KMP as obtained from the period of 4.0-7.2 ns of the simulations (black lines). RMSF values based on crystallographic B-factors (grey lines) were obtained by using the conversion formula: RMSF $=\sqrt{(3B/8\pi^2)}$.

6.3.2 Loop stability and fluctuation

Since the behaviour of the loop (residues 154–169), that appears to lock the substrate in place, is particularly interesting, a more detailled view of the dynamic behaviour of residues 150–180 is obatined during the 15.2 ns simulations. The RMSF values of the backbone atoms averaged over intervals of 3 ns for a series of consecutive intervals are shown in Figure 6.9, including the RMSF values calculated from the crystallographic B-factors. The MD data mirror the gross variation in B-factor as obtained from the XRD

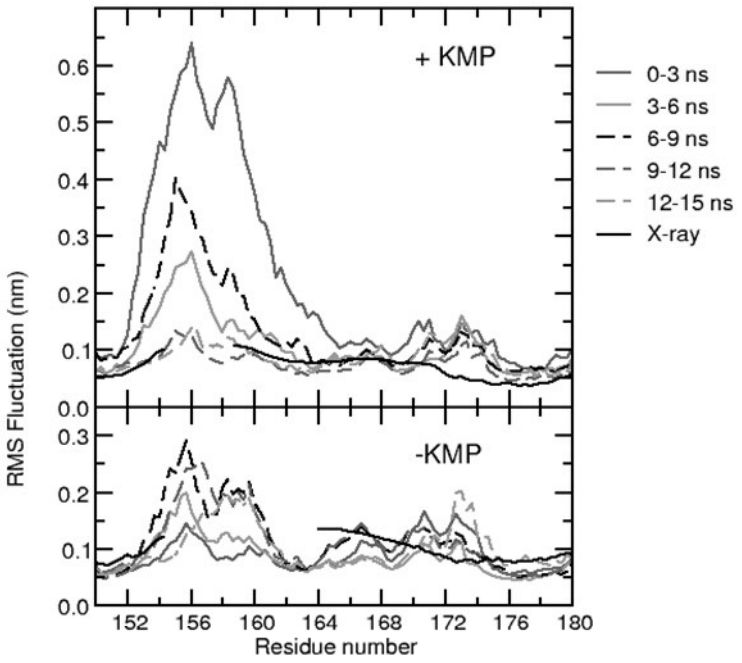


Figure 6.9. Root Mean Square Fluctuation of residues 150-180 in the presence (top) and in the absence (bottom) of KMP averaged over consecutive 3 ns periods of the simulations.

analysis. The striking feature in Figures 6.8 and 6.9 is the variation in RMSF in relation to the presence or absence of KMP. The very large RMSF values that are seen at the beginning of the simulation with KMP (Figure 6.9) in the region of residues 154–164 gradually diminish with time. The structure of this region needs several nanoseconds to reach equilibrium. Residues 155–158 maintain a high RMSF compared to the rest of the loop (159–169) after 6–9 ns MD-simulation. These are exactly the residues that remain invisible in the XRD electron density map of the E.S. Also, the region of residues 172–176 remains fairly mobile, but here the mobility is higher than expected on the basis of the crystallographic B-factors. As pointed out above, the B-factors may be lowered due to monomer–monomer contacts in the crystal.

When KMP is taken out of the E.S, a conspicuous increase in RMSF is observed in the region of residues 154–176 (Figure 6.9), so the loop becomes more mobile, with

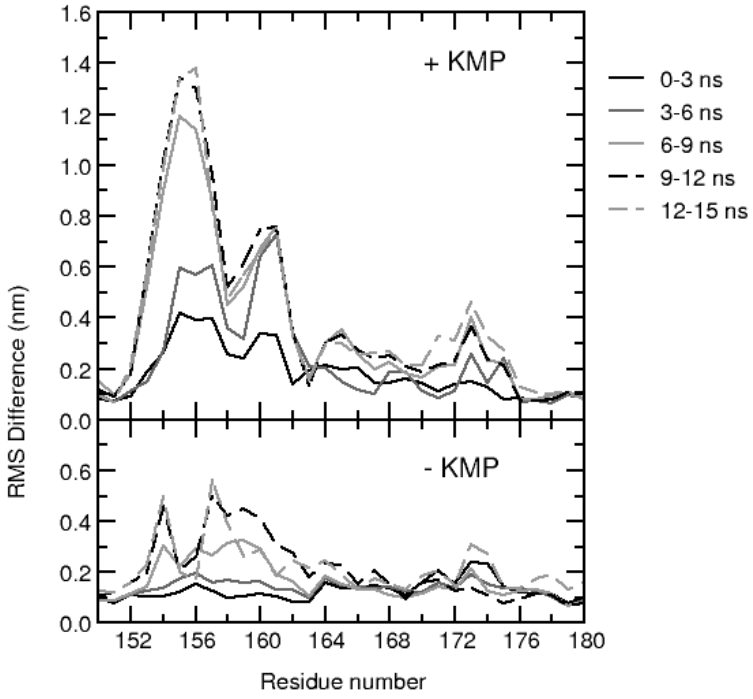


Figure 6.10. Root Mean Square Displacement of the Cα atoms of residues 150-180 averaged over consecutive 3 ns intervals in the presence (top) and in the absence (bottom) of KMP with respect to the structure at time zero.

particularly large variations in the regions 154–162 and 165–176. This parallels the variation in the crystallographic B-factors. In the simulation, the breaking up of the α-helix (164–176) into two slightly differently oriented helices also seems recognizable in Figure 6.9, where residue 172 shows less motion. It is interesting to note that the RMSF of Pro164 is very small. Pro164 has been ascribed a mechanistic role in that it positions the substrate in the substrate cavity by van der Waals contacts.

Apart from the mobility of the loop, the change in its average position as a function of substrate presence is of interest. This is illustrated in Figure 6.10, where the RMSD of the Cα atoms from the starting structures of the respective 15.2 ns simulations has been averaged over intervals of 3 ns and plotted as a function of residue number for five consecutive intervals. It can be seen (Figure 6.10, top panel) that this part of the protein

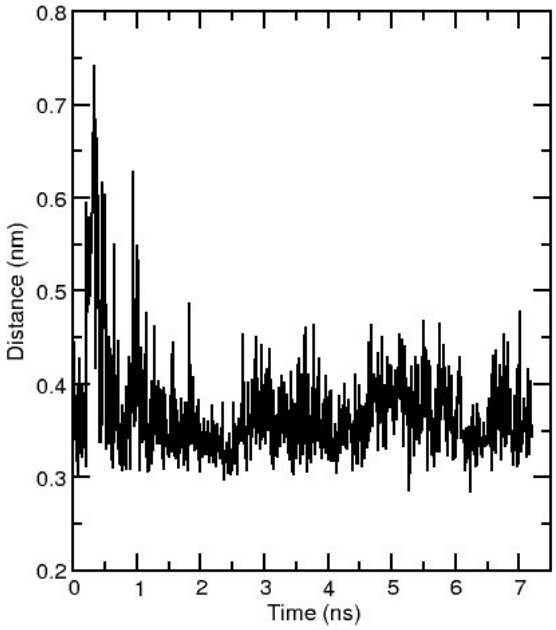


Figure 6.11. The distance between the Cδ atom of Pro164 and the O7 atom of KMP in the E.S complex as a function of time.

in the E.S is stable for the duration of the simulation. It is clear that the structure needs at least 3–6 ns to equilibrate. Again, the mobile character of part of the loop (residues 154–160) is visible, although it should be kept in mind that the starting structure of the loop was obtained by simple energy minimisation. Therefore, the initial motion of the loop may reflect, in part, the fact that the minimised initial structure was not in equilibrium. When the substrate is taken out, this part of the loop again makes a large movement, while the α-helical region (residues 164–176) undergoes relatively large displacements (up to 0.3 nm). This is the region where large structural differences (0.6–0.7 nm) are seen in the crystal structure, indeed, between the E.S and the free enzyme.

6.3.3 Proline 164

So far, it has transpired that Pro164 shows relatively little positional fluctuation (Figure 6.9) and that, of all the residues in the loop, it shows one of the smallest displacements upon removal of substrate (Figure 6.10). To analyse the dynamic behaviour of this

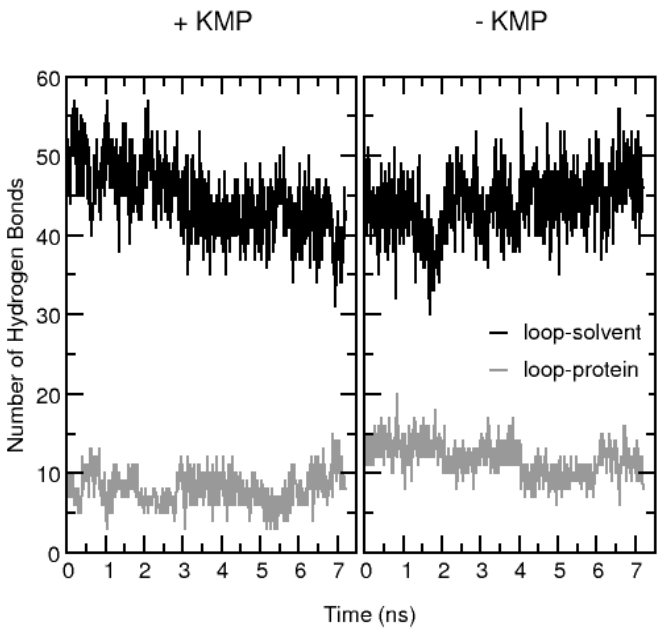


Figure 6.12. Number of hydrogen bonds during the MD-simulations between the loop (residues 154-169) and the remaining protein (grey) and between the loop and the solvent (black) in presence (left) and in absence (right) of KMP.

residue further the distance between the Cδ atom of this residue and the O7 atom of KMP as a function of time was verified (Figure 6.11). Apart from some large variations in the beginning of the run, this distance is stable at about 0.33–0.37 nm, indicating that Pro164 and KMP are continuously in van der Waals contact. When looking at the hydrogen bonds made by the loop (residues 154–169) (Figure 6.12), it is clear thatin the simulation of the E.S the number of hydrogen bonds with the solvent water molecules decreases over time, while the number of hydrogen bonds of the loop to the rest of the protein increases slightly. This is in agreement with the ordering of the loop. The reverse occurs when the KMP is removed, which implies that the loop becomes more solvent-exposed.

6.3.4 Driving force

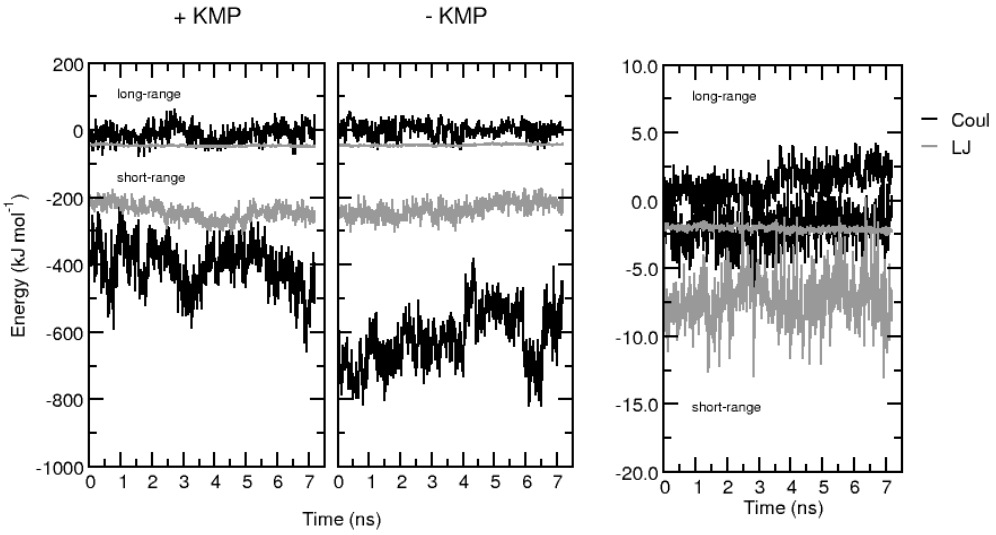


Figure 6.13. (a). Short Range (SR) and Long Range (LR) non-bonded interaction (Coulomb and Lennard-Jones) between the loop and the remaining protein in presence (left panel) and in absence (right panel) of KMP as a function of time. (b) non-bonded energy between the loop and KMP as a function of time.

In an attempt to trace the driving force behind the motion of the loop upon binding of substrate, the non-bonding energy of interaction of the loop with the rest of the protein has been plotted as a function of time in Figure 6.13(a) (in the presence of KMP, left panel; in the absence of KMP, right panel), while the interaction energy of the loop with KMP is plotted in Figure 6.13(b). It appears that immobilisation of the loop in the E.S is favoured by the loop–protein interactions, which is in agreement with the increase of hydrogen bonds between loop and protein, and disfavoured to a lesser extent by the loop–KMP interactions. The movement of the loop away from its position in the E.S, when the KMP is taken out, is not favoured by the loop–protein interactions, but solvation and hydrophobic effects seem to play an important role in driving this conformational change.

6.3.5 Presence and behaviour of water molecules inside the cavity.

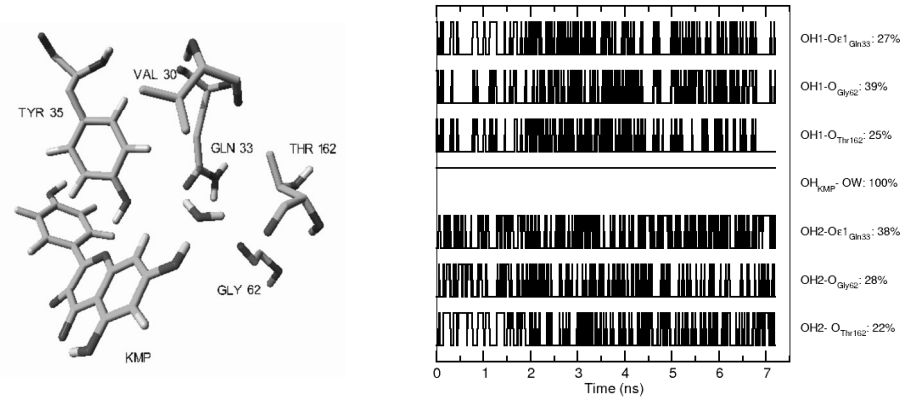


Figure 6.14. (a). View of structure around water molecule SOL5890. (b). The presence of hydrogen bonds from and to water molecule SOL5890 (oxygen OW, hydrogen OH1 and OH2) as a function of time.

Finally, an analysis of the behaviour of the water molecules inside the substrate cavity was done. In the crystal structure of the E.S, two water molecules are visible inside the substrate cavity that connect the hydroxy groups at the C7 and C40 positions of the KMP through hydrogen bridges to the protein framework. For the simulation of the E.S, water molecules were inserted in the initial configuration by the program GROMACS. During the simulation, two water molecules interact with the KMP, i.e. one (SOL3663) close to the Cu atom (0.27 nm) where it formed hydrogen bridges with $O\varepsilon1_{Glu73}$ and $O3_{KMP}$ for 70% and 80% of the time, respectively, and one (SOL5890; see Figure 6.14(a)) that forms hydrogen bonds with the O7 H7 of the KMP during the whole simulation (see Figure 6.14(b)) and with residue Gly62 for approximately 50% of the time. It was found from the MD analysis that the hydrogen atoms of the two solvent molecules alternate position continuously during the simulation. The protons interchange on a picosecond timescale.

When the KMP was taken out of the E.S, the cavity filled with water molecules. This can be seen in Figure 6.15, where the radial distribution function of the water atoms (O and H atoms) is presented as a function of the distance from the Cu atom together with the corresponding cumulative number of water molecules. As mentioned in the previous paragraph, for the E.S, one water molecule is seen to be present at a distance of 0.27 nm from the Cu atom, while the second water molecule (the one hydrogen bonded to the O7

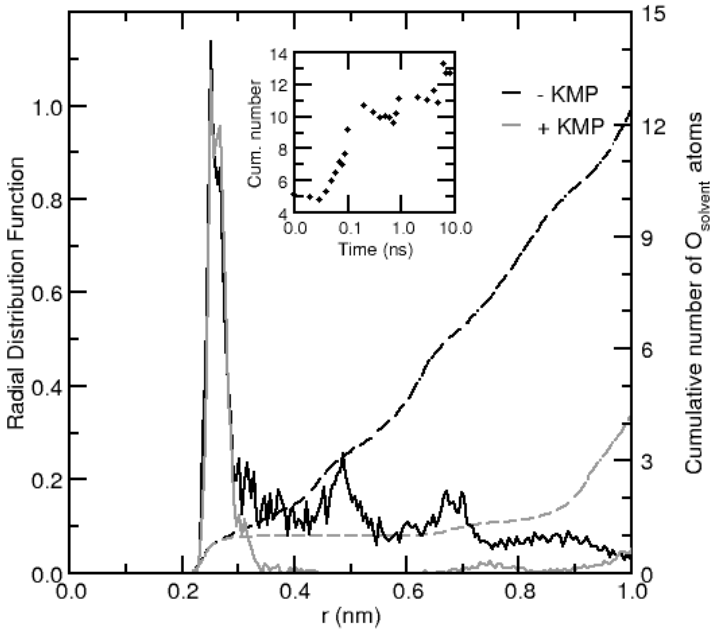


Figure 6.15. Radial distribution of water oxygen atoms around the Cu-atom. The broken lines represent the cumulative number of water molecules within a certain distance from the Cu atom in presence (grey) and absence (black) of KMP as a function of distance. The inset shows the amount of water molecules present within 1 nm of the Cu-atom as a function of time after removal of KMP including the equilibration period of 1 ns.

atom of KMP) is seen at a distance between 0.8 and 1.0 nm from the Cu atom. When the KMP is taken out, the radial distribution function is seen to change rapidly, with water entering the outer parts of the cavity first and subsequently moving to the Cu atom within 0.2 ns. At the end of the simulation the cavity (i.e. within a radius of 1 nm from the Cu atom) is filled with 13 water molecules.

6.4 Conclusion

The results presented above show that the structures as obtained from the crystallographic studies are stable in MD-simulations. This holds for the free enzyme as well as for the E.S. Secondly, in the MD-simulation of the free enzyme, the linker

connecting the two domains of the enzyme monomer shows strongly enhanced mobility in the loop region that is close to the entrance to the substrate cavity (residues 154–169), which is in agreement with data on the crystallographic B-factors. It is found that in the simulations the movement of the loop takes place on the 5–10 ns timescale, which makes the phenomenon, in principle, accessible for systematic analysis with present-day MD-simulation techniques. When substrate binds into the cavity, the loop orders remarkably. Still, mobility is seen for residues 155–158, which is in agreement with the XRD data. Some regions of the loop (residues 154–160 and 164–176) move over a considerable distance and approach the substrate closely, reinforcing the idea that they lock the substrate in the substrate cavity (Steiner *et al.*, 2002). Upon loop closure and ordering of Pro164 with respect to the substrate, more hydrogen bonds are formed between the loop and the protein, while the number of hydrogen bonds between the loop and the solvent decreases. The non-bonded energy terms confirm this, the short-range (SR) Coulomb interaction between loop and protein becomes more negative.

To pinpoint the driving force for the loop motion is beyond the power of present-day MD-simulations, but it appears that at least the enthalpic component of the interaction of the loop with the protein and the KMP favours the locking of the substrate. Further analysis is needed to elucidate the possible role of (de)solvation of substrate, loop and cavity.

Finally, the simulations allowed for an analysis of the behaviour of solvent molecules inside the substrate cavity. Two water molecules were found immobilised in the cavity, one of which exhibited rotation at the picosecond timescale. Removal of the substrate led to a filling of the cavity within 0.2 ns. In conclusion, this study shows that MD-simulation can complement X-ray crystallographic work regarding enzyme–substrate binding by providing a dynamical picture of protein, substrate and solvent at the atomic level, which is inaccessible to experiment.