Chapter 4

An extended MD simulation of wild-type haloalkane dehalogenase

4.1 Introduction

For wild-type dehalogenase, the rate-limiting step in the overall kinetics of dehalogenation of 1,2-dichloroethane is the release of the halide ion [14], a product of formation of the intermediate ester. To find out whether the presence of a halide ion, in the active site cavity, has any influence on the dynamics of the protein, two MD simulations were performed. In one of these simulations a chloride ion was bound in the active site, and this calculation will be the subject of the next chapter. This chapter deals with the simulation of dehalogenase in the absence of halide. We will have a closer look at the stability of the molecule and investigate the types of motion that occurred during the simulation.
4.2 Methods

Simulations were performed using the GROMOS force field [15] with adjusted Lennard-Jones interactions between aliphatic carbon atoms and water oxygen atoms [42] and explicit hydrogens added to aromatic residues [43]. Polar hydrogens were included explicitly whereas nonpolar hydrogens were implicitly included by the use of united atoms. As a starting structure, the X-ray structure at pH 8.2 and a resolution of 1.9 Å (Brookhaven data bank entry 1EDE) [3, 7, 44] was chosen. All histidines were assumed to be electrically neutral. To find out which ring nitrogen was protonated, the X-ray structure was checked for the presence of possible hydrogen bonds by looking at the closest hydrogen bond acceptor. Histidine residues 37, 54 and 289 were protonated at N$_{\text{E1}}$ positions, residues 102 and 305 at the N$_{\text{E2}}$ positions. Periodic boundary conditions were applied by the use a truncated octahedron filled with equilibrated water molecules. Each water molecule of which the oxygen atom had a distance to any non-hydrogen protein atom (or water molecule present in the X-ray structure) of less then .23 nm was removed. The minimum distance between protein atoms and the walls of the box was taken as 0.75 nm. The volume of the box was 176 nm$^3$. The protein possessed a negative charge of -17 e. To obtain an electrically neutral system, 17 sodium ions were added as counterions. This was done by calculating the electric potential at the positions of all oxygen atoms of water molecules and replacing the 17 water molecules with the lowest potential by a sodium ion. In total the system contained 16287 atoms of which 3169 were part of the protein. Subsequently the energy of the system was minimized by the steepest descent method for 100 steps without position restraining.

During the first 10 ps of simulation, harmonic position restraining was applied to all protein atoms, using a force constant of 9000 kJ nm$^{-1}$ mol$^{-1}$. After that, the position restraints were removed and the system was simulated for 1 ns. Initial velocities were taken from a Maxwell-Boltzmann distribution at 298 K. The temperature was kept constant by coupling solute and solvent separately to a thermal bath [37] at 298 K.
with a coupling constant $\tau_T=0.1$ ps. Pressure was kept constant by coupling to a pressure bath at 1.0 bar [37] using a coupling constant $\tau_P=0.5$ ps. Bond lengths were constrained using the SHAKE method [16] with a relative tolerance of 0.0001. Nonbonded interactions were calculated using a twin cut-off radius: within a short cut-off radius of 0.8 nm interactions were calculated every timestep, all other interactions within a radius of 1.2 nm were only calculated every 10 steps. The atom pair list for the short range interactions was updated every 10 steps. The stepsize was 0.002 ps.

Simulations were performed using the GROMOS87 software package [15]. Nonbonded routines were rewritten by the author to be used on a Cray J90 parallel computer. The simulation took ~200 hrs of real time. Analysis of secondary structure elements, rmsd, B-factors and radius of gyration was performed with GROMACS [45].

4.3 Results and discussion

To investigate the stability of the enzyme, three properties were investigated: the stability of secondary structure elements, the rmsd of the protein with respect to the crystal structure and the radius of gyration of the protein. Fig. 4.1 shows the behaviour of the secondary structure elements, as calculated with the program DSSP [46]. For clarity, the elements are subdivided into only three categories: $\alpha$-helices, $\beta$-sheets and coils (including $\beta$-bridges, bends, turns, $\pi$-helices and $3_{10}$-helices). Although there are fluctuations, in general the secondary elements remain intact. One exception is helix 5 (residues 171-181), where the helicity is lost after Trp175. The radius of gyration $R_g$ as a function of time is shown in fig 4.2. If unfolding occurs then this is often revealed by an increase of $R_g$. Initially $R_g$ increases slightly, but stabilizes after 600 ps, resulting in a deviation of less than 2% from the X-ray structure. The root mean square deviation (rmsd), of the backbone $C_{\alpha}$ atoms, from the X-ray structure is shown in fig 4.3. Here we notice that a relatively large
rmsd arises, but again there appears to be a stabilization after 600 ps. To find out which regions are involved in this large rmsd, B-factors can be derived from the trajectory. B-factors are defined as:

$$B = \frac{8}{3} \pi^2 \langle |\Delta r|^2 \rangle$$

(4.1)

where $\langle |\Delta r|^2 \rangle$ is the mean square atomic displacement. Fig 4.4 shows the B-factors from the simulation (for C$_\alpha$ atoms only) compared to the crystallographic ones. For several regions, the fluctuations are one order of
Figure 4.2: Radius of gyration vs. time

Figure 4.3: Root mean square deviation from the crystal structure vs. time
magnitude larger in the simulation compared to the X-ray structure. But peaks in the simulation always correspond to peaks in the X-ray structure. Fig 4.4 also shows that the region where the largest fluctuations occur, the cap domain (residues 156-229), is also rich in crystallographic contacts, especially close to peaks.

The trajectory was further investigated by applying Essential Dynamics (ED) analysis [6] described in section 2.3.1. In our case we constructed a covariance matrix from the 930 Cα coordinates in the time interval 100-1000 ps. The ten largest eigenvalues from this matrix are shown in fig 4.5 (in decreasing order of magnitude). The projections along the first six eigenvectors are shown in fig 4.6. One might look at
the motions produced by projection of the trajectory on separate eigenvectors. It has been suggested however [47] that the motion within the essential subspace is of a diffusive nature. We found further support for this by analysing the trajectory of a multidimensional Brownian particle (appendix C). For this reason it is not obvious that the eigenvector with the largest eigenvalue, which corresponds to the direction in which the largest fluctuation happens to occur, also corresponds to a direction in which the largest fluctuation is possible. So we chose to look at the motion, produced by projection of the trajectory on the hyperplane defined by the first three eigenvectors. The total mean square displacement (msd) of the Cα carbon atoms was 3.78 nm², whereas the sum of the first three eigenvalues was 2.21 nm². This means that in this way 58% of the total msd was retained. Fig. 4.7 shows the B-factors for the Cα atoms, as produced by the first three eigenvectors as well as the B-factors produced by the total motion along all other directions. Especially the motion in the cap-domain (residues 156-229) is mainly produced by the first three eigenvectors. It can also be seen that the residues, involved in catalysis (Asp124, Trp125, Trp175, His289 and Asp260) all have little
motion except for Trp175. But even here the B-factor lies in a minimum with respect to its direct environment. There is a small peak at Asp260, probably caused by the fact that the hydrogen bond between this residue and His289 is lost. Fig. 4.8 shows the motions along the first three eigenvectors as they appear in three dimensional physical space.

After this we performed rigid body analysis (section 2.4) on the trajectory. In principle, this can be done on the full trajectory, as well as on the trajectory obtained by projection on the first three eigenvectors. In fig 4.9 the scores for both, from the time interval 900-1000 ps are shown. There is no significant difference between them. To interprete these data we have to remember that the magnitude of the distance fluctuations between two Cα atoms is related to the absolute value of the difference between their rigidity scores. This means that a group of atoms having similar scores also has small interatomic distance fluctuations and can be considered as being rigid. Based on the rigidity scores, we can roughly

---

**Figure 4.6:** projections along the first 6 eigenvectors
Figure 4.7: B-factors derived from the total motion in the hyperplane of the first three eigenvectors (thin line) and of the total motion in all other directions (thick line). Asp124, Trp125, Trp175 and His289 are catalytic residues. There is a sharp peak at Pro168.

divide the protein into three regions, A, B, and C, with respectively positive, neutral, and negative scores. The boundaries between them are indicated in the figure. The distance fluctuations between regions A and C are largest. In fig 4.10 the regions are visualized in the three-dimensional structure. Apparently, segment 183-200 should be considered as a separate region. So we subdivide A into $A_I$ (residues 183-200) and $A_{II}$ (residues 1-3,24-37 and 100-114). The segment between residues 149 and 182, assigned to region C, has a rather irregular pattern, meaning that there are also considerable internal distance fluctuations present. A strong negative peak is observed at Pro168 indicating that this residue
Figure 4.8: 11 structures taken from the projection of the trajectory on the first three eigenvectors; thick solid line: projection of the starting structure, thick dashed line: projection at 1ns, thin dashed lines: nine projections separated by 100ps. (drawn with WHATIF [48]).

is rather flexible with respect to any other region in the protein. Trp175 (involved in halide ion binding), has a score that is most similar to the one of region B, which covers most of the main domain, and also contains other residues that are involved in catalysis. We also looked for a possible tunnel, connecting the active site cavity with the exterior and serving as an entrance for the substrate. In the crystal structure such a tunnel was found [7] but it was blocked by the sidechain of Leu262. The solvent accessible surface of the structure, after 1ns, was calculated with WHATIF [48], and the result was visually inspected on the presence of a connection of the active site with the surface of the protein. No such opening was found.
4.4 Conclusions

During the simulation the protein appeared to possess a large flexibility, giving rise to large rms deviations from the crystal structure that stabilized after $\sim 600$ ps. Based on atomic distance fluctuations, the molecule can roughly be subdivided into four regions which were designated as $A_I$, $A_{II}$, B and C. The term rigid body does however not apply to segment 149-183, as this part of the molecule appears to possess a relatively large internal flexibility. It is interesting to note that Trp175, which is part of this region, has only small distance fluctuations with respect to all other catalytic residues that are found in region B which covers most of the main domain. The fact that also the B-factors derived from the simulation are small for all the catalytic residues may further indicate that, in spite of the large rms for the complete structure, the protein is not unfolding.
Figure 4.9: Rigidity scores, derived from the time interval 900-1000 ps of the full trajectory (solid line) and the projection on the first three eigenvectors (dotted line). 0, + and - indicate neutral, positive or negative scores. Trp175 lies in the middle of a region with negative scores, but has itself a neutral score. Pro168 shows irregular behaviour and has a strong negative peak. Based on visual inspection, region A is subdivided into $A_I$ and $A_{II}$.
Figure 4.10: Regions A (thick dashed line), B (thick solid line) and C (thin solid line), obtained by subdividing the molecule based on rigid body scores. The largest distance fluctuations are found between regions A and C. Region A consists of two separate regions. Compared to regions A and B, region C has a relatively large internal flexibility: Trp175 has a neutral score (and should therefore be assigned to region B), whereas Pro168 has a strongly negative score and appears to be flexible relative to any other part of the protein.
CHAPTER 4. MD OF DEHALOGENASE