Supplementary Material (ESI) for Chemical Communications

Towards understanding phosphonoacetaldehyde hydrolase: an alternative mechanism involving proton transfer that triggers P–C bond cleavage

ELECTRONIC SUPPLEMENTARY INFORMATION

Borys Szefczyk

Institute of Physical and Theoretical Chemistry, Wrocław University of Technology, Wybrzeże Wyspiańskiego 27, 50-370, Wrocław, Poland

1 Description of the computational procedure

QM/MM model of the phosphonatase used in these calculations was build using coordinates from the crystal structure (PDB entry code 1RQL) of *Bacillus cereus* phosphonatase complexed with inhibitor, vinyl sulfonate. [1] This is a homodimeric structure, with some terminal amino acids missing, but they do not fall into the sphere investigated here. The inhibitor is bound only in one of the two active sites (chain B). The structure of PALD from earlier calculations [2] was aligned with the inhibitor, which was removed afterwards. QM/MM calculations were performed in CHARMM [3, 4] interfaced with GAMESS. [5] For the QM part, the Density Functional Theory was applied, with the B3LYP functional and 6-31G(d) basis set. The MM part of the system was treated with the CHARMM22 force field. [6] All titrable groups were verified and its protonation state was set according to the possibility to form hydrogen bonds with neighbouring amino acids and contact with the solvent (Table 1). Hydrogens were built into the structure using CHARMM parameters and optimized afterwards, keeping heavy atoms fixed and bond lengths held by SHAKE procedure. Optimization was performed using the Steepest Descent (SD) method (400 steps) and then the Adopted Basis Newton-Raphson (ABNR) method until the gradient RMS of 0.01 kcal·mol^{-1}·Å^{-1} was reached (89 steps). As the PALD is not a standard residue, it is worth to note, that the topology file for this residue was made, using ESP charges calculated at the HF/6-31G(d) level with the CHELPG method which has been found to provide reasonable charge transfer values compared to the experiment. [7] Bonding force field parameters for PALD atoms were not needed, as the PALD molecule was treated entirely quantum-mechanically. All calculations were done using the non-bonded cutoff of 25 Å. The structure was solvated, by superimposing a pre-equilibrated box of water molecules on the protein. The box, $60 \times 60 \times 60$ Å, containing 8000 TIP3P water molecules, was centered at the phosphorus atom of PALD. Next, all water molecules overlapping with other atoms were deleted, using the distance between heavy atoms less than 2.6 Å as a criterion. Then, all water molecules and residues having no atom closer than 20 Å from the phosphorus atom were deleted. The resulting model had 4135 atoms. Position of the water molecules was shortly optimized using the SD method (200 steps, final gradient RMS 0.76) with other atoms kept fixed and then optimized using the ABNR method (2000 steps, final gradient RMS 0.06). Next, water molecules were heated from 100 to 300 K over 1 ps, using the SHAKE procedure for bond constraints and equilibrated over 19 ps at 300 K. After that, another portion

Table 1: Protonation states of the titrable groups in phosphonatase. The same protonation states were used in chain A and B.

Residue type	Residue numbers
ARG protonated	3, 36, 46, 58, 65, 72, 75, 100, 116, 118, 130, 160, 195, 227, 234, 236
ASP deprotonated	2, 12, 19, 55, 82, 134, 148, 154, 155, 186, 190, 222
HIS N_ε protonated	34, 243
HIS doubly protonated	56, 180, 258
GLU deprotonated	7, 27, 31, 43, 44, 62, 69, 80, 85, 88, 89, 91, 92, 110, 117, 131, 139, 172, 193, 210, 215, 216, 217, 219, 225, 228, 231, 239, 247, 251, 253, 257, 260, 263
LYS protonated	5, 35, 47, 53, 109, 121, 138, 146, 168, 183, 192, 229, 261

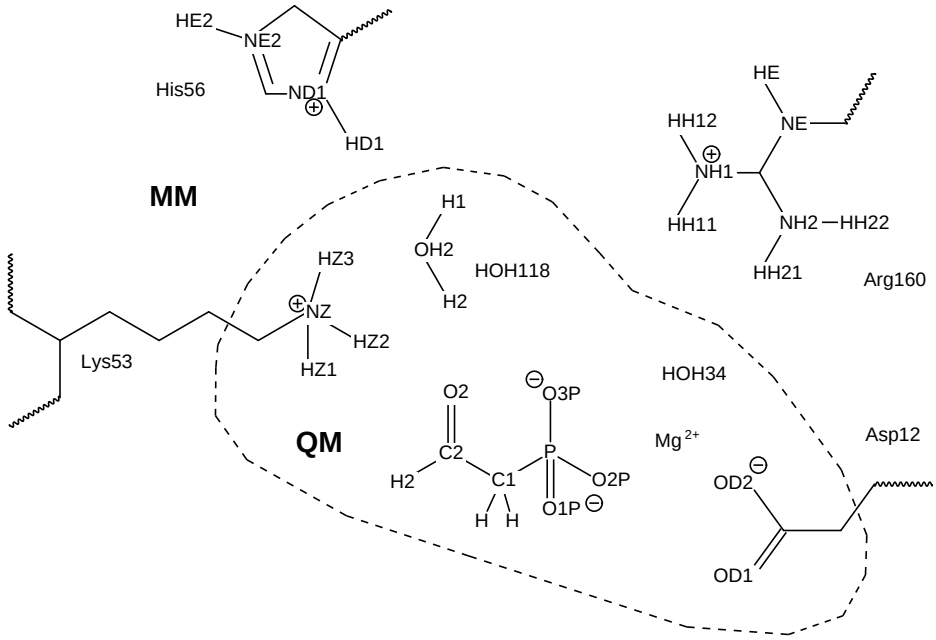

Figure 1: Division of the system into QM and MM regions.

of water was added using the procedure described above (including optimization). All water molecules were again heated and equilibrated over 20 ps in total and optimized in 200 steps of SD method and 197 steps of ABNR (until gradient RMS 0.01 was reached). Next, the system was divided into QM and MM regions shown in Figure 1. Our earlier, unpublished, QM/MM calculations using Hartree-Fock and 3-21G(d) basis set have shown, that significant geometrical changes happen only in the active site, whereas the remaining part of the structure is more or less rigid during the reaction. Therefore, in the present calculation most part of the system was fixed and only the active site residues were allowed to move, including: Phe11, Asp12, Trp13, Met49, Leu52, Lys53, Ile54, Hsp56, Thr126, Tyr128, Arg160, Asp186 – all from chain B. The QM subsystem was described by 280 basis functions. Following protonation states were assumed: Asp12(-1), Lys53(+1), Mg(+2), PALD(-2), resulting in a neutral charge of the QM subsystem. In the QM/MM model, 241 atoms were free to move and 30 atoms were included in the QM

part of the system. The QM region boundary intersects two covalent bonds, therefore two link-atoms had to be introduced into the system. In this way, in the QM part of the calculation, the lysine Lys53 became a methyl-ammonium cation and the aspartate Asp12 became an acetate anion. The magnesium ion and water molecule were also included in the QM region. Similar procedure (except for the fixing of atoms) was successfully applied in former calculations on chorismate mutase [8, 9]. The whole MM part of the system was thoroughly optimized (QM part was fixed) using the SD method (100 steps) and then the ABNR method, until gradient RMS and energy reached the machine precision. Up to this point, all calculations were done using the Molecular Mechanics only, with PALD atoms fixed. PALD binds in the active site, with the phosphonyl group directed to Asp12, held by the magnesium ion, the water molecule and hydrogen bonds formed by hydroxyl group of Thr126 and proton in the peptide bond between Trp13 and Ala14. The phosphorous atom is exposed to the oxygen atom in Asp12, ready for the nucleophilic attack. Carbonyl oxygen in PALD forms hydrogen bonds with Lys53 and the water molecule HOH120, present in the x-ray structure and may have two orientations differing by the rotation about C1–C2 bond in PALD. Therefore, the reactant-enzyme complex can be in R1 or R2 conformation. The structure of these complexes are in perfect agreement with experimental predictions. Optimization of the R1 and R2 complexes was done using the QM/MM method, initially using SD method (100 steps) and then using ABNR method until gradient RMS 10^{-5} kcal·mol^{-1}·Å^{-1} was reached.

Product complexes were constructed by introducing manually some changes in the structure of appropriate reactant (for example, by shifting a proton from Lys53 to PALD) and then by performing QM/MM optimization of such initial model. All geometries were optimized using SD and ABNR methods until gradient RMS of 10^{-5} kcal·mol^{-1}·Å^{-1} was reached, however in some cases the energy change reached machine precision first. Two conformations of the reactant complex (R1 and R2) and two conformations of the product complex (Pd and Pm) were optimized. From these reactant- and product complexes four possible paths were constructed and investigated using TREK module [10] in CHARMM. In brief, the method implemented in TREK relies on interpolation between provided geometries and then optimization of the initial path until the saddle points are found. In the resulting path, the saddle points satisfy the criterion on gradient RMS deviation of 10^{-3} kcal·mol^{-1}·Å^{-1}, however, due to many minimization cycles performed, the final convergence on RMS value is usually better. The points being not minima or saddles are optimized only to the extent that the path is monotonic between the extrema.

2 Key distances

Comparison of the geometrical parameters of QM/MM models of the reactant, TS and product complexes (distances in Å). Labeling of atoms is shown in Figure 1.

Distance	R1	R2	TSm1	TSm2	TSd1	TSd2	Pd	Pm
PALD:C1 – PALD:P	1.971	2.001	2.052	2.118	2.318	2.391	3.840	3.471
PALD:C1 – PALD:C2	1.434	1.432	1.405	1.400	1.370	1.371	1.332	1.332
PALD:C2 – PALD:O2	1.261	1.262	1.282	1.283	1.324	1.316	1.375	1.369
PALD:O2 – Lys53:NZ	2.675	2.625	3.107	3.218	2.524	2.479	2.681	3.806
PALD:O2 – Lys53:HZ1	1.982	1.780	3.098	2.896	1.097	1.178	1.011	3.754
PALD:P – Asp12:OD1	3.101	3.135	3.069	3.140	2.971	2.955	1.886	1.936
PALD:O2P – Mg	1.948	1.964	1.967	1.953	1.967	1.990	1.949	1.940
Asp12:OD2 – Mg	2.018	2.032	2.023	2.030	2.020	2.033	2.079	2.067
PALD:O2 – HOH120:OH2	2.650	2.667	2.392	2.408	2.764	2.721	2.694	2.723
PALD:O2 – HOH120:H2	1.733	1.732	1.192	1.213	1.857	1.777	1.743	0.985
PALD:O3P – Arg160:NH1	2.684	2.672	2.686	2.724	2.712	2.685	2.589	2.591
PALD:O3P – Arg160:HH11	1.775	1.749	1.772	1.832	1.806	1.759	1.601	1.602
Lys53:NZ – HOH120:OH2	2.743	2.769	2.524	2.518	3.055	2.942	2.970	2.765
Lys53:NZ – Lys53:HZ3	1.052	1.046	1.225	1.216	1.018	1.022	1.020	1.783

References

[1] M. C. Morais, G. Zhang, W. Zhang, D. B. Olsen, D. Dunaway-Mariano and K. N. Allen, *J. Biol. Chem.*, 2004, **279**, 9353–9361.

[2] B. Szefczyk, P. Kedzierski, W. A. Sokalski and J. Leszczynski, *Mol. Phys.*, 2006, **104**, 2203–2211.

[3] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan and M. Karplus, *J. Comp. Chem.*, 1983, **4**, 187–217.

[4] P. D. Lyne, M. Hodoscek and M. Karplus, *J. Phys. Chem. A*, 1999, **103**, 3462–3471.

[5] M. W. Schmidt, K. K. Baldridge, J. A. Boatz, S. T. Elbert, M. S. Gordon, J. H. Jensen, S. Koseki, N. Matsunaga, K. A. Nguyen, S. J. Su, T. L. Windus and M. Dupuis, *J. Comput. Chem.*, 1993, **14**, 1347–1363.

[6] A. D. MacKerell, D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin and M. Karplus, *J. Phys. Chem. B*, 1998, **102**, 3586–3616.

[7] B. Szefczyk, W. A. Sokalski and J. Leszczynski, *J. Chem. Phys.*, 2002, **117**, 6952–6958.

[8] K. E. Ranaghan, L. Ridder, B. Szefczyk, W. A. Sokalski, J. C. Hermann and A. J. Mulholland, *Mol. Phys.*, 2003, **101**, 2695–2714.

[9] K. E. Ranaghan, L. Ridder, B. Szefczyk, W. A. Sokalski, J. C. Hermann and A. J. Mulholland, *Org. Biomol. Chem.*, 2004, **2**, 968–980.

[10] S. Fischer and M. Karplus, *Chem. Phys. Lett.*, 1992, **194**, 252–261.