Conformational Flexibility Influences Structure-Function Relationships in Nucleic Acid N-Methyl Demethylases

Sodiq O. Waheed,[a,&] Rajeev Ramanan,[a,&] Shobhit S. Chaturvedi,[a] Jon Ainsley,[b] Martin Evison,[b] Jennifer M. Ames,[c] Christopher J. Schofield,[d, *] Christo Z. Christov,[a,b *] and Tatyana G. Karabencheva-Christova[a,b *]

[a] Department of Chemistry, Michigan Technological University, Houghton, Michigan 49931, United States.

[b] Faculty of Health and Life Sciences, Northumbria University, Newcastle upon Tyne, NE1 8ST, United Kingdom.

[c] Centre for Research in Biosciences, Faculty of Health and Applied Sciences, University of the West of England, Bristol, BS16 1QY, United Kingdom. Present address: Jenny Ames Consulting Ltd, Reading, RG6 5PY, United Kingdom.

[d] The Chemistry Research Laboratory, The Department of Chemistry, Mansfield Road, University of Oxford, Oxford, OX1 3TA, United Kingdom.

[&]S.O.W., and R.R. made equal contributions to this article.

[*]C.J.S., C.Z.C and T.G.K.C. are joint corresponding authors.

Table of Contents

	Methods	
1	System Preparation	S4
2	Details of MCPB	S4
3	MD Simulations	S4
4	QM Cluster calculations	S5
5	QM/MM calculations	S5
6	Molecular Mechanics/Generalized Born Surface Area (MM/GBSA)	S6

List of Figures

List of Tables

Methods

System Preparation

An X-ray crystal structure of FTO (PDB ID code: 4IDZ[1]) was used for building the FTO model used in the computational studies. Missing (i.e. not observed by crystallography) residues from loop regions were added using Modeller.[2] The substrate 3-methylthymidine (3-meT) was modelled into the 4IDZ[1] based structure using another structure (3LFM[3]) by aligning the two structures using Maestro (Schrodinger LLC, New York). This was followed by replacement of Ni(II) with Fe(II) and N-oxalylglycine with 2-oxoglutarate (2OG) using GaussView 5.0.[4] A crystal structure of the AlKB (PDB: 4NID[5] in complex with Mn(II), 2OG and double stranded DNA containing N6-methyladenine (6MA) was used in computational studies. The Mn(II) was replaced with Fe(II) using GaussView 5.0.[4] AlkB was also modelled with the nucleoside only (6MA (nsd)) instead of double stranded DNA to compare with the FTO single base simulations and to investigate how the DNA effect the conformational dynamics of AlkB. The Amber parameters for N6-methyladenine were developed using the appropriate CIF file and the AM1-BCC charge model available in Antechamber and prepgen from AmberTools15. A wildtype structure of AlkB (PDB: 3BIE[6]) with substrate N1-methyladenine (m1A) was also used in the simulation studies. The parameters for the substrate were developed using Antechamber.

The likely protonation states of ionisable residue sidechains were assessed using the H++ server.[7] Histidine residues coordinating with the metal centre were assigned protonation states based on visual inspection of their local environment. The cofactor analogue N-oxalylglycine was modelled to 2OG by replacing its NH with a methylene using GaussView 5.0.[4] Hydrogen atoms were added to 2OG and 3-methylthymidine using the reduce program in Amber.[8] The amber parameters for 2OG and 3-methylthymidine were developed using the General Amber force field (GAFF) [9] using Antechamber. The atomic charges of the cofactors were calculated based on the electrostatic potential from single point HF/6-31G* calculations using Gaussian09.[10] The restrained electrostatic potential (RESP)[11] method was used for the charge fitting procedure.

MCPB: The Amber parameters for the active site containing Iron (Fe(II) high spin S=2, M=5 , ground state [12-19] and the coordinating ligands (2OG (bidentate ligation), histidine and aspartic acid (both monodentate ligation) were prepared using the Metal Centre Parameter Builder (MCPB) and MCPB.py v1.0 Beta2[21] for a 5-coordinate (5C) distorted square pyramidal active site. The binding of the substrate in the vicinity of the Fe(II) in the presence of 2OG results in the dissociation of coordinated water giving a 5C distorted square pyramidal geometry.[12,15] The metal centre parameters were derived based on the bonded and electrostatic model approach in which the coordinating ligands are connected to metal through covalent bonds. The bond and the angle force constants were derived using the Seminario method; point charge parameters for the electrostatic potential were obtained using the ChgModB method. Pabis *et al.* have applied the MCPB tools for description of the mononuclear non-heme iron centre and iron-sulfur Rieske cluster.[22] Molecular dynamics simulations run using these parameters have successfully reproduced the crystallographically observed geometry of metal-ligand complexes.[22]

MD Simulations: Molecular dynamics simulations were performed using the GPU version[23] of the PMEMD engine integrated with Amber 14. [24] The FF14SB[25] force field was used in all the simulations and the Leap module was used to add missing hydrogen atoms and counter ions for neutralisation of the protein system. All the systems were immersed into a truncated octahedral box with TIP3P water molecules[26], such that no protein atom was within 10 Å of any box edge. Periodic boundary conditions were employed in all the simulations. Long-range electrostatic interactions were calculated using the Particle Mesh Ewald (PME) method[27] with a direct space and vdW cut-off of 8 Å. The various systems were subjected to energy

minimization using first steepest descent (5000 steps) followed by conjugate gradient (5000 steps) to eliminate clashes. Solute molecules were restrained using a restrained potential of 100 kcal mol^{-1} Å^2; only solvent and ions were allowed to minimize. This was followed by full minimization of the entire system with both steepest descent (5000 steps) and conjugate gradient (5000 steps) treatments to relax the system prior to productive simulation. All the energy minimization, heating and equilibration steps were performed with the CPU version of PMEMED. The systems were then subjected to controlled heating from 0 to 300K at constant volume using Langevin thermostat[28] with a collision frequency of 1 ps^{-1} using a canonical ensemble (NVT) MD simulation for 400 ps. The solute molecules were restrained using harmonic potential of 10 kcal mol^{-1} Å^2 during the heating process. The SHAKE algorithm[29] was used to constrain bonds involving hydrogen. This was followed by equilibration at 300K in an NPT ensemble for 1 ns without restraints on solute molecules; pressure was maintained at 1 bar using Berendsen barostat.[30] A productive MD run with explicit solvent for continuous 1μs was performed in a NPT ensemble with a target pressure set at 1 bar and constant pressure coupling of 2ps. The frames from the productive run were saved every 10 ps.

Trajectories were analysed using CPPTRAJ,[31] VMD,[32] UCSF Chimera,[33] and R (Bio3D[34]). The Root Mean Square Deviation (RMSD) of Cα atoms of the protein with respect to minimized crystal structure, Root mean square fluctuations (RMSF), electrostatic interactions, hydrogen bonding, solvent accessible surface area (SASA), and cluster analysis were performed. The Bio3D package[34] in R was used to produce PCA and domain cross correlations as described by Singh *et al*.[35]

QM cluster calculations: Snapshots of structures were obtained from the equilibrated MD trajectory of the systems described above. GaussView 5.0[4] was used to set up QM calculations and Gaussian09 code [10] was used to run all QM calculations. In all the calculations, the Fe(II) (high spin S=2, M=5, ground state [12-19]) and the coordinating ligands (2OG (bidentate ligation), histidine and aspartic acid(both with monodentate ligation) were used. The histidine and aspartic acid residues were truncated and restrained at their Cβ positions; hydrogen atoms were added to saturate bonds. The geometry optimization, frequency calculations and single point calculations were performed with Density Functional Theory (DFT) using unrestricted UBP86 functional with 10 % exact HF (Hartree Fork) exchange with 6-311G* basis set of the Fe and its coordinating atoms (oxygen and nitrogen) from the ligands; for the rest of the atoms we employed 6-31G*[36]. A conductor-like polarizable continuum model (CPCM) with ε=4.3 (diethyl ether as solvent) [37] was used in the QM calculations to mimic the hydrophobic active site.[38-40]

QM/MM calculations: Snapshots for the QM/MM calculations were obtained from the MD simulations on all the systems. In particular, snapshots were taken from the minimized crystal structure, 800ns, 700ns, 300ns, 200ns for FTO and minimized crystal structure, 200ns, 300ns, 400ns, 500ns for AlkB. The residues of all the enzymes including the water molecules which are within 35 Å of Fe(II) (expect for AlkB where whole protein and DNA were used and water up to 35 Å) were involved in QM/MM optimization. These snapshots were first subjected to energy minimization for 10,000 steps by using both steepest descend (5000) and conjugate gradient (5000) algorithms in Amber14. Active site residues were restrained with a restrained potential of 100 kcal mol^{-1} Å^2 in the energy minimization, in order to maintain the geometry of the active site. The energy minimized snapshots of all the enzymes were prepared using the Schlegel's toolkit TAO[41] for ONIOM[42-46] calculation in Gaussian09.[10] Residues within 20 Å of Fe(II), including water molecules, were allowed to move freely during geometry optimization; the rest of the system was frozen during geometry optimization in ONIOM. The QM/MM system was prepared using GaussView 5.0; all calculations were run using Gaussion09. Residues were assigned with the standard bonded and non-bonded terms available from the ff99SB force field available in Gaussion09. The electrostatic embedding scheme was used in the geometry optimization; however, we also used the mechanical embedding scheme for some snapshots. The non-bonded van der Waals parameters for the Fe(II) were obtained using

the method of Li *et al.*[47] The QM region in the QM/MM calculation is consistent with the QM calculation performed above and link atoms were used to saturate the dangling bond in the QM/MM calculation.[48]

Molecular Mechanics/Generalized Born Surface Area (MM/GBSA): The binding free energy calculations with AlkB were performed using the Molecular Mechanics/Generalized Born Surface Area (MM/GBSA) approach.[49-51] The binding free energy was calculated taking into account 2000 snapshots of equilibrated trajectory from a 1μs molecular dynamic production run. The following set of equations describes the calculation of the binding free energy:

$$\Delta G = G_{complex} - G_{receptor} - G_{ligand} \tag{1}$$

$$\Delta G_{bind} = E_{gas} + G_{sol} - T\Delta S \tag{2}$$

$$E_{gas} = E_{int} + E_{vdw} + E_{ele} \tag{3}$$

$$G_{sol} = G_{GB} + G_{SA} \tag{4}$$

$$G_{SA} = \gamma SASA \tag{5}$$

E_{gas} signifies the gas-phase energy; E_{int} signifies internal energy; and E_{ele} and E_{vdw} signify the electrostatic and van der Waals contributions, respectively. E_{gas} is evaluated directly from the FF14SB force field terms. The solvation free energy, denoted by G_{sol}, can be decomposed into polar and nonpolar contribution states. The polar solvation contribution, G_{GB}, is determined by solving the GB equation, whereas, G_{SA}, the nonpolar solvation contribution is estimated from the solvent accessible surface area (SASA) determined using a water probe radius of 1.4 Å. T and S correspond to temperature and total solute entropy, respectively.

A

B
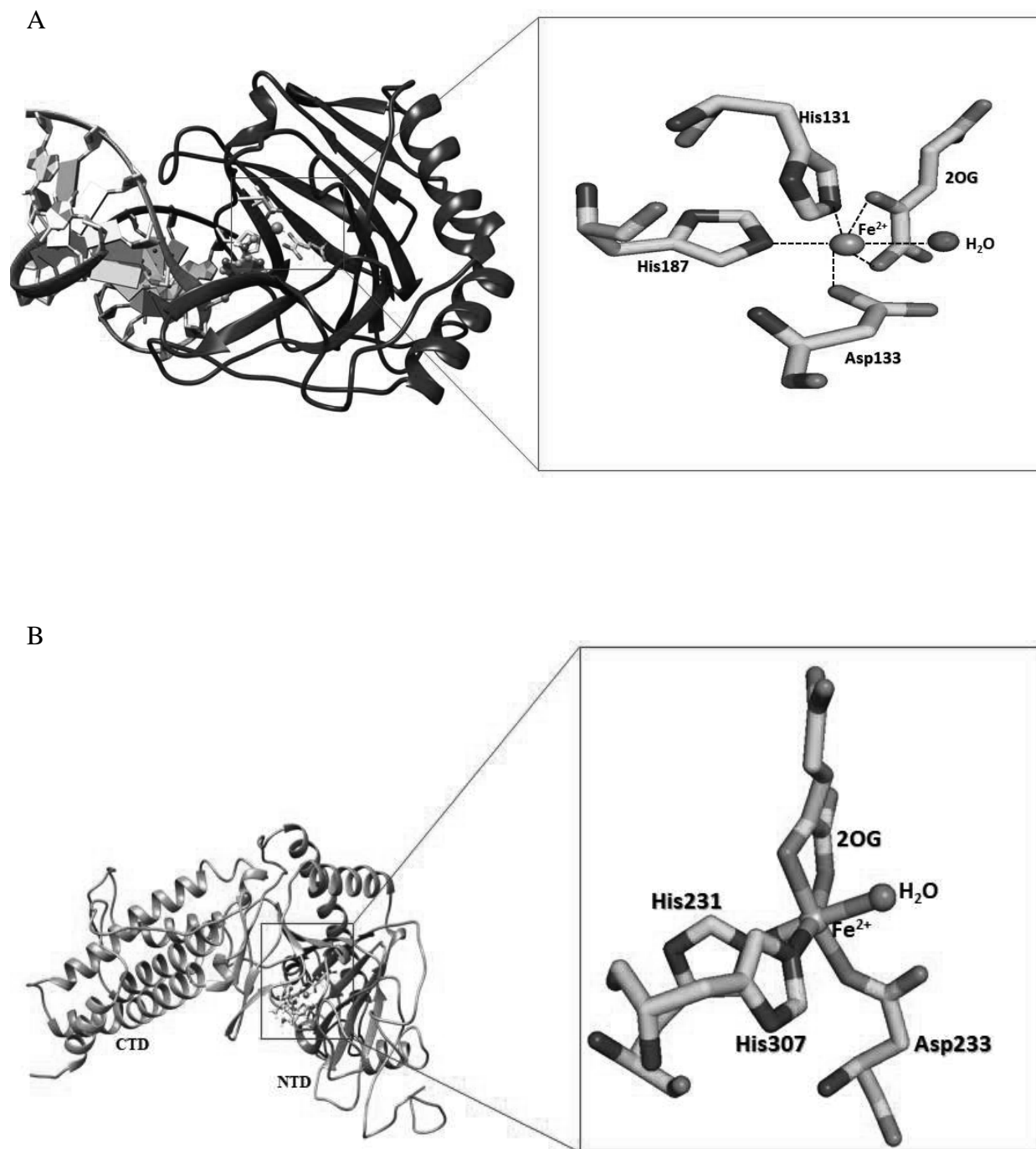
Fig. S1. Views from the active sites of AlkB (A) and FTO (B). Note the conserved triad of metal binding residues (HDH), bidentate binding of the co-substrate 2OG, and that octahedral coordination of Fe(II) is completed by a water molecule.
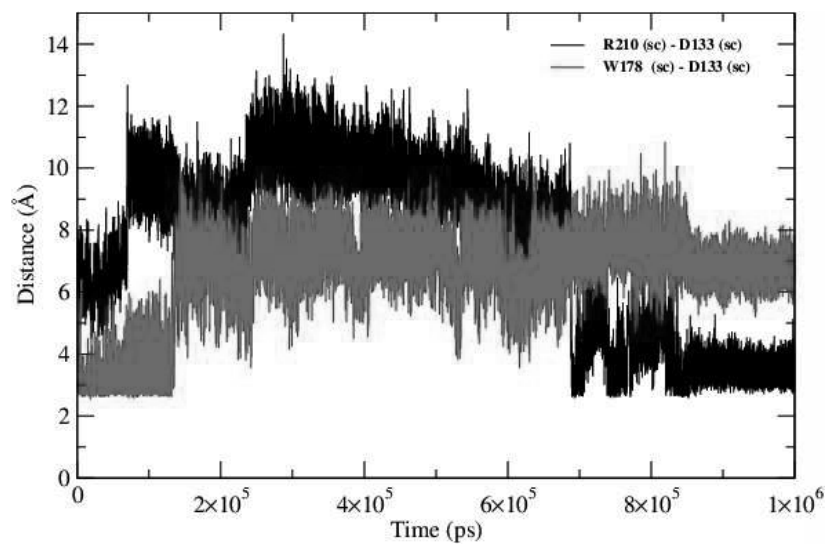
Fig. S2. Interactions of the AlkB R210 and W178 sidechains with the sidechain of D133 in AlkB-DNA complex MD simulation. Distances were measured between the centres of masses of the atoms forming the guanidino group of arginine (R210), the indole group of tryptophan (W178) and carboxyl group of aspartate (D133). Note W178 interacts more strongly with D133 during the last 200 ns of the simulation than in the crystal structure.
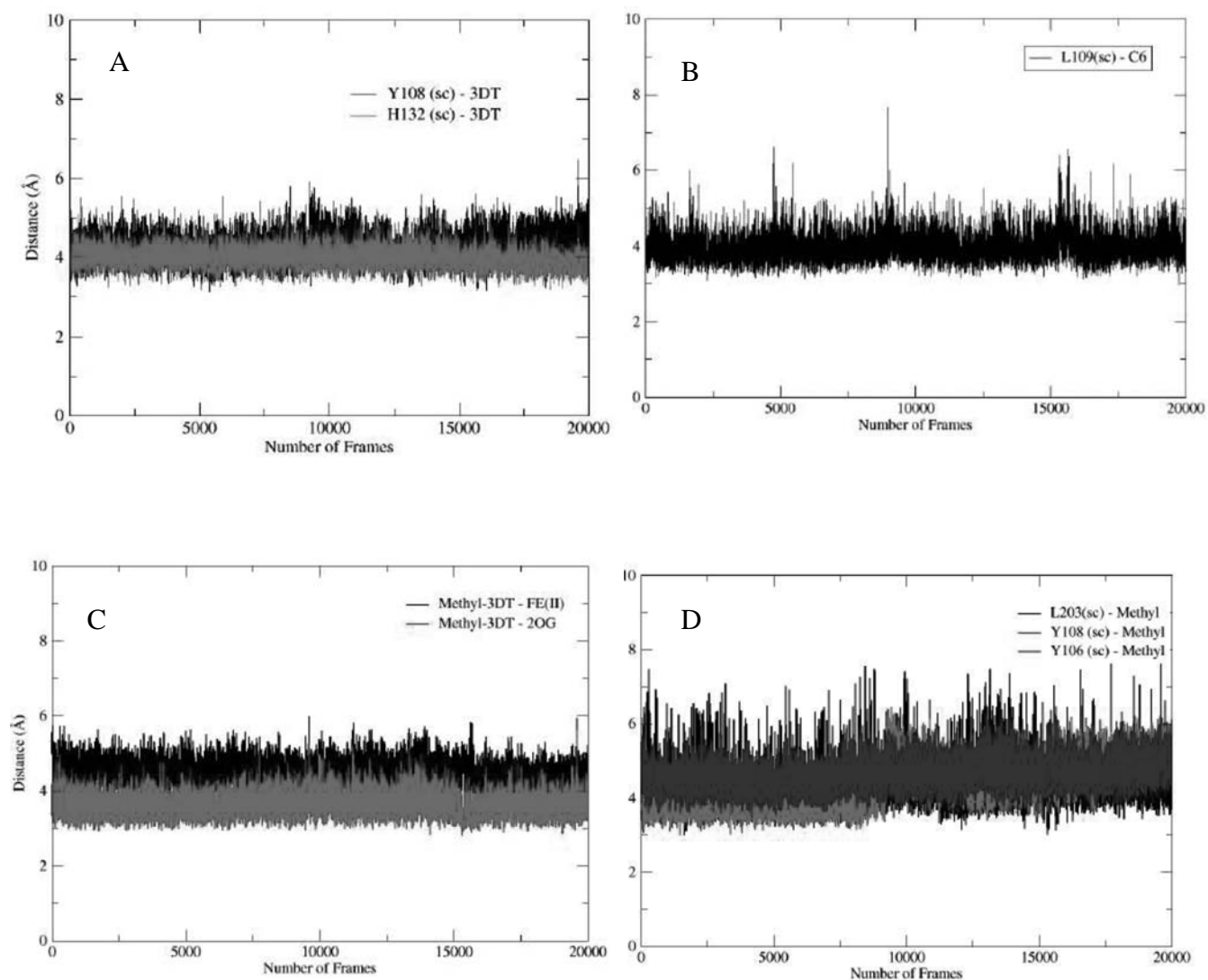
**Fig. S3.** The interaction of 3-meT in FTO during the 1 μs trajectory. (A) Hydrophobic interactions of the side chains of Y108 and H231 with the thymidine ring of 3-meT; (B) Hydrophobic interactions between the side chain of L109 and the sugar ring of 3-meT; (C) Interaction of the substrate methyl group (to be de-methylated) of 3-meT; (D) van der Waals interactions of the neighboring residues with the methyl group of 3-meT.
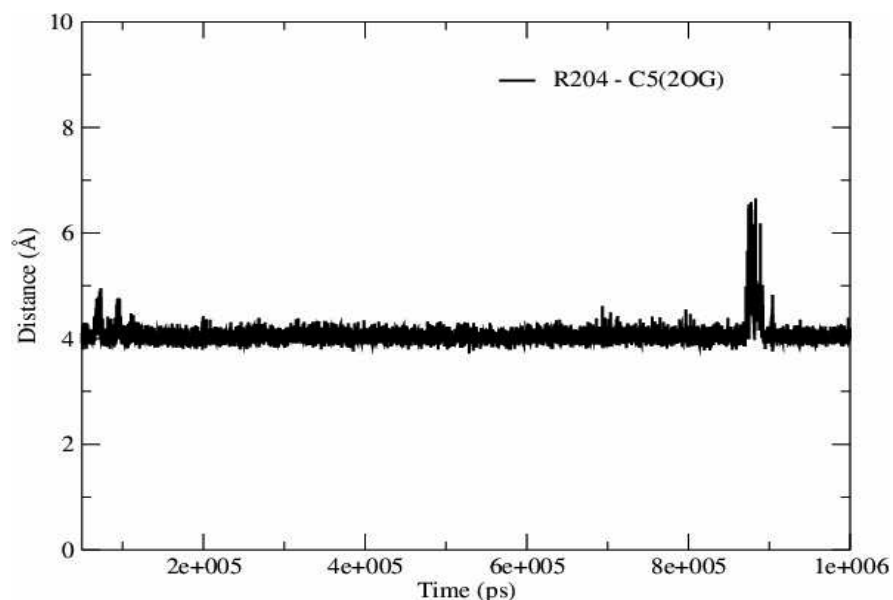
Fig. S4. Electrostatic interactions of the side chain of R204 of with the non-metal coordinating C5 carboxylate of 2OG. Distances were measured between the centres of masses of the atoms of the guanidino group of arginine (R204) and the C5 carboxylate of 2OG.
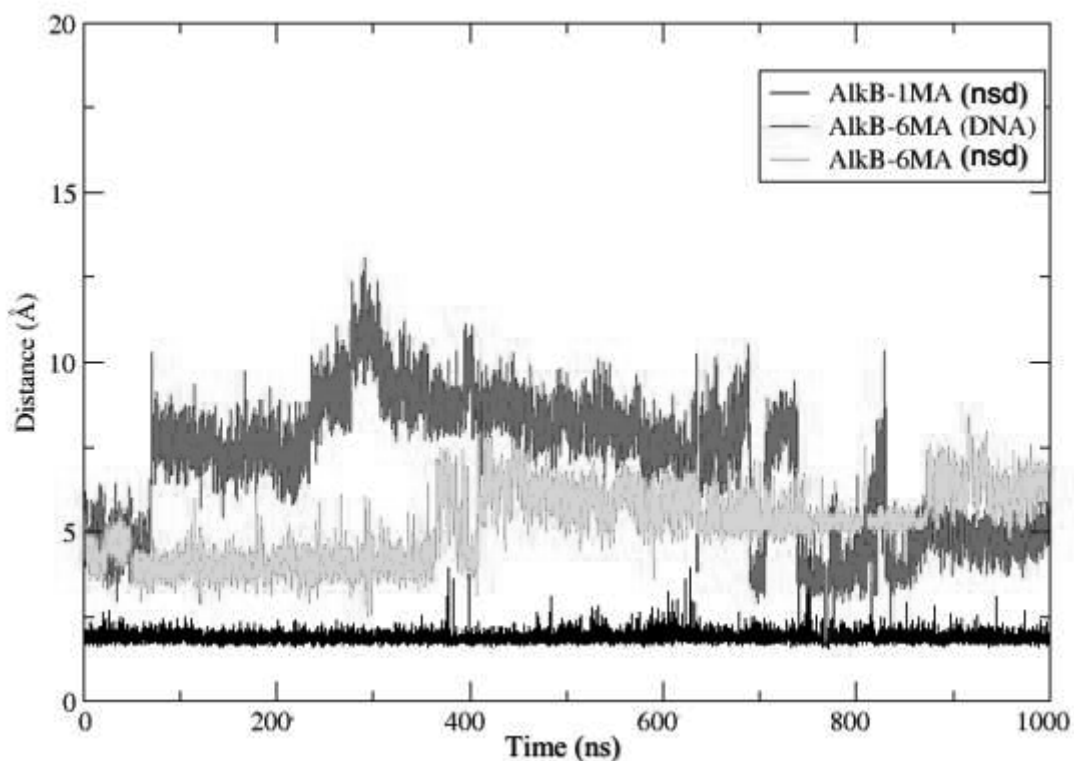
Fig. S5. Hydrogen bond interactions of the side chain of R210 with the non-metal coordinating carboxylate oxygen of D133. Distances were measured between the centres of masses of the atoms of the guanidino group of R210 and the carboxylate of D133.

S10
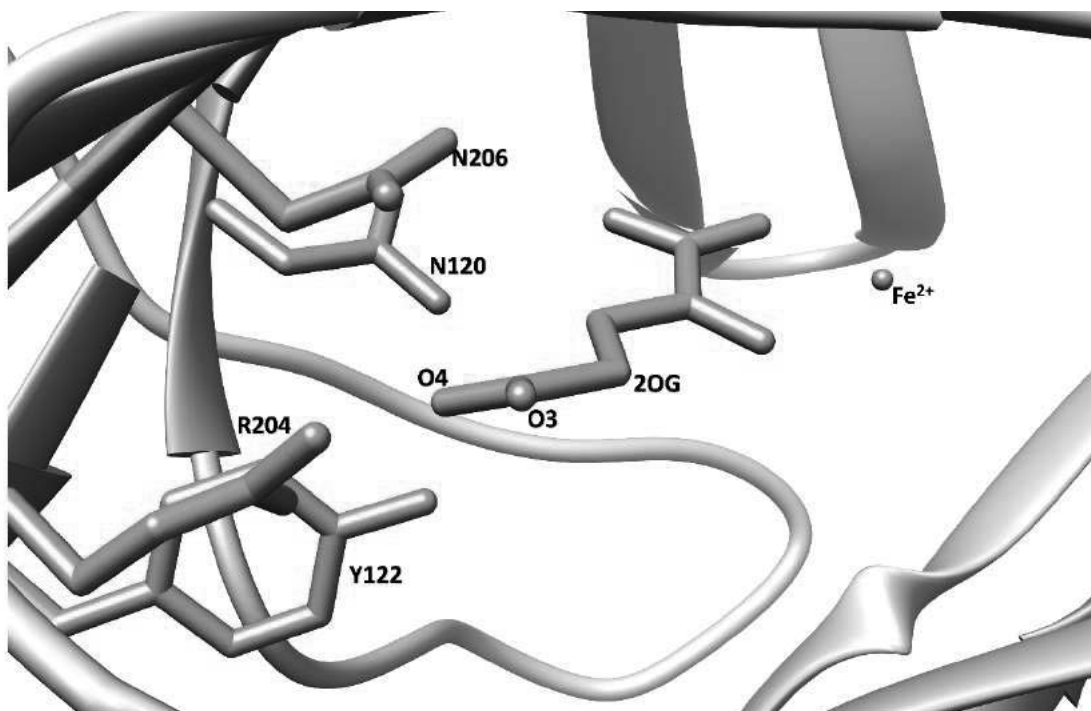
Fig. S6. The 2OG co-substrate in AlkB and residues that are involved in its binding.
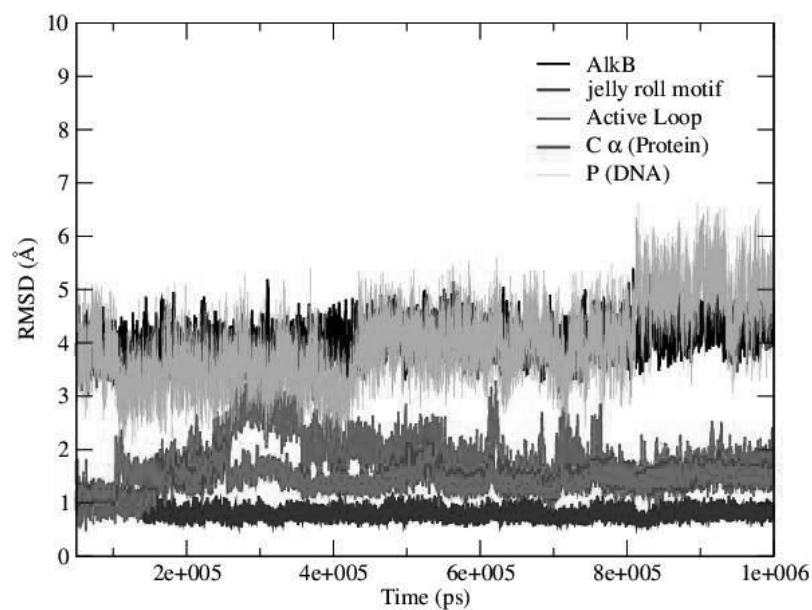
Fig. S7. The RMSD profile of the full AlkB complex (AlkB-6MA(DNA)) along with the active site bordering loop (residues 133-139), DSBH core motif (residues 187-195), and DNA. The overall RMSD of the full AlkB complex was performed using the Cα and P atoms of protein and DNA, respectively. Note that different parts of the complex are characterized by different flexibilities.
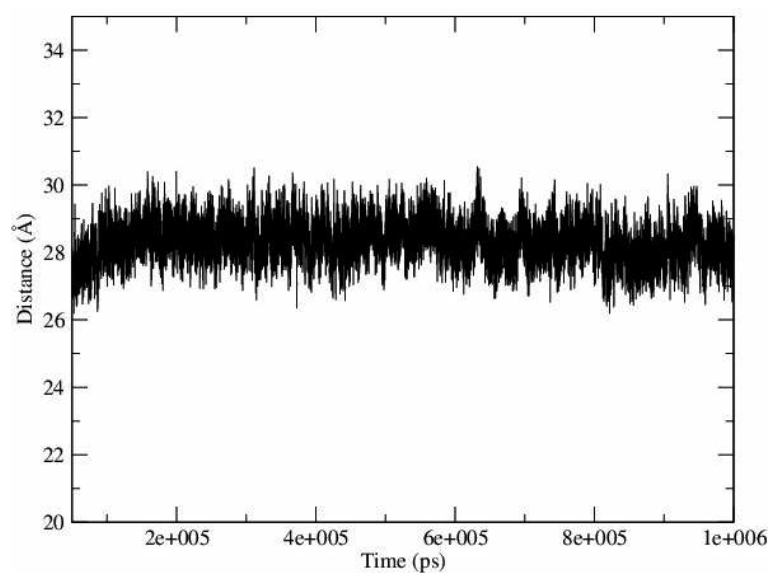
Fig. S8. The distance between centre of mass of the protein and DNA in the AlkB-6MA(DNA) complex during the MD simulation.
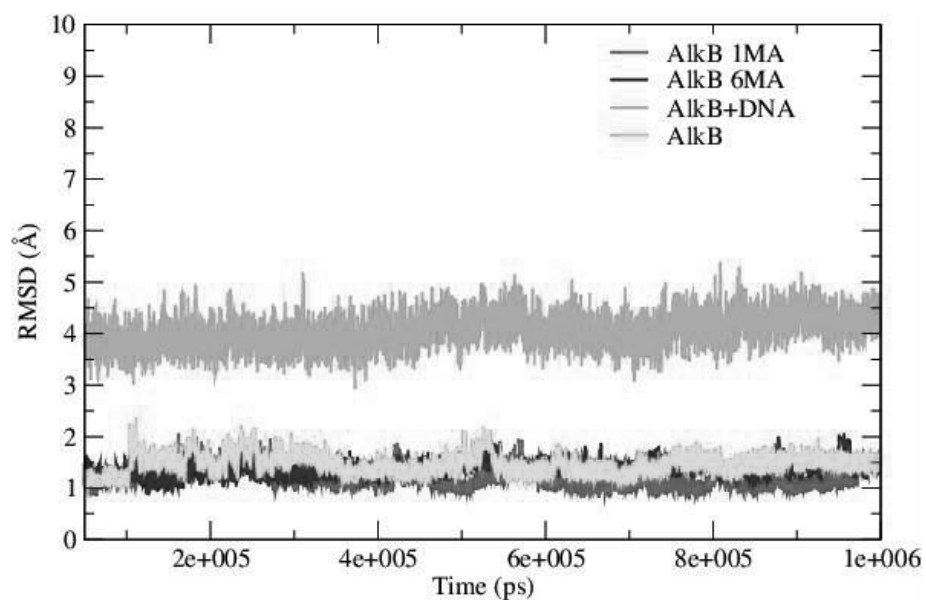
Fig. S9. RMSD analysis of the AlkB-1MA in comparison to AlkB-6MA and AlkB-6MA(DNA) complexes. Note, that the bulk DNA is the most flexible component of the system.
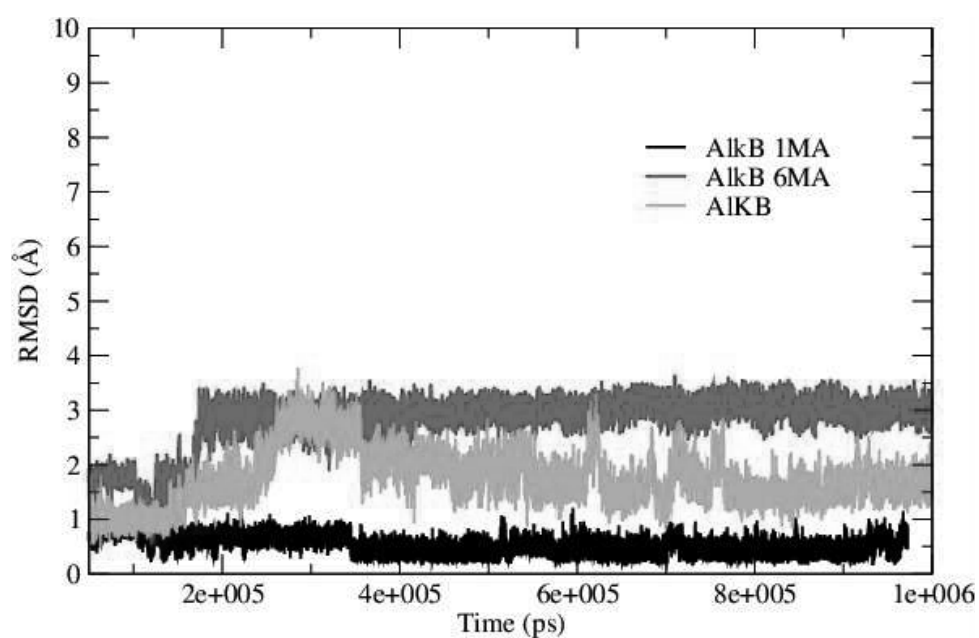
Fig. S10. The RMSD of the active site region (residues 132-187) of wildtype AlkB and its variants, during the MD simulation.
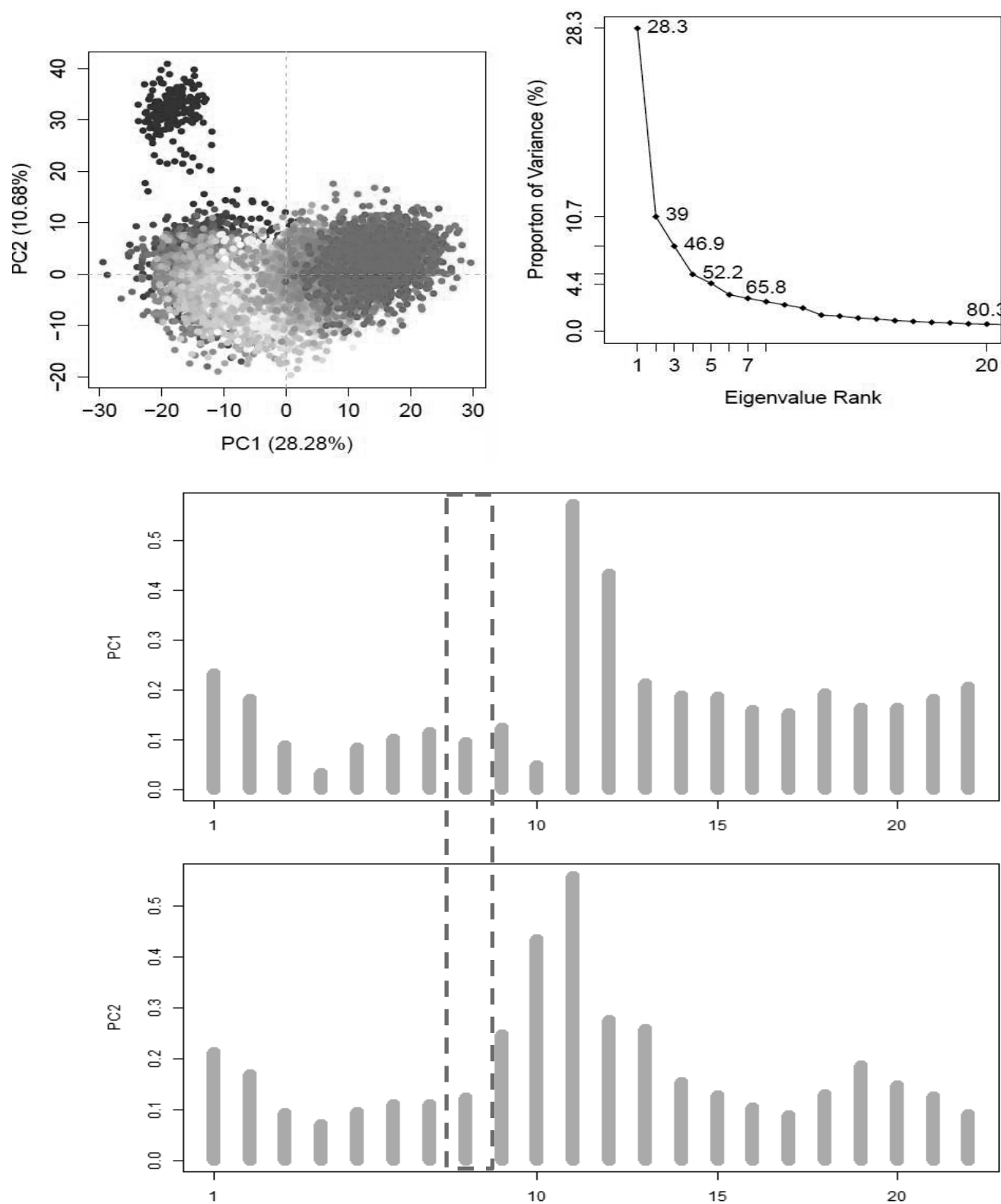
Fig. S11. The contribution of individual DNA bases from the PC1 and PC2 in the AlkB-6MA(DNA). The region marked by the red dotted line is the methylated base 6MA. PC1 and PC2 are shown as a function of residue numbers.

S14
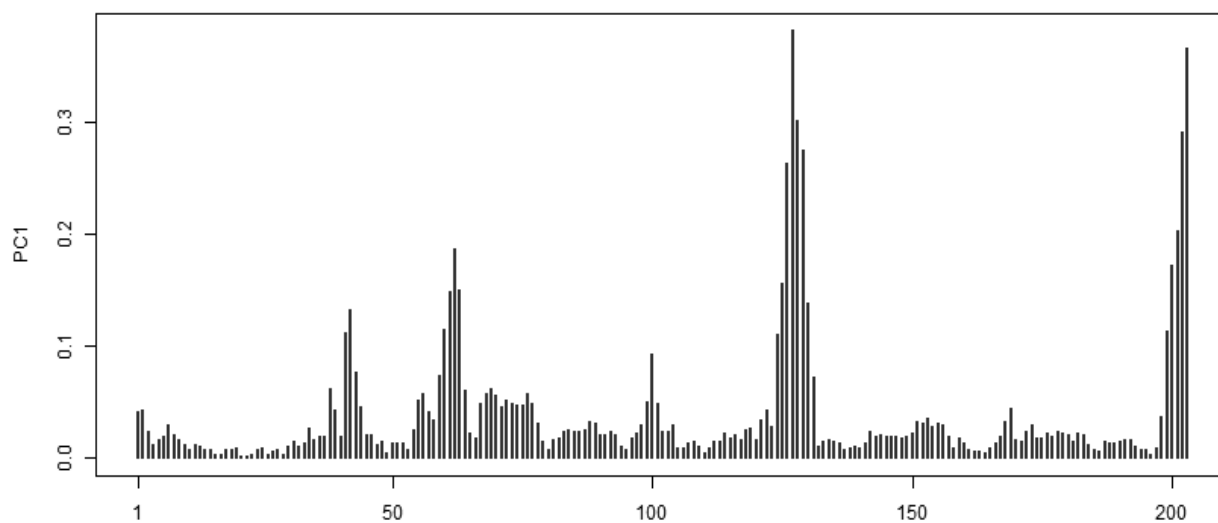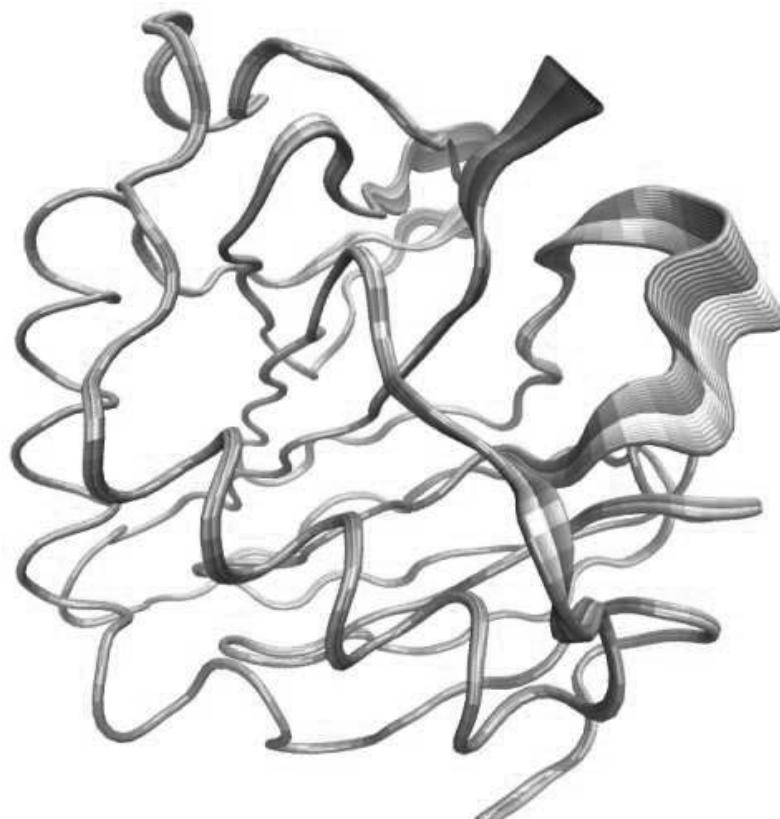
Fig. S12. PCA analysis of the AlkB-6MA(nsd) complex, showing projection of the PC1 on the enzyme residues indicating the major motions in the protein by removing the noise from translational and rotational data from the MD trajectory. The PC1 is shown as a function of residue numbers.
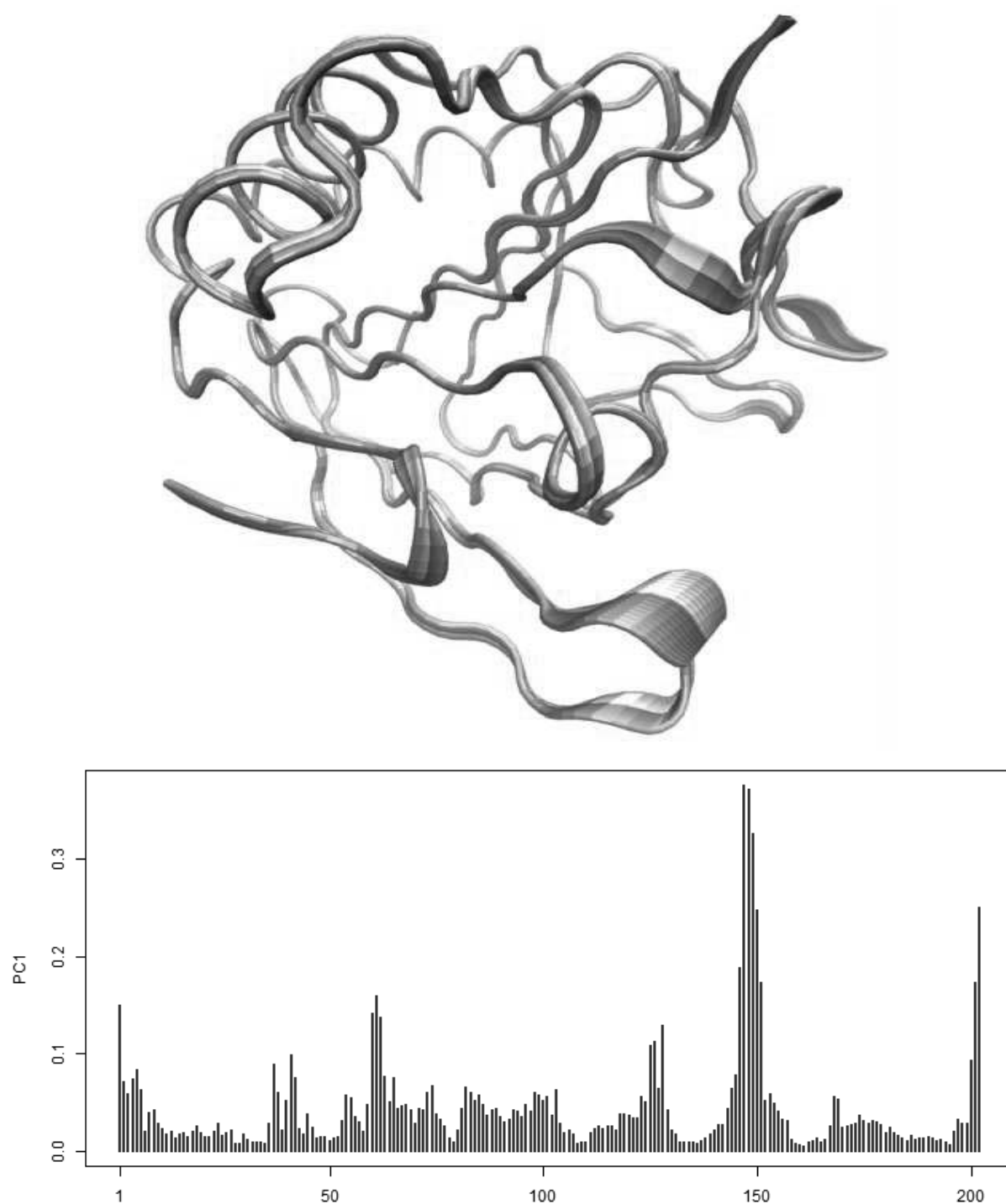
Fig. S13. PC1 of the AlkB-1MA system. The PC1 is shown as a function of residue numbers.
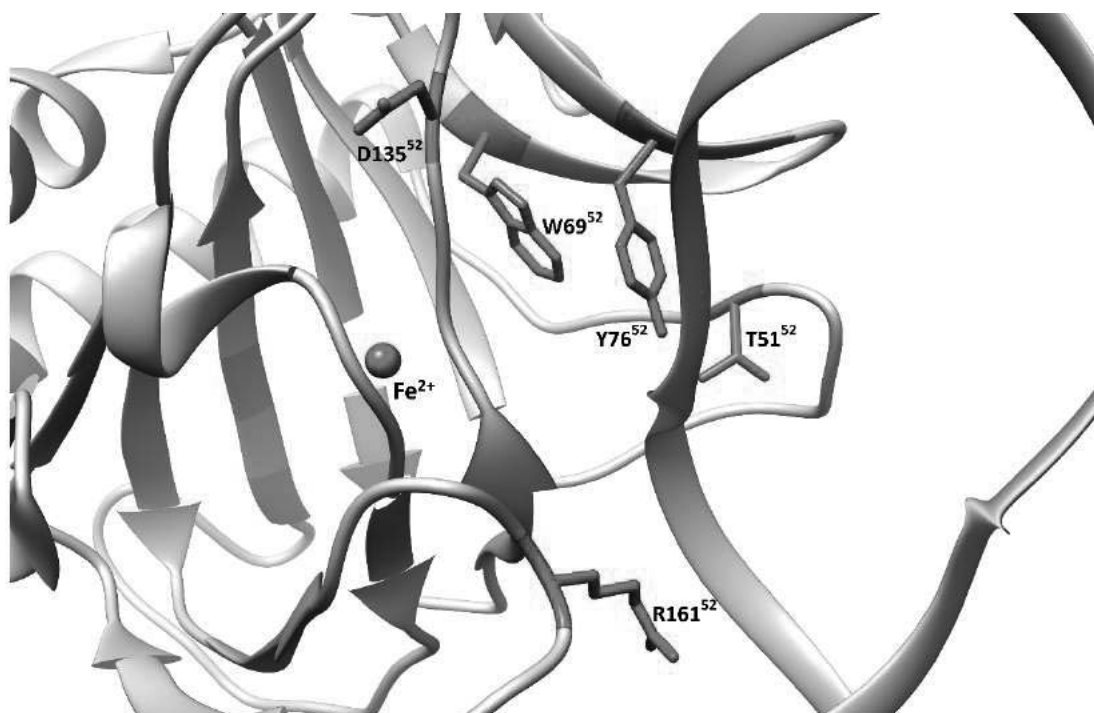
Fig. S14. The positions of AlkB residues whose substitution is reported to influence the activity of the enzyme.[52]
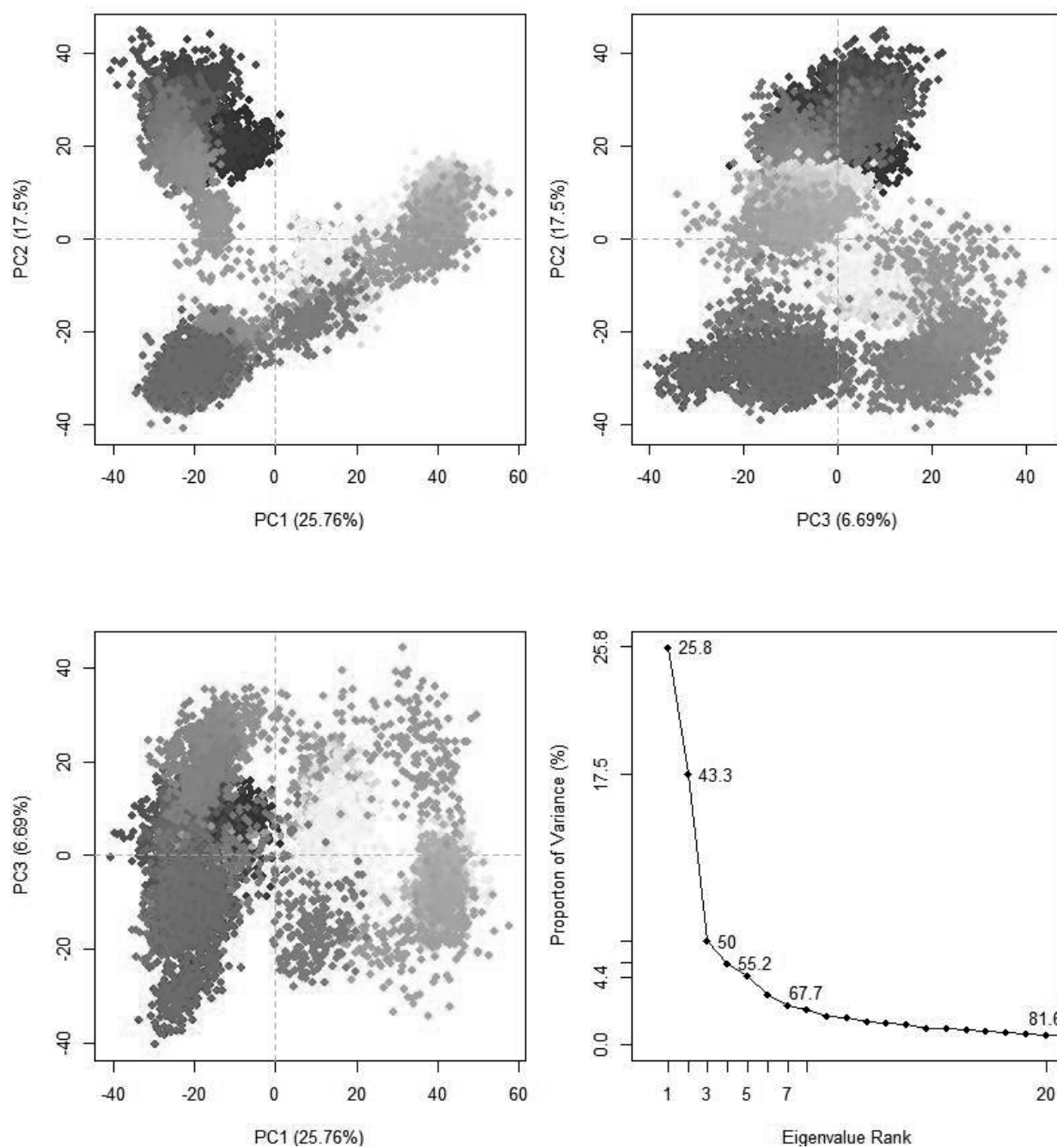
Fig. S15. The projection of the first three principal components for the last 900 ns FTO equilibrated trajectories. The first three eigenvectors represent half of the overall variance in the data set.
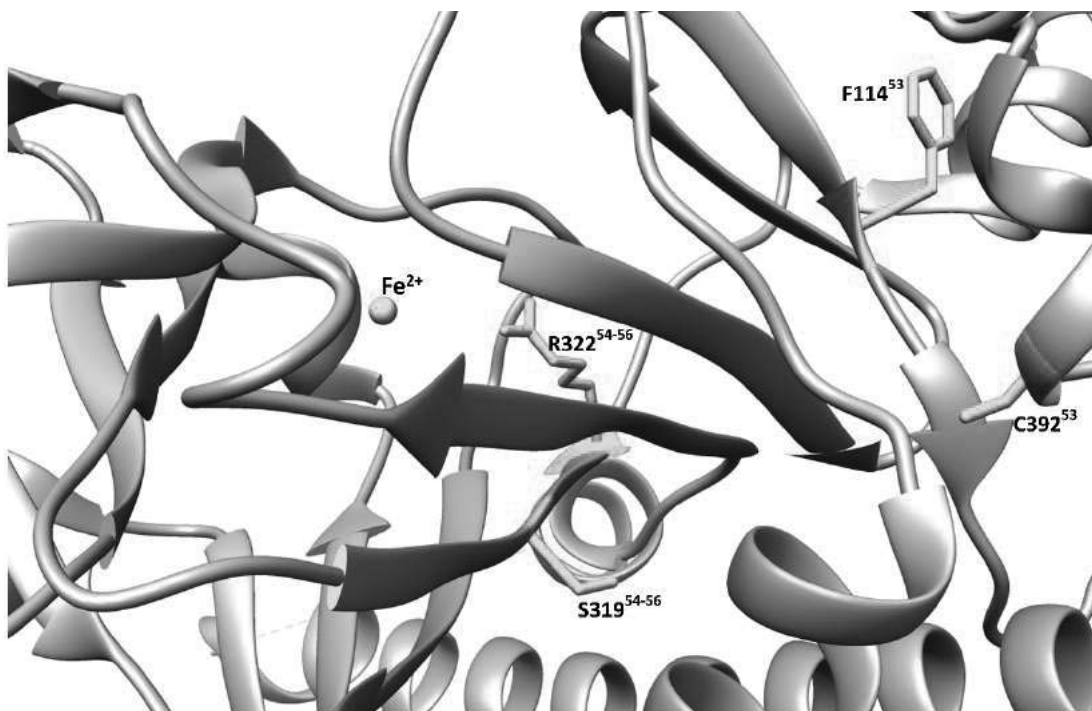
Fig. S16. The positions of FTO residues whose substitutions influence the enzyme activity[53] (F114 and C392) as well as those correlated with pathological changes (R322 and S319).[54-56]
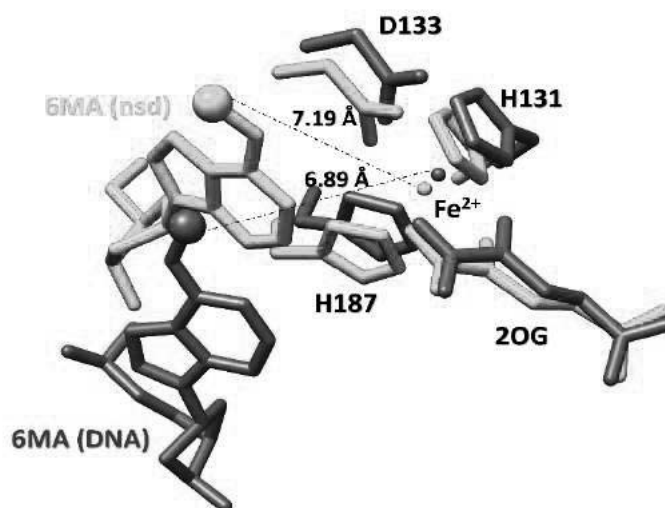
Fig. S17. Overlaid QM/MM optimized structures of AlkB-6MA (nsd) (green) and AlkB-6MA (DNA) (purple). The methyl group is shown as the larger sphere.
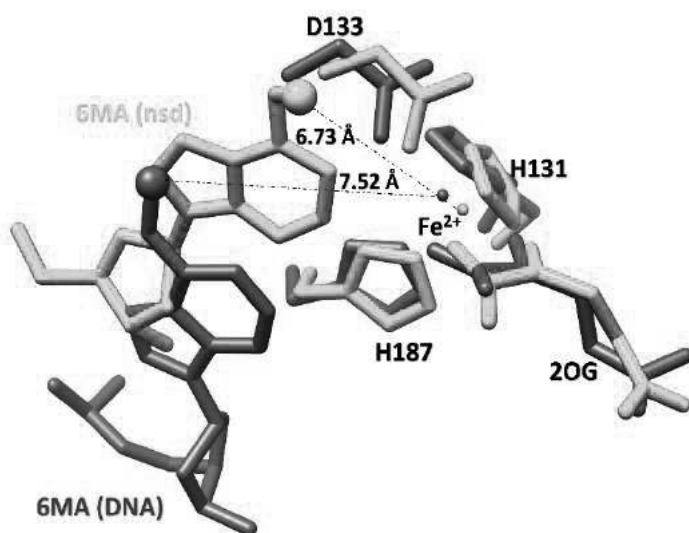
Fig. S18. The overlaid MD structures of AlkB-6MA (nsd) (green) and AlkB-6MA (DNA) (purple). The methyl group is shown as the larger sphere.

Table S1. The relative free energy (in kcal/mol) of binding of 6MA and 1MA to AlkB enzymes calculated using MMGBSA.

Energy Component	AlkB 6MA	AlkB 1MA
ΔE_{vdW}	-34.2588	-34.9231
ΔE_{ele}	-14.5910	39.4729
$\Delta E_{ele, sol}$ (GB)	29.1608	-27.2597
ESURF	-3.6987	-3.9800
ΔG gas	-48.8498	4.5497
ΔG solv	25.4621	-31.2397
TΔS	**-21.3820**	**-21.8505**
ΔH(GB)	-23.3877	-26.6900
ΔG_{pred} (GB)	**-2.0057**	**-4.8395**

Table S2. QM/MM analysis of distances (in Å) to the non-heme Fe (II) of the ligands in AlKB complexed with DNA. Distances of the metal coordinating nitrogen of His and oxygen of Asp from the Fe (II) are given.

Name	Equatorial His 131	Axial His 187	ASP 133	2OG (O1)	2OG (O5)	Base (methyl)
QM Cluster minimized	**2.13**	**2.14**	**2.07**	**2.08**	**2.13**	**7.31**
QM/MM Minimized crystal structure (3BIE 1MA, base only)	**2.12**	**2.11**	**2.05**	**2.01**	**2.36**	**4.88**
QM/MM 200ns (6MA, base only)	**2.10**	**2.13**	**2.06**	**2.09**	**2.27**	**7.21**
QM/MM Minimized crystal structure Minimized	2.16	2.18	2.34	2.04	2.29	5.66
QM/MM 200ns	2.18	2.15	2.15	2.06	2.27	7.25
QM/MM 400ns	2.19	2.17	2.11	2.04	2.27	6.69
QM/MM 500ns	2.18	2.18	2.13	2.02	2.27	8.50
QM/MM (Average)	**2.18**	**2.17**	**2.18**	**2.04**	**2.27**	**7.025**
MD (Average)	**2.30**	**2.13**	**2.10**	**2.00**	**2.19**	**7.44**

Table S3. QM/MM analysis of second sphere interactions within the active site of AlkB (PDB: 4NID, 3BIE). Distances are in Angstrom (Å) and were measured between the centres of masses of the guanidino groups of arginine (R204 and R210), the phenol ring of a tyrosine (Y122), the amido groups of asparagines (N206, N120), the indole group of a tryptophan (W178), the carboxyl group of an aspartate (D133) and the C5 carboxyl group of 2OG.

Name	R204 (sc) – C5 (2OG)	Y122 (sc) – C5-(O3) 2OG	N206 (sc)-C5-(O4) 2OG	N120 (sc) O1 2OG	T208 (sc) – O1 2OG	W178 (sc) – D133 (OD2)	R210(sc) – D133 (OD2)
Minimized (3BIE, 1MA base only)	4.05	2.62	3.06	5.16	3.44	5.49	2.85
200ns – (6MA-base only)	4.01	2.59	3.25	3.10	4.00	6.78	5.53
QM/MM Minimized	3.96	2.65	2.89	2.87	2.90	2.90	5.67
QM/MM 200ns	3.99	2.69	3.52	2.90	4.08	5.19	2.87
QM/MM 400ns	3.96	2.74	3.58	2.86	3.95	3.97	3.87
QM/MM 500ns	3.99	2.71	3.52	2.82	3.50	5.45	7.17
QM/MM (Average)	3.97	2.70	3.37	2.86	3.60	4.37	4.89
MD (Average)	4.06	3.02	3.47	2.98	4.18	6.54	7.48

Table S4. QM/MM analysis of distances to the Fe(II) of FTO. Distances (in Å) of the coordinating nitrogen of His and oxygen of Asp from the Fe(II) are given.

Name	Equatorial His 231	Axial His 307	ASP 233	2OG (O1)	2OG (O5)	Base (methyl)
QM Cluster minimized	**2.15**	**2.13**	**2.08**	**2.08**	**2.15**	**5.51**
QM/MM Minimized	2.17	2.13	2.11	2.11	2.23	4.67
* **QM/MM Minimized (EE)**	**2.18**	**2.12**	**2.13**	**2.09**	**2.16**	**4.71**
QM/MM Minimized (Lanl2dz-ECP)	**2.25**	**2.20**	**2.15**	**2.08**	**2.32**	**4.66**
* **QM/MM Minimized (EE) (Lanl2dz-ECP)**	**2.20**	**2.16**	**2.20**	**2.07**	**2.40**	**4.36**
QM/MM 200ns	2.12	2.13	2.13	2.15	2.13	4.65
QM/MM 300ns	2.12	2.16	2.21	2.05	2.19	5.01
QM/MM 700ns	2.09	2.17	2.15	2.13	2.15	5.34
QM/MM 800ns	2.14	2.12	2.14	2.14	2.16	5.82
QM/MM (Average)	**2.13**	**2.14**	**2.15**	**2.12**	**2.17**	**5.1**
MD (Average)	**2.18**	**2.14**	**1.98**	**2.15**	**2.33**	**4.50**

References

1. W. Aik, M. Demetriades, M.K.K. Hamdan, E.A.L. Bagg, K.K. Yeoh, C. Lejeune, Z. Zhang, M.A. McDonoug and C.J. Schofield, *J. Med. Chem.* 2013, **56,** 3680-3688.

2. A. Fiser and A. Sali, *Method enzymology.* 2003, **374,** 461.491.

3. Z. Han, T. Niu, J. Chang, X. Lei, M. Zhao, Q. Wang, W. Cheng, J. Wang, Y. Feng, and J. Chai, *Nature*, 2010, **464,** 1205-1209.

4. R.K. Dennington, and T. Millam, *J. Semichem Inc,* 2009.

5. C. Zhu and C. Yi, *Ang. Chemie Intnal Ed.* 2014, **53,** 3659-3662.

6. C.G. Yang et al., *Nature,* 2008, **452,** 961-965.

7. J.C. Gordon et al., *Nuclei Acids Research,* 2005**, 33,** W368-371.

8. J.M Word, S.C. Lovell, J.S. Richardson and D.C. Richardson, *J. Mol. Biol.* 1999, **285,** 1735-1747.

9. J. Wang, R.M. Wolf, J.W. Caldwell, P.A. Kollman, and D.A. Case, *J. Comput. Chem.,* 2004, **25,** 1157-1174.

10. M.J.T. Frisch, G. W. Schlegel, H. B. Scuseria, G. E. Robb, M. A. Cheeseman, J. R. Scalmani, G. Barone, V. Mennucci, B. Petersson, G. A. Nakatsuji, H. Caricato, M. Li, X. Hratchian, H. P. Izmaylov, A. F. Bloino, J. Zheng, G. Sonnenberg, J. L. Hada, M. Ehara, M. Toyota, K. Fukuda, R. Hasegawa, J. Ishida, M. Nakajima, T. Honda, Y. Kitao, O. Nakai, H. Vreven, T. Montgomery, J. A., Jr. Peralta, J. E. Ogliaro, F. Bearpark, M. Heyd, J. J. Brothers, E. Kudin, K. N. Staroverov, V. N. Kobayashi, R. Normand, J. Raghavachari, K. Rendell, A. Burant, J. C. Iyengar, S. S. Tomasi, J. Cossi, M. Rega, N. Millam, N. J. Klene, M. Knox, J. E. Cross, J. B. Bakken, V. Adamo, C. Jaramillo, J. Gomperts, R. Stratmann, R. E. Yazyev, O. Austin, A. J. Cammi, R. Pomelli, C. Ochterski, J. W. Martin, R. L. Morokuma, K. Zakrzewski, V. G. Voth, G. A. Salvador, P. Dannenberg, J. J. Dapprich, S. Daniels, A. D. Farkas, Ö. Foresman, J. B. Ortiz, J. V. Cioslowski, J. Fox, Gaussian 09. *Gaussian, Inc* Revision D.01(Wallingford CT) 2009.

11. W.D. Cornell, P. Cieplak, C.I. Bayly, and P.A. Kollmann, *J. Am. Chem. Soc.* 1993, **115,** 9620-9631.

12. E. I. Solomon, A. Decker and N. Lehnert, *Proc. Natl. Acad. Sci. U.S.A.,* 2003, **100,** 3589-3594.

13. E.I. Solomon, S. Goudarzi and K.D. Sutherlin, *Biochem.* 2016, **55,** 6363-6374.

14. R.P. Hausinger, *Crit. Rev. Biochem. Mol. Biol.,* 2004, **39,** 21-68.

15. E.I. Solomon, T.C. Brunold, M.I. Davis, J.N. Kemsley, S. Lee, N. Lehnert, F. Neese, A.J Skulan, Y. Yang and J. Zhou, *Chem. Rev.* 2000, **100,** 235-350.

16. R.H. Holm and E.I. Solomon, *Chem. Rev.* 2014, **114,** 4039-4040.

17. B. Wang, Z. Cao, D.A. Sharon, and S. Shaik, *ACS Catalysis* 2015, **5,** 7077-7090.

18. E.G. Pavel *et al.* *J. Am. Chem. Soc.* 1998, **120,** 743-753.

19. J. Zhou *et al., J. Am. Chem. Soc.* 2001, **123,** 7388-7398.

20. J. Zhou, M. Gunsior, B.O. Bachmann, C.A. Townsend, and E.I. Solomon, *J. Am. Chem. Soc.* 1998, **120,** 13539-13540.

21. P. Li and K.M. Merz, *J. Chem. Info and Model* 2016, **56,** 599-604.

22. A. Pabis, I. Geronimo, D.M. York, and P. Paneth, *J. Chem. Theo. Comput* 2014, **10,** 2246-2254.

23. R. Salomon-Ferrer, A.W. Götz, D. Poole, S. Le Grand, amd R.C. Walker, *J. chem. Theo comput* 2013, **9,** 3878-3888.

24. D. Case, V. Babin, J. Berryman, R. Betz, Q. Cai, D. Cerutti, T. Cheatham III, T. Darden, R. Duke and H. Gohlke, Amber, 2014, **14,** 29-31.

25. J.A. Maier*, et al. J. chem. Theo. comput* 2015, **11,** 3696-3713.

26. W.L. Jorgensen, J. Chandrasekhar, J.D. Madura, R.W. Impey, and M.L. Klein, *J. chem. Phys.* 1983, **79,** 926-935.

27. T. Darden, D, York, and L. Pedersen, *J. chem. Phys.* 1993, **98,** 10089-10092.

28. R. Davidchack, R. Handel and M.V. Tretyakov, *J. Chem. Phys.* 2009, **130,** 234101.

29. J.P. Ryckaert, G. Ciccotti, and H.J. Berendsen, *J. Comput. Phys.* 1977, **23,** 327-341.

30. H.J. Berendsen, J.V. Postma, W.F. van Gunsteren, A. DiNola, and J. Haak, *J. chem. Phys.* 1984, **81,** 3684-3690.

31. D.R. Roe and T.E. Cheatham III, *J. chem. theo comput.* 2013 **9,** 3084-3095.

32. W. Humphrey, A. Dalke and K. Schulten, *J. Mol. Graph.* 1996, **14,** 27-38.

33. E.F. Pettersen*, et al.*, *J Comput Chem* 2004, **25,** 1605-1612.

34. B.J. Grant, A.P. Rodrigues, K.M. ElSawy, J.A. McCammon, and L.S. Caves, *Bioinformatics (Oxford, England)* 2003, **22,** 2695-2696.

35. W. Singh, G. Fields, C. Christov and T.G. Karabencheva-Christova, *RSC Advances 2016*.

36. G. Schenk, M.Y. Pau, and E.I. Solomon, *J Am Chem Soc* 2004, 126, 505-515.

37. V. Barone and M. Cossi, *J. Phys. Chem A* 1998, **102,** 1995-2001.

38. P. Walter, J. Metzger, C. Thiel, and V. Helms, *PloS one* 2013, **8,** e58583.

39. P.E. Siegbahn and T. Borowski, *Acct. chem. Resear.* 2006, **39,** 729-738.

40. P.E. Siegbahn and F. Himo *Wiley Interdisciplinary Reviews: Computational Molecular Science* 2011, **1,** 323-336.

41. P. Tao and H.B. Schlegel, *J. Comput. Chem.* 2010, **31,** 2363-2369.

42. F. Maseras and K. Morokuma, *J. Comput. Chem.* 1995, **16,** 1170-1179.

43. M. Svensson, S. Humbel, and K. Morokuma, *J. chem. phys.* 1996, **105,** 3654-3661.

44. S. Dapprich, I. Komáromi, K.S. Byun, K. Morokuma and M.J. Frisch, *J. Mol. Struc: THEOCHEM* 1999, **461,** 1-21.

45. T. Vreven, K. Morokuma, O. Farkas, H.B. Schlegel, and M.J. Frisch, *J Comput Chem* 2003, **24,** 760-769.

46. T. Vreven and K. Morokuma, *J. Comput. Chem.* 2000, **21,** 1419-1432.

47. P. Li, B.P. Roberts, D.K. Chakravorty, and K.M. Merz Jr., *J Chem Theory Comput* 2013, **9,** 2733-2748.

48. H.M. Senn and W. Thiel, *Angewandte Chemie International Edition* 2009, **48,** 1198-1229.

49. P.A. Kollman, I. Massova, C. Reyes, B. Kuhn, S. Huo, L. Chong, M. Lee, T. Lee, Y. Duan, W. Wang, O. Donini, P. Cieplak, J. Srinivasan, D.A. Case, and T.E. 3rd Cheatham, *Acc. Chem. Res* 2000, **33,** 889-897.

50. I. Massova and P.A. Kollman, *Perspectives in drug discovery and design* 2000, **18,** 113-135.

51. V. Tsui and D.A. Case, *Biopolymers* 2000, **56,** 275-291.

52. P.J. Holland, T. Hollis, *PLoS One.* 2010, **5,** e8680.

53. Z. Han, T. Niu, J. Chang, X. Lei, M. Zhao, Q. Wang, W. Cheng, J. Wang, Y. Feng, and J. Chai, *Nature* 2010, **464,** 1205-1209.

54. G. Jia, Y. Fu, X. Zhao, Q. Dai, G. Zheng, Y. Yang, C. Yi, T. Lindahl, T. Pan, Y.G. Yang and C. He, *Nat. Chem. Biol.,* 2011, **7,** 885-887.

55.   L. Rohena, M. Lawson, E. Guzman, M. Ganapathi, M.T. Cho, E. Haverfield and K. Anyane-Yeboa, *Am. J. Med. Genet A.,* 2016, **170,** 1023-1028.

56.   H. Daoud, D. Zhang, F. McMurray, A. Yu, S.M. Luco, J. Vanstone, O. Jarinova, N. Caron, J. Wickens, S. Shishodia, H. Choi, M.A. McDonough, C.J Schofield, M.E. Harper, D.A. Dyment and C.M. Armour, *J. Med. Genet.,* 2016, **53,** 200-207.