Supporting Information

METHODS

Starting Structures. The wild type DfHase crystal structure (PDB-id 1FRF) has been used as starting structure for the DfHase simulation, while comparative homology modeling was used to build the starting structure for AaHase. The amino acid sequences of the two soluble AaHase domains have been taken from the GenBank id AAC06862.1 (small subunit) and AAC06861.1 (large subunit). MODELLER v9.10[1,2] was used to build the starting structure using the following templates: Desulfovibrio desulfuricans (1E3D)[3], Desulfovibrio vulgaris Miyazaki (1WUI)[4], Desulfovibrio gigas (1FRV)[5], Desulfovibrio fructosovorans (1YQW)[6], Hydrogenovibrio marinus (3AYX)[7], Ralstonia eutropha (3RGW)[8] and Allochromatium vinosum (3MYR)[9]. The structures were superimposed using Chimera[10,11] and used as input for MODELLER. Regarding the small subunit, only the residues for the soluble domain (i.e. from P47 to G316), were retained. The metallic centers were obtained from the oxygen tolerant hydrogenase of Hydrogenovibrio marinus[7] and were treated as rigid blocks during the modeling procedure. The loops originating from gaps in the alignment were energy-optimized in order to obtain a refined, energetically favorable structure according to the dope-score of the Modeller v9.10 software. The secondary structure regularity and the degree of steric clashes in the model were evaluated using the MolProbity (http://molprobity.biochem.duke.edu) web server[12] at two different stages namely on the starting model and after the equilibration.

Metal centre parametrization. In this paper we attempt to study the hydrogenase enzymes in the oxidation state prior to H_2 processing. According to Niu and coworkers[13], the active site has been assumed to be in the Ni-SIa state (total charge of -2) while the FeS clusters were treated as oxidized. Parameters for the [4Fe4S] proximal cluster in DfHase were obtained from[14], while the medial clusters were parametrized according to ref.[15]. Charges for

the remaining metallic centers (i.e. active site, the distal and medial clusters and the proximal cluster of AaHase) were obtained through the R.E.D. Server[16] using the RESP-X1 charge model and are reported in Table 1 of this Supplementary Information. To reduce the complexity of the calculations, only the side chains have been retained. The Cβ has been treated as a methyl group and the neutrality has been imposed to these four atoms. The metallic centers were maintained rigid during the simulations through an elastic network (spring force constant 200 kJ) between the metallic atoms and the carbons of the side chains. It is worth stressing that the primary aim of this study was the global characterization of the protein motion that is not supposed to be influenced by the approximations in the parameterization of the metal centers. During the writing of this paper the metallic centers of oxygen-tolerant hydrogenases were more accurately parametrized[17], providing a starting point for future studies. Among the possible states of the AaHase proximal cluster (reduced, oxidized and superoxidized), the oxidized state has been used for two reasons. The first is in order to assign the same charge (-2) in both simulations. The second is that deriving the parameters is less ambiguous because in this state the N of $C26^S$ is protonated.

System parametrization. The proteins were solvated in a cubic box of water and afterwards Na^+ and Cl^- ions were added at a concentration of 0.150 M. The systems were represented using CHARM27 with CMAP corrections[18] and the standard TIP3P water model as implemented in the GROMACS software[19] was used. The protonation state has been chosen according to the most probable one at pH 7.

Simulation set-up. The two systems were simulated using GROMACS 4.5.3[20]. The systems were initially energy minimized for 2000 steps, then equilibrated by simulating for 1 ns in the NVT ensemble and subsequently for 1 ns in the NPT ensemble, prior to starting production runs of 2 microsecond duration. The temperature of protein and solvent (water and ions) was separately regulated using the Velocity Rescaling[21] method with a reference

temperature of 300K and a coupling constant of 1ps for the two groups. The Parinello-Rahman[22,23] algorithm was adopted to maintain the pressure at 1 atm. The SETTLE[24] algorithm was employed to maintain the water rigidity and LINCS[25,26] to constrain the covalent bonds involving hydrogen atoms. The two optimized and relaxed systems were simulated for 2 microseconds, with a 2.0 fs time step, in periodic boundary conditions. Van der Waals interactions were switched off between 1.0 to 1.2 nm, updating the neighbor pair lists every 10 steps while the long range electrostatic interactions were evaluated through the particle-mesh Ewald method[27] using a cut-off of 1.2 nm.

Analysis. Unless otherwise stated, the trajectories were skipped every 5 ps and analyzed with GROMACS 4.5.5 analysis tools. All the plots were depicted through Matplotlib[28], while structures have been depicted using VMD 1.9.1[29] and Chimera[11]. Since the RMSD reaches a stable plateau within 200 ns, the analysis has been carried out on the last 1.8 μs, considering the initial 200 ns as equilibration period.

The electrostatic potentials have been calculated using the conformation at time 2 μs using the APBS 1.3 software[30] assigning the charges used in the simulation.

The dipole moment has been calculated through the Dipole Watcher VMD plugin and afterwards the vector components have been used to obtain the θ and φ angles. θ defines the dipole moment inclination with respect to the Z axis, which goes through the distal FeS cluster and the [NiFe] active site (see Fig. 1b), while the φ angle depends on the dipole moment orientation with respect to [NiFe]-Mg^{2+} ion axis.

Modeling of the AaHase trans-membrane helix. The recently published E.coli-cytochrome structure (4GD3)[31] has been used as template to build a homology model of the full length AaHase (Fig.4a). To assess the influence of the trans-membrane helix on the overall dipole moment, two additional helix conformations have been generated using the sculpting

module implemented in Pymol. In the former conformation, the helix is oriented toward the large subunit (Fig.4b) while in the latter it is oriented toward the small subunit (Fig.4c).

Quality assessement. The quality of the homology model has been assessed through MolProbity[12]. This software provides a score resuming the model quality in terms of steric clashes and deviation of the backbone dihedral angles from any of the Ramachandran regions describing α-helices and ß-sheets. The analysis performed on the starting model displays that 96.6% of the residues belong to the favored region and 99.5% are in the allowed region; at the end of the thermalization (i.e. 0 ns) several residues on the border between helices and loop became unstructured so the percentage of the residues belonging to favored and allowed regions slightly decreased to 92.9% and 98.5% respectively. The residues of the long loop ranging from Tyr142 to Lys161 from the large subunit retained the unstructured conformation and compose, together with the other loops, the residues in the forbidden region.

The final structure of AaHase can be downloaded as a pdb file at the following address: http://www.lojou.fr/biopac/?page=./common/bibliotheque.php&display=0

References:

[1] A. Fiser, A. Sali, in Methods Enzymol., Elsevier, **2003**, pp. 461–491.

[2] A. Fiser, R. K. Do, A. Sali, Protein Sci. Publ. Protein Soc. **2000**, 9, 1753–1773.

[3] P. M. Matias, C. M. Soares, L. M. Saraiva, R. Coelho, J. Morais, J. Le Gall, M. A. Carrondo, J. Biol. Inorg. Chem. JBIC Publ. Soc. Biol. Inorg. Chem. **2001**, 6, 63–81.

[4] H. Ogata, S. Hirota, A. Nakahara, H. Komori, N. Shibata, T. Kato, K. Kano, Y. Higuchi, Structure **2005**, 13, 1635–1642.

[5] A. Volbeda, M. H. Charon, C. Piras, E. C. Hatchikian, M. Frey, J. C. Fontecilla-Camps, Nature **1995**, 373, 580–587.

[6] A. Volbeda, L. Martin, C. Cavazza, M. Matho, B. Faber, W. Roseboom, S. Albracht, E. Garcin, M. Rousset, J. Fontecilla-Camps, JBIC J. Biol. Inorg. Chem. **2005**, 10, 239–249.

[7] Y. Shomura, K.-S. S. Yoon, H. Nishihara, Y. Higuchi, Nature **2011**, 479, 253–256.

[8] J. Fritsch, P. Scheerer, S. Frielingsdorf, S. Kroschinsky, B. Friedrich, O. Lenz, C. M.

Spahn, Nature **2011**, 479, 249–252.

[9] H. Ogata, P. Kellers, W. Lubitz, J. Mol. Biol. **2010**, 402, 428–444.

[10] G. S. Couch, D. K. Hendrix, T. E. Ferrin, Nucleic Acids Res. **2006**, 34, e29.

[11] E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, T. E. Ferrin, J. Comput. Chem. **2004**, 25, 1605–1612.

[12] I. W. Davis, A. Leaver-Fay, V. B. Chen, J. N. Block, G. J. Kapral, X. Wang, L. W. Murray, W. B. Arendall, J. Snoeyink, J. S. Richardson, et al., Nucleic Acids Res. **2007**, 35, W375–W383.

[13] S. Niu, L. M. Thomson, M. B. Hall, J. Am. Chem. Soc. **1999**, 121, 4000–4007.

[14] C. H. Chang, K. Kim, J. Chem. Theory Comput. **2009**, 5, 1137–1145.

[15] M. Meuwly, M. Karplus, Faraday Discuss. **2003**, 124.

[16] F.-Y. Y. Dupradeau, A. Pigache, T. Zaffran, C. Savineau, R. Lelong, N. Grivel, D. Lelong, W. Rosanski, P. Cieplak, Phys. Chem. Chem. Phys. PCCP **2010**, 12, 7821–7839.

[17] D. M. A. Smith, Y. Xiong, T. P. Straatsma, K. M. Rosso, T. C. Squier, J. Chem. Theory Comput. **2012**, 8, 2103–2114.

[18] A. D. MacKerell, M. Feig, C. L. Brooks, J. Am. Chem. Soc. **2003**, 126, 698–699.

[19] P. Bjelkmar, P. Larsson, M. A. Cuendet, B. Hess, E. Lindahl, J. Chem. Theory Comput. **2010**, 6, 459–466.

[20] B. Hess, C. Kutzner, D. van der Spoel, E. Lindahl, J. Chem. Theory Comput. **2008**, 4, 435–447.

[21] G. Bussi, M. Parrinello, Phys. Rev. E **2007**, 75, 056707+.

[22] S. Nosé, M. L. Klein, Mol. Phys. **1983**, 50, 1055–1076.

[23] M. Parrinello, A. Rahman, J. Appl. Phys. **1981**, 52, 7182–7190.

[24] S. Miyamoto, P. A. Kollman, J Comput Chem **1992**, 13, 952–962.

[25] B. Hess, J. Chem. Theory Comput. **2007**, 4, 116–122.

[26] B. Hess, H. Bekker, H. J. C. Berendsen, J. G. E. M. Fraaije, J Comput Chem **1997**, 18, 1463–1472.

[27] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, L. G. Pedersen, J. Chem. Phys. **1995**, 103, 8577–8593.

[28] J. D. Hunter, Comput. Sci. Eng. **2007**, 9, 90–95.

[29] W. Humphrey, A. Dalke, K. Schulten, J. Mol. Graph. **1996**, 14.

[30] N. A. Baker, D. Sept, S. Joseph, M. J. Holst, J. A. McCammon, Proc. Natl. Acad. Sci. **2001**, 98, 10037–10041.

[31] A. Volbeda, C. Darnault, A. Parkin, F. Sargent, F. A. Armstrong, J. C. Fontecilla-Camps, Struct. Lond. Engl. 1993 **2013**, 21, 184–190.

[32] X. Luo, M. Brugna, P. Tron-Infossi, M. T. T. Giudici-Orticoni, E. Lojou, J. Biol. Inorg. Chem. JBIC Publ. Soc. Biol. Inorg. Chem. **2009**, 14, 1275–1288.

FIGURE LEGENDS

Supplementary figure 1. Parameterized metallic centers. The atomic name is reported only for the inorganic parts in order to allow the comparison with table 1 of Supplementary Information. Panel A represents the active site, panel B the proximal FeS cluster while panel C reports the distal FeS cluster.

Supplementary figure 2. Surface characteristics for AaHase (left) and DfHase (right) from six different views (panel I to VI). The top rows recall the orientation of the protein, the middle one shows the electrostatic potential (at 2 kT) as identified in Table 2 of Supplementary Information.

Supplementary figure 3. Dipole moment strength as a function of time for AaHase (black line) and DfHase (red line).

Supplementary figure 4. Dipole moment strength as a function of time for the small (black line) and large (red line) subunit for AaHase (Panel A) and DfHase (Panel B).

Supplementary figure 5. Angle between the dipole moment contributions of the large and small subunits as a function of time in AaHase (black line) and DfHAse (red line)

Supplementary figure 6. RMSF for each residue of the small (Panel A) and large (Panel B) subunit for AaHAse (black) and DfHase (red). The residues have been aligned based on their similarity. The Cysteines coordinating the active site, proximal, medial and distal cluster have been denoted as AS, P, M and D respectively. The blue rectangle represents the long loop. The horizontal line indicates the threshold value chosen for the selection of the fluctuating residues (0.1 nm).

Center	Atom	Charge
Active Site	Ni	0.40
	Fe	0.23
	C	0.22
	O	-0.294
	C	0.11
	N	-0.61
Cys65/610	CB	-0.03
	HB1	0.08
	HB2	0.08
	SG	-0.39
Cys62/607t	CB	-0.05
	HB1	0.051
	HB2	0.051
	SG	-0.57
4Fe3S Proximal FeS cluster	Fe1	0.4349
	Fe2	0.3710
	Fe3	0.5092
	Fe4	0.3965
	S1	-0.3661
	S2	-0.3728
	S3	-0.3076
Cys59	CB	-0.0751
	HB1	0.065
	HB2	0.065
	SG	-0.5387
Cys61	CB	-0.1877
	HB1	0.126
	HB2	0.126
	SG	-0.2863
Cys62	CB	-0.0884
	HB1	0.0693
	HB2	0.0693
	SG	-0.6067
Cys191	CB	0.0656
	HB1	-0.0044
	HB2	-0.0044
	SG	-0.5784
	CB	-0.0247

Cys162	HB1	0.0448
	HB2	0.0448
	SG	-0.5055
Distal FeS cluster	Fe1	0.0505
	Fe2	0.3725
	Fe3	0.3102
	Fe4	0.3965
	S1	-0.3057
	S2	-0.2484
	S3	-0.2352
	S4	-0.2266
Cys232	CB	-0.0455
	HB1	0.0645
	HB2	0.0645
	SG	-0.5437
Cys257	CB	0.0187
	HB1	0.0453
	HB2	0.0453
	SG	-0.5820
Cys263	CB	-0.0176
	HB1	0.056
	HB2	0.056
	SG	-0.5302
Hys229	CB	-0.1479
	HB1	0.099
	HB2	0.099
	CG	0.0620
	ND1	-0.0286
	CE1	-0.0292
	HE1	0.1507
	NE2	-0.2535
	HE2	0.3341
	CD2	-0.1953
	HD2	0.1646

Table 1: Charges used in this work: The atom names of the amino acids follow the CHARMM convention, while for the inorganic parts the name agrees with the Fig.SI1 of Supporting Information.

	Lobe	L	S
AaHase	A	D7 E19 D31 D410 E411 E580 E581 D585 E596	E96 E99 E125 D126 E139 E143
	B	D116 E122 D127 E140 E147 D151 D195 E205	E85 E108
	C	D219 E223 D252 D253 E261 E267 D268 E275 D278 D454 E455	
	D	D82 E359 E362 D512 E519 D522	D245 D248 D249
	E	**K160 K161 K164 R171 K175 K176 K178 K179**	
	F		**K187 R210 K221**
	G		**K178 K181**
	H	**K531 K550**	**K167 K260 R308**
	I	**K330 K331 K351 K358 K444**	
DfHase	A	D298 D299 E445 D453 E461	E135 D252 D255 E262 D170 E171 E181 D185 E194 E197 E250 D203 E262
	B	D19 D363 D364 D366	E54 E62 E65 D68 E96
	C	D149 E264 D405	
	D	K135 K140 K143 K189	
	E	K245 K256 K444 K452	

Table 2: Residues lining the electrostatic lobes in Fig.SI2 of Supporting Information.

Figure SI1

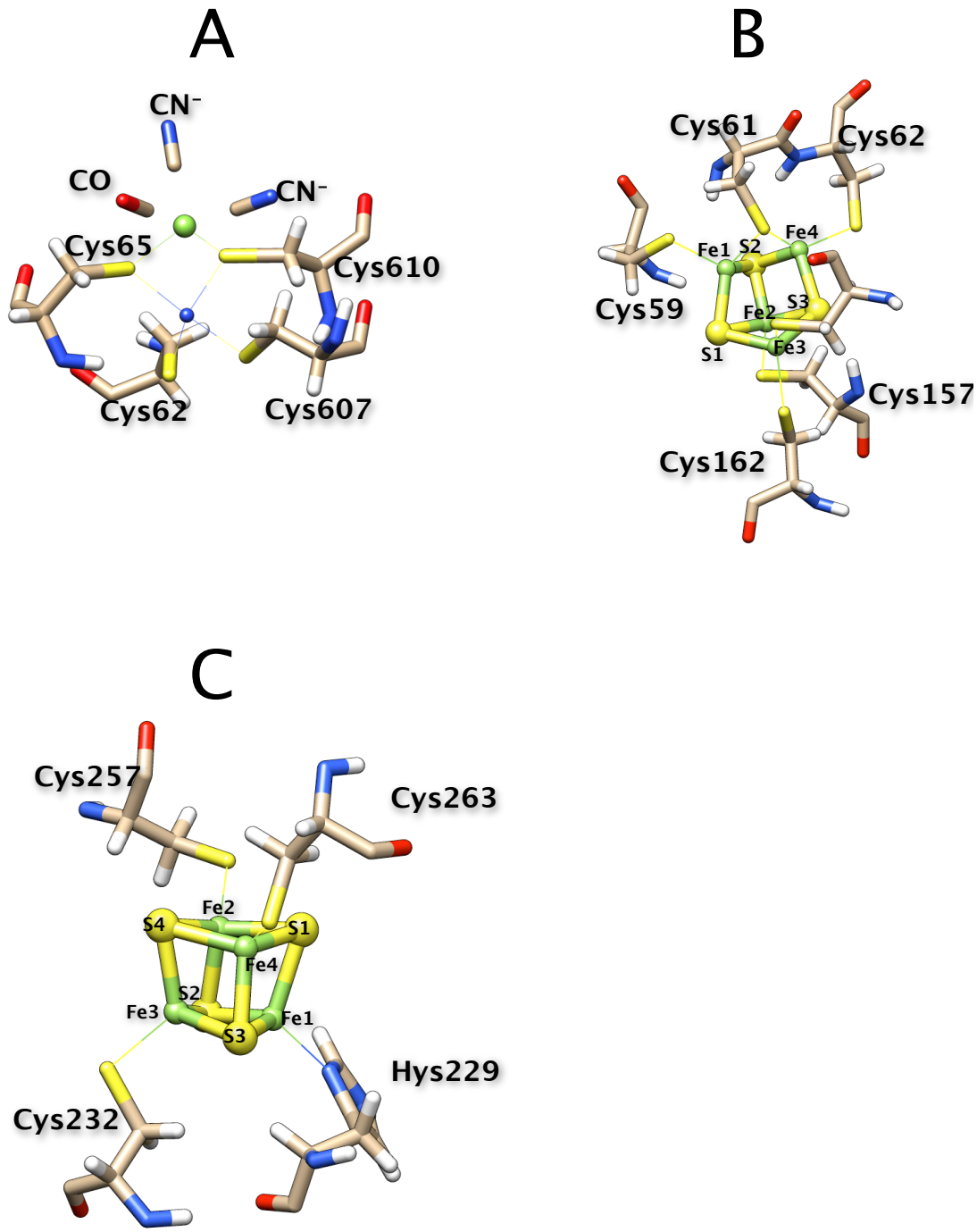
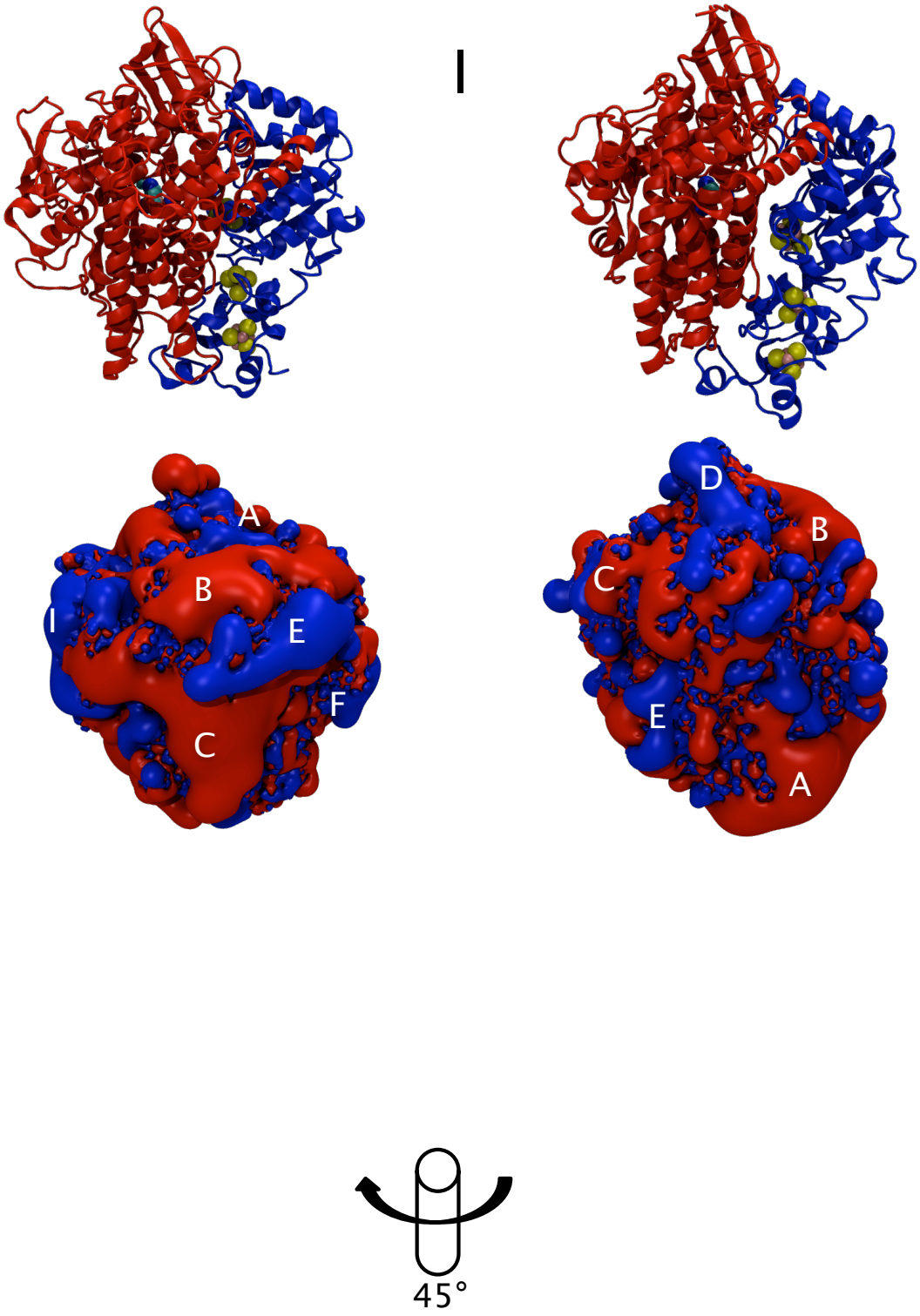
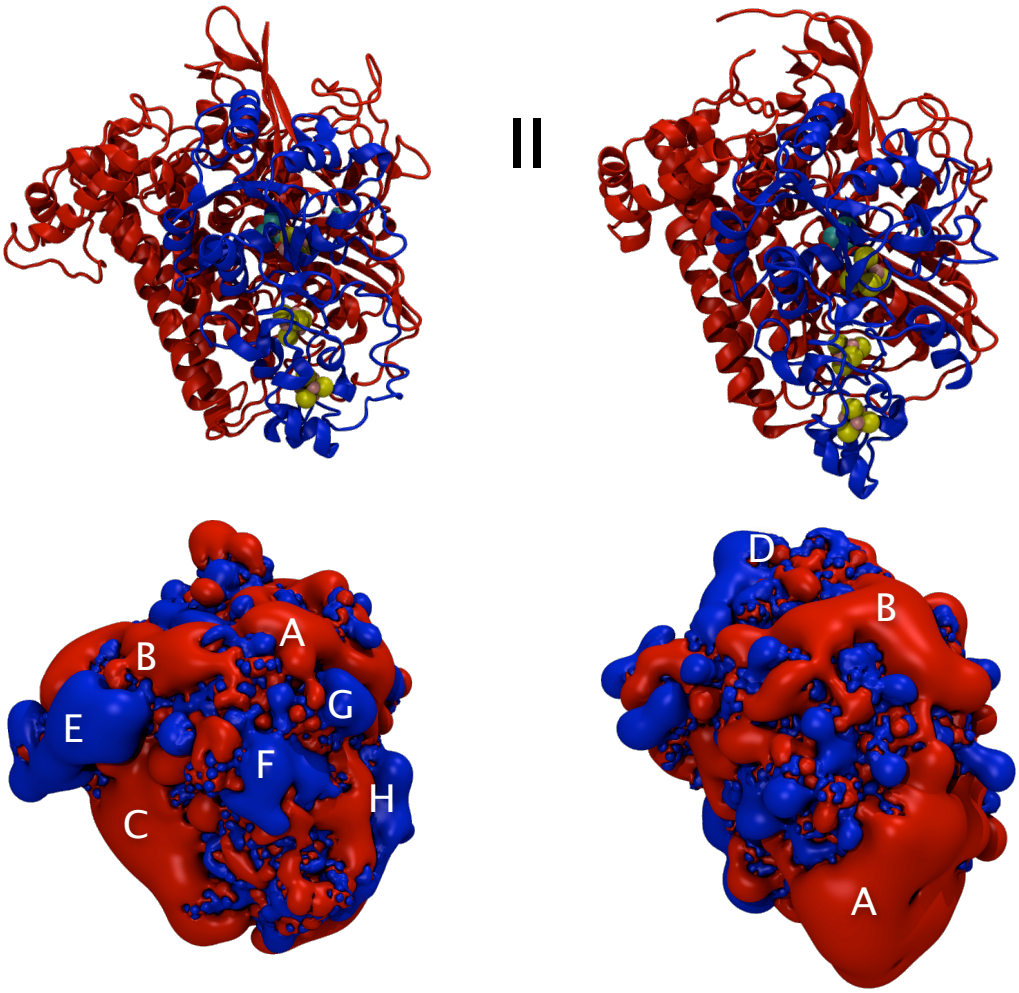
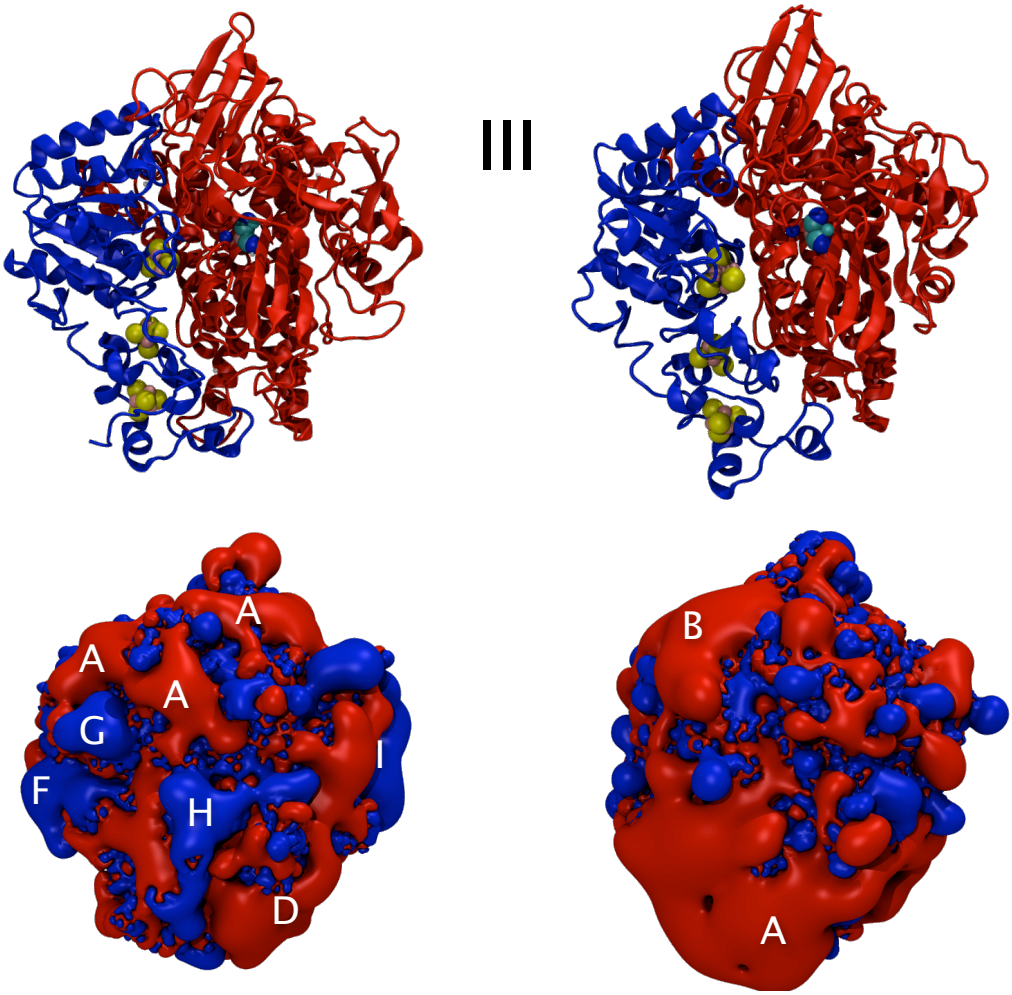
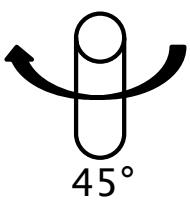

A
CN⁻
CO
CN⁻
Cys65
Cys610
Cys62
Cys607
B
Cys61
Cys62
Fe4
Fe1
S2
Cys59
Fe2
S3
S1
Fe3
Cys157
Cys162
C
Cys257
Cys263
Fe2
S4
S1
Fe4
S2
Fe3
Fe1
S3
Hys229
Cys232

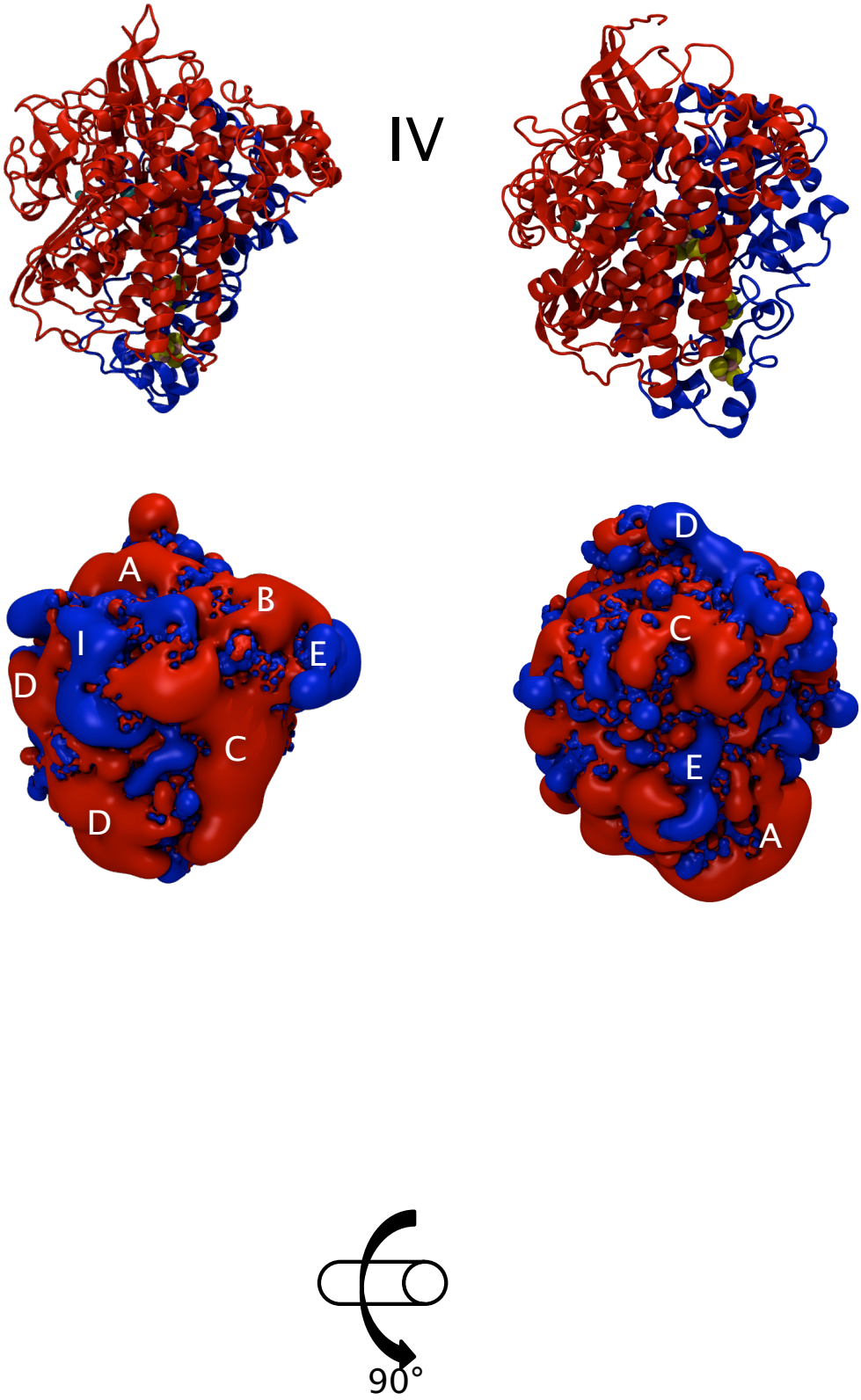
Figure SI2

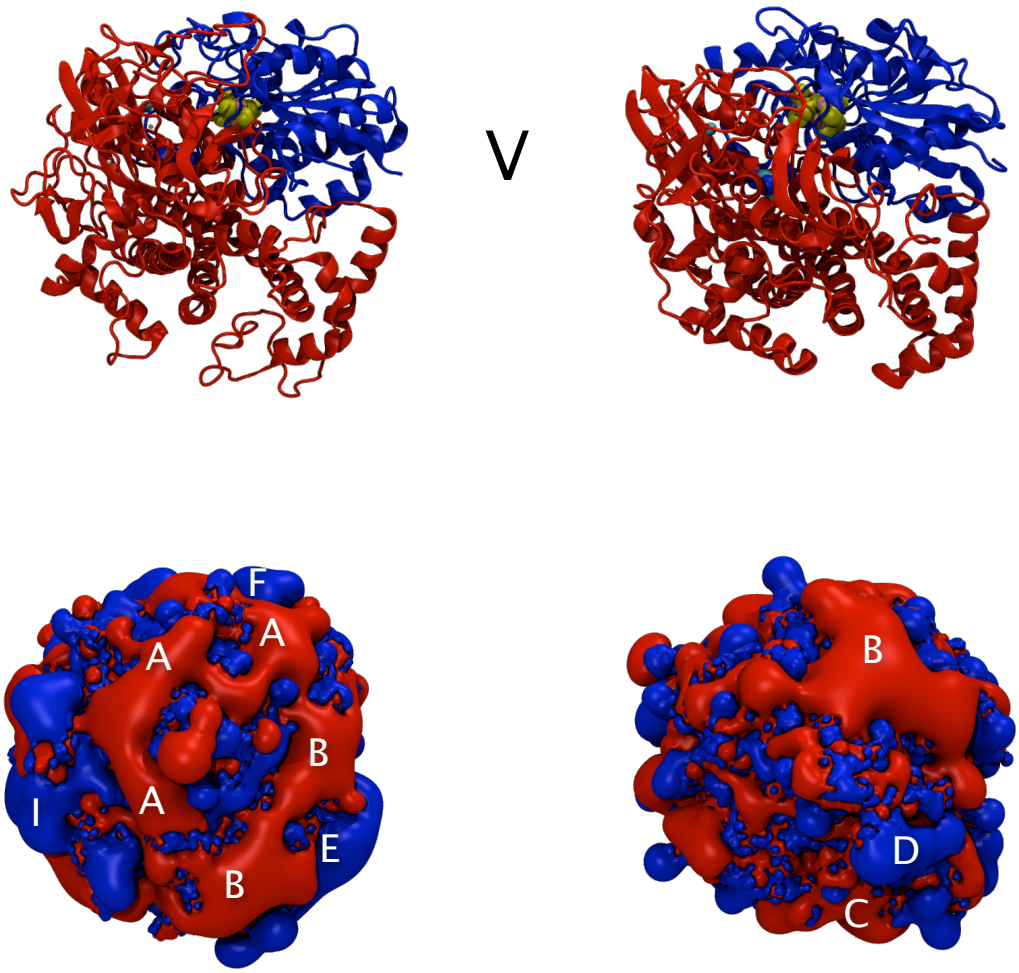


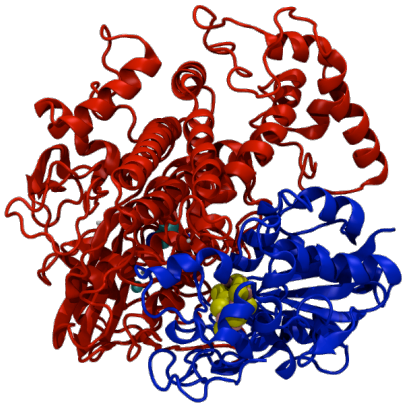

II
A
B
G
E
F
H
C
D
B
A
45°


A
A
A
G
I
F
H
D
B
A

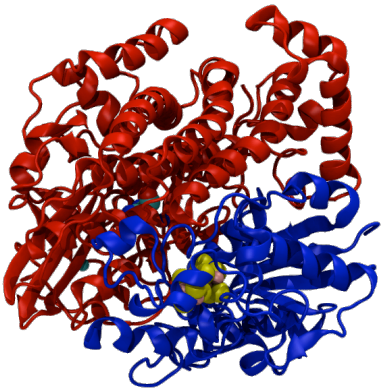

45°

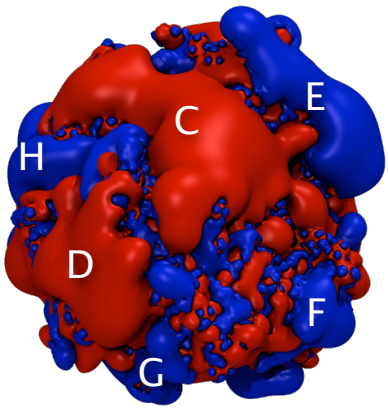

IV
A
B
I
E
D
C
D
D
C
E
A
90°


V
F
A
A
B
I
A
E
B
B
D
C
180°



VI





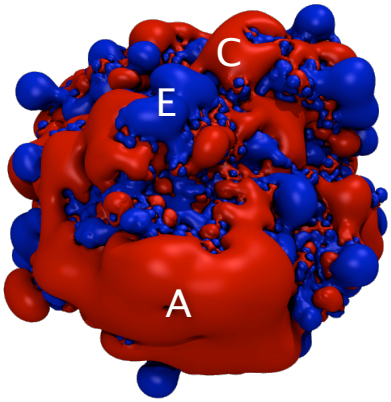


Figure SI3

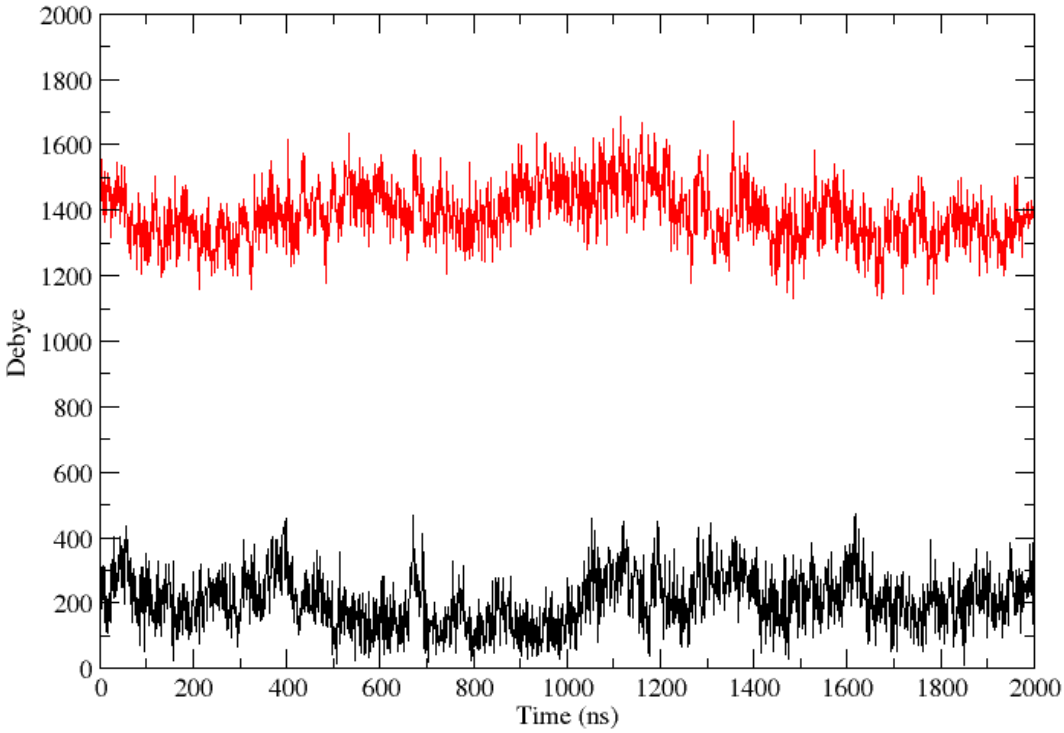
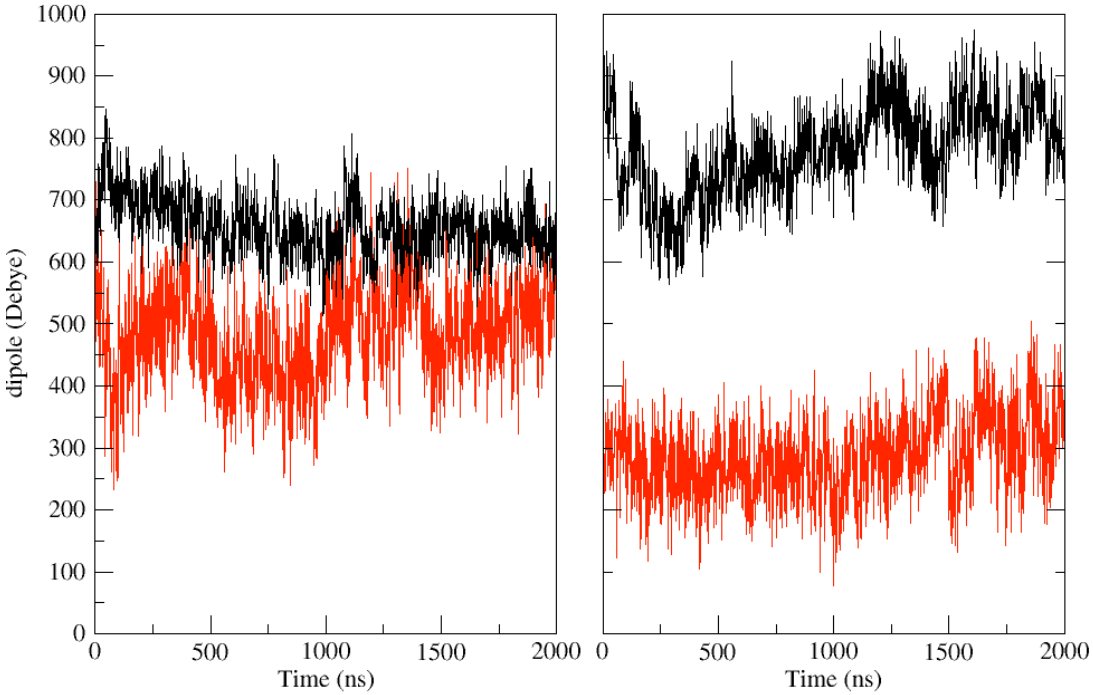


Figure SI4



A B

Fig. SI5

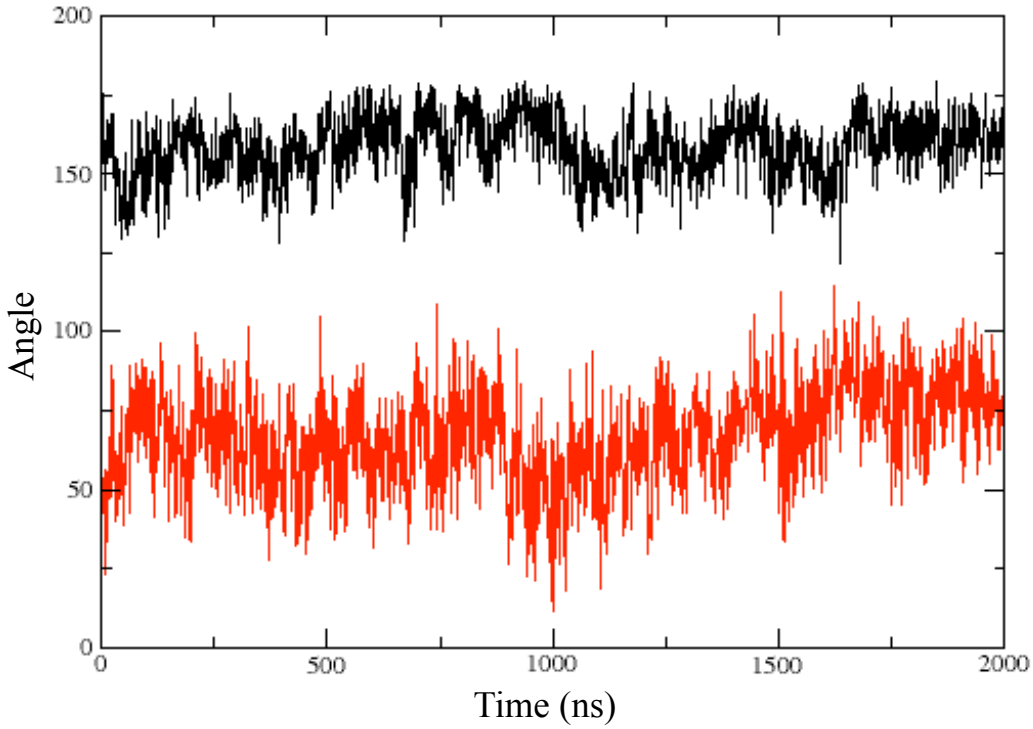
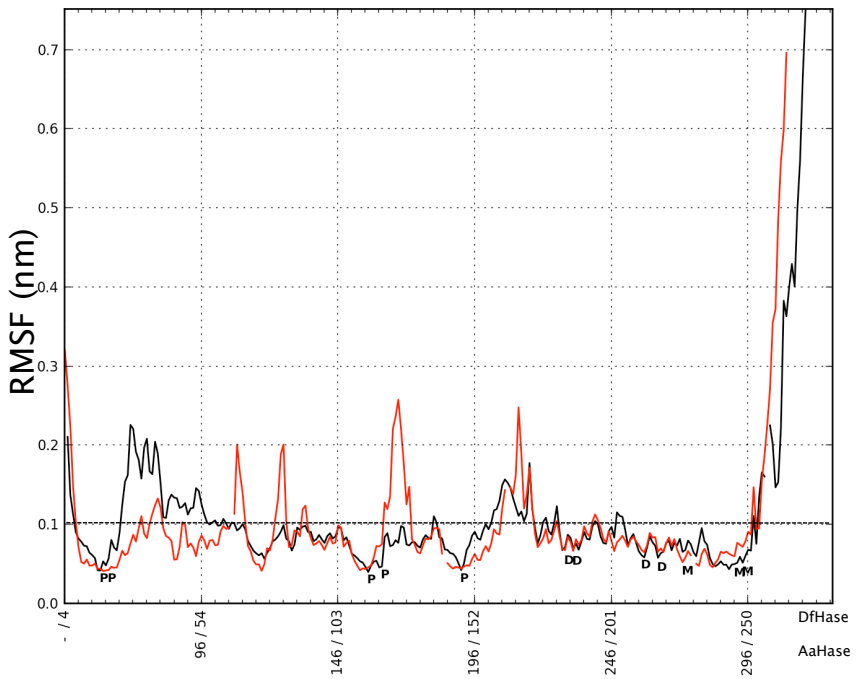


Figure SI6

A



B