Electronic Supplementary Information

A QM/MM Study on the Origin of Retro-Aldolase Activity of a Catalytic Antibody.

Daria De Raffele, Sergio Martí,* Vicent Moliner.*

Departament de Química Física i Analítica; Universitat Jaume I,

12071 Castellón (Spain)

*correspondence should be addressed:

S. Martí: smarti@uji.es

V. Moliner: moliner@uji.es

COMPUTATIONAL METHODS

The starting structure for the construction of our CA model was the X-ray crystal structure of the aldolase antibody 33F12 Fab' (PDB code 3FO9). The structure was manually modified using the VMD[1] program in order to obtain a favorable conformation. The hapten 1,3-diketone was substituted by (R)-methodol (see Fig. 2). Hydrogen atoms were added to the X-ray structure at physiological pH, according to the pK_a values of titratable residues calculated with the empirical PROPKA 3.1 program of Jensen et al.[2,3] A total of three counterions (Cl^-) were placed into optimal electrostatic positions around the system, to obtain electroneutrality. Finally, the system was placed in a $121.72 \times 74.83 \times 61.86$ Å^3 orthorhombic box of water molecules. A cutoff for nonbonded interactions was applied using the switching-force scheme, with a range of radius from 14.5 Å to 16 Å, as well as periodic boundary conditions. The system was relaxed employing 30 ns of classical molecular dynamics (MD) at 300 K, using the NVT ensemble and the Langevin−Verlet[4] integrator, with a time step of 1 fs. This MD simulation was performed with the parallel code NAMD,[5] using the CHARMM36[6–8] force field for the protein and substrate, while the water molecules were described with the TIP3P[7] force field. The time evolution analysis of the root-mean-square deviation for the backbone atoms of the protein confirmed that the system became equilibrated during the MD simulation. Once the molecular model was set up and equilibrated, potential energy surfaces (PESs) for every proposed chemical step of the reaction were explored, using the fDYNAMO library.[9] The structures generated on the PESs were then used as starting points to compute the corresponding free energy surfaces (FESs), in terms of potentials of mean force (PMF). This technique requires the use of a limited number of distinguished reaction coordinates in order to recover the FES, using the combination of the weighted histogram analysis (WHAM)[10] and the umbrella sampling (US) methods,[11,12] as implemented in the fDYNAMO library.[9] Different reaction mechanisms have been studied, until the most energetically favorable was found which consisted of seven reaction steps. The first and second steps were studied by means of a 2D-PMF, based on 1053 series of simulation windows, using the antisymmetric combination of the bond-breaking and the bond-forming distances describing the hydrogen transfer from Tyr36 to the O1 of the methodol, $\xi_1 = d(O_{Y36}-H_{Y36}) - d(H_{Y36}-O_1)$, and the bond-forming distance $\xi_2 = d(N_{K93}-C_2)$, as distinguished reaction coordinates. The following step was explored with a 1D-PMF using just the antisymmetric combination of the bond-breaking and the bond-forming distances $\xi_1 = d(N_{K93}-H_{K93}) - (H_{K93}-O_{Y36})$, with a total of 39 simulation windows. A 2D-PMF was traced for the fourth step, built from 1521 simulation windows, using as coordinates the antisymmetric combination of the bond-breaking and the bond-forming distance $\xi_1 = d(O_2-H) - d(H-O_{Y36})$, and the antisymmetric combination of the bond-breaking and the bond-forming distance $\xi_2 = d(O_{Y36}-H_{K93}) - d(H_{K93}-O_1)$.

Throughout the course of the fourth reaction step the 6-MNA is released from the active site. The successive fifth and sixth steps were studied with a 2D-PMF, using the antisymmetric combination of the bond-breaking and the bond-forming distances $\xi_1 = d(O_{Y36}-H) - d(H-C_3)$ and the bond-forming distance $\xi_2 = d(O_1-C_2)$, from a total of 819 simulation windows. The final production of acetone from I_6 was carried out by computing a 2D-PMF, using 1521 simulation windows with the combination of $\xi_1 = d(O_{Y36}-H_{K93}) - d(H_{K93}-O_{Y36})$ and $\xi_2 = d(O_1-H_{Y36}) - d(H_{Y36}-O_{Y36})$ as reaction coordinates.

The atoms being described quantum mechanically in the first six steps comprised the substrate and the side chains of Lys93H and Tyr36L residues (63 atoms in total, see Fig. 3). Once the 6-MNA was released, the QM region was reduced to 38 atoms, which included the enamine, the side chains of Lys93H and Tyr36L, and the generated water molecule. The rest of the protein−solvent system (59175 atoms) was represented classically by means of the OPLS-AA[13–15] and the TIP3P[16] water-model force fields. We introduced two quantum link atoms[17] to saturate the valence of the QM-MM frontiers, as depicted in Fig. 3. Because of the large number of gradient vectors that must be evaluated during the QM/MM FESs calculations, the QM subset of atoms is usually represented by a semi-empirical Hamiltonian, being the AM1[18] the one selected in present work. Then, to reduce the errors associated with the semiempirical quantum level of theory employed in our simulations, a cubic spline was introduced to interpolate a correction term. This correction is defined for any value of the reaction coordinates ξ_1 (and ξ_2 in the case of two-dimensional PMFs),[19-21] based on the seminal work of Truhlar and coworkers.[22–24] In this way, a continuous energy correction function is added to the potential energy (and its gradient) as:

$$E_{corr} = S\left[\Delta E_{LL}^{HL}\left(\xi_1, \xi_2\right)\right]$$ (1)

where S is a one- or two-dimensional spline function providing a parametric energy difference between high-level (HL) and low-level (LL) perturbed Hamiltonians of the QM subsystem, and which is derived from QM/MM single-point energy calculations. In the present paper, the LL was the AM1 method; meanwhile, the hybrid M06-2X functional with the 6-31+G(d,p) basis set was selected as the HL method.[25] These calculations were performed using the Gaussian09 family of programs.[26] In order to carry out a detailed analysis of how the different amino acids of the protein participate and affect the reaction mechanism, the catalytic role of the different residues was measured by computing the activation interaction energy, in terms of difference of the electrostatic and Lennard-Jones interaction ($\Delta E_{int} = E_{int-TS} - E_{int-R}$), decomposed by residue. The calculations were done in the

antibody, the efforts were focused in the key chemical steps, as shown in the Figure S9 (a,b,c,d), in which step 1,2, 4 and 5 are respectively depicted .
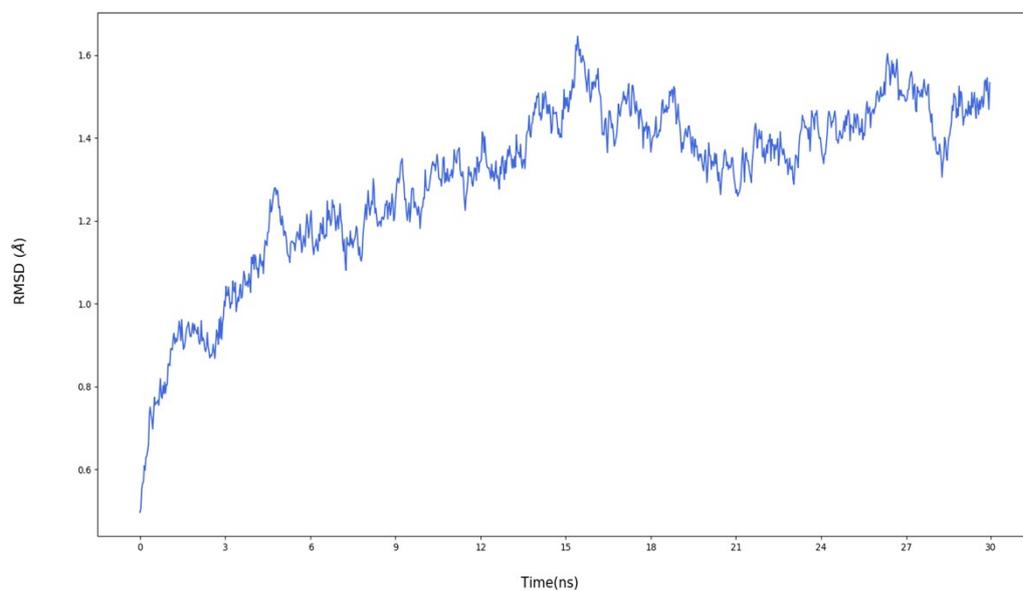
Figure S1. Time evolution of the root-mean-square deviation (RMSD) of the backbone atoms of the protein along the 30 ns of the QM/MM MD simulation
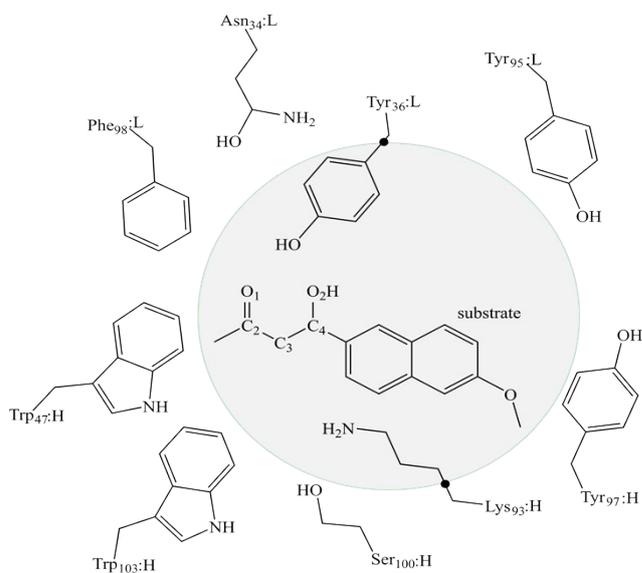
Figure S2. Schematic representation of the active site of the catalytic antibody 33F12. The gray region contains the atoms treated quantum mechanically. Link atoms are represented as black dots.
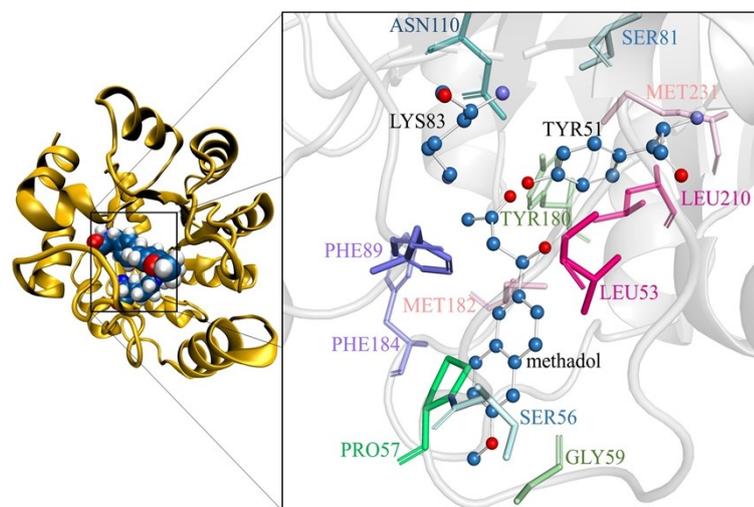
Figure S3. Representation of the X-ray structure RA95.5-8F and detail of the active site in which VDW representation refers to residues that take part in the reaction.

(a)
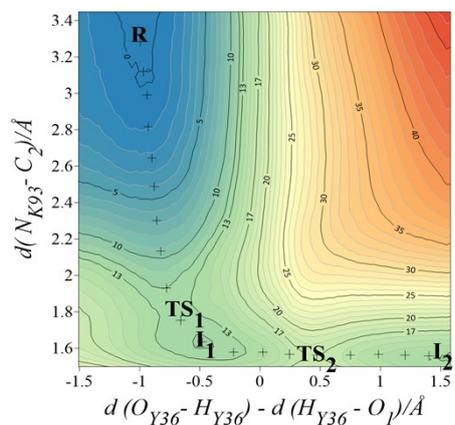
(b)
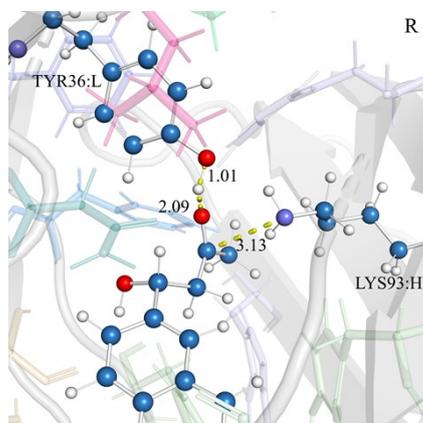
(d)
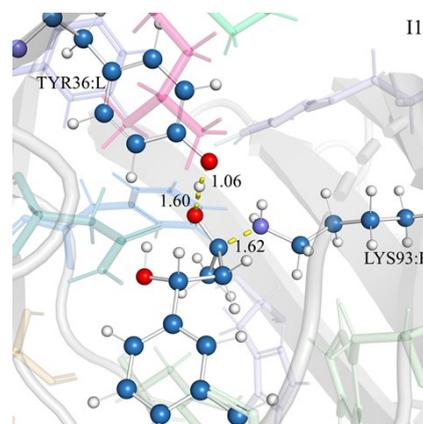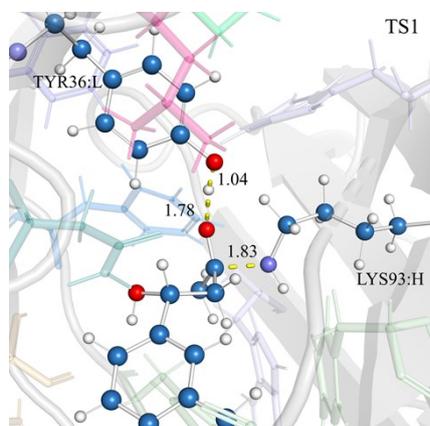
(e)
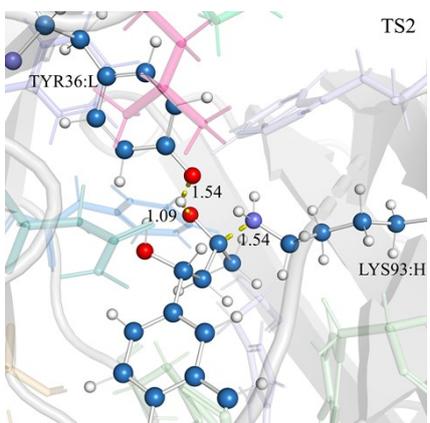
(f)
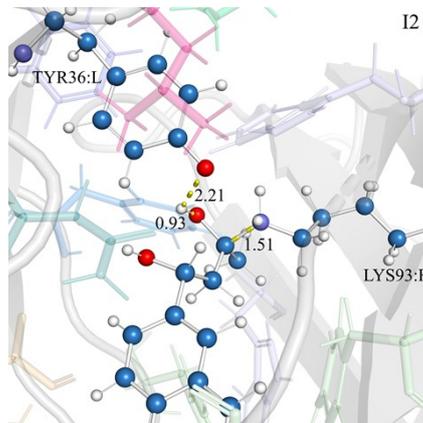
Figure S4. (a) M06-2X:AM1/MM FES for the first and second step of the reaction computed at 300 K. The path indicated as "+" corresponds to the minimum energy path through the TS_1 and TS_2. Representative snapshots of (b) R, (c) TS_1, (d) I_1 (e) TS_2 and (f) I_2. The values of energies are given in kcal mol^{-1} and distances in Å.

(a)

(b)

(c)
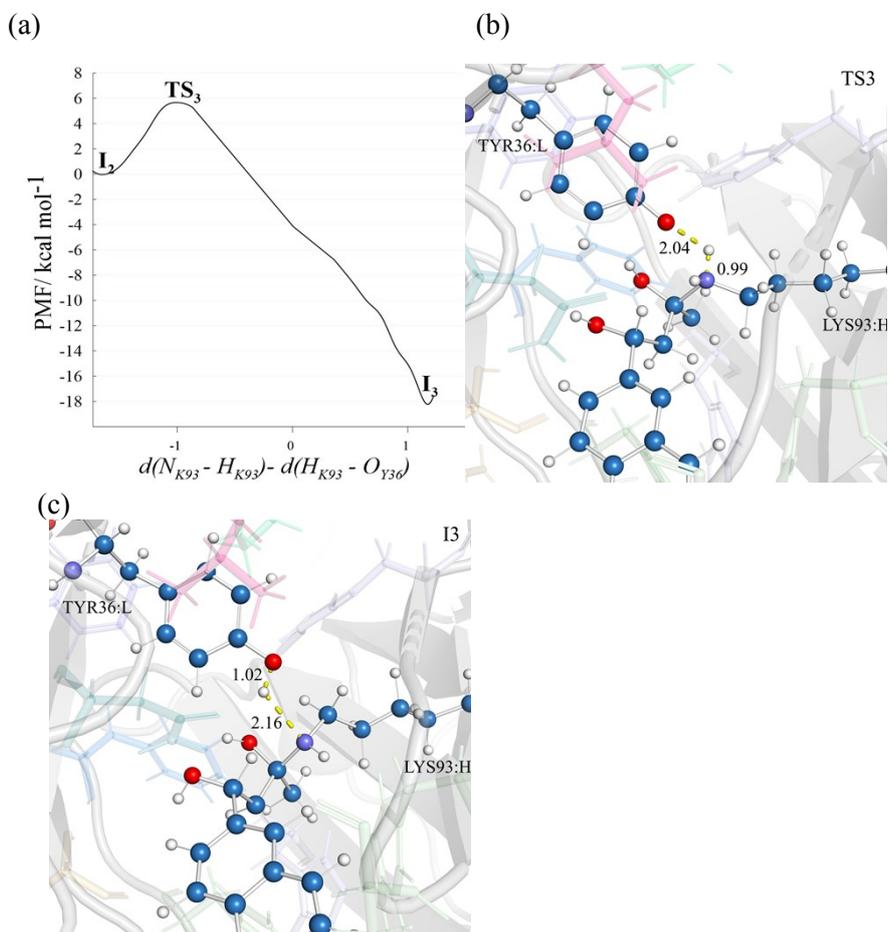
Figure S5. (a) M06-2X:AM1/MM FES for the third step of the reaction computed at 300 K. Representative snapshots of (b) TS_3, (c) I_3. The values of energies are given in kcal mol^{-1} and distances in Å.
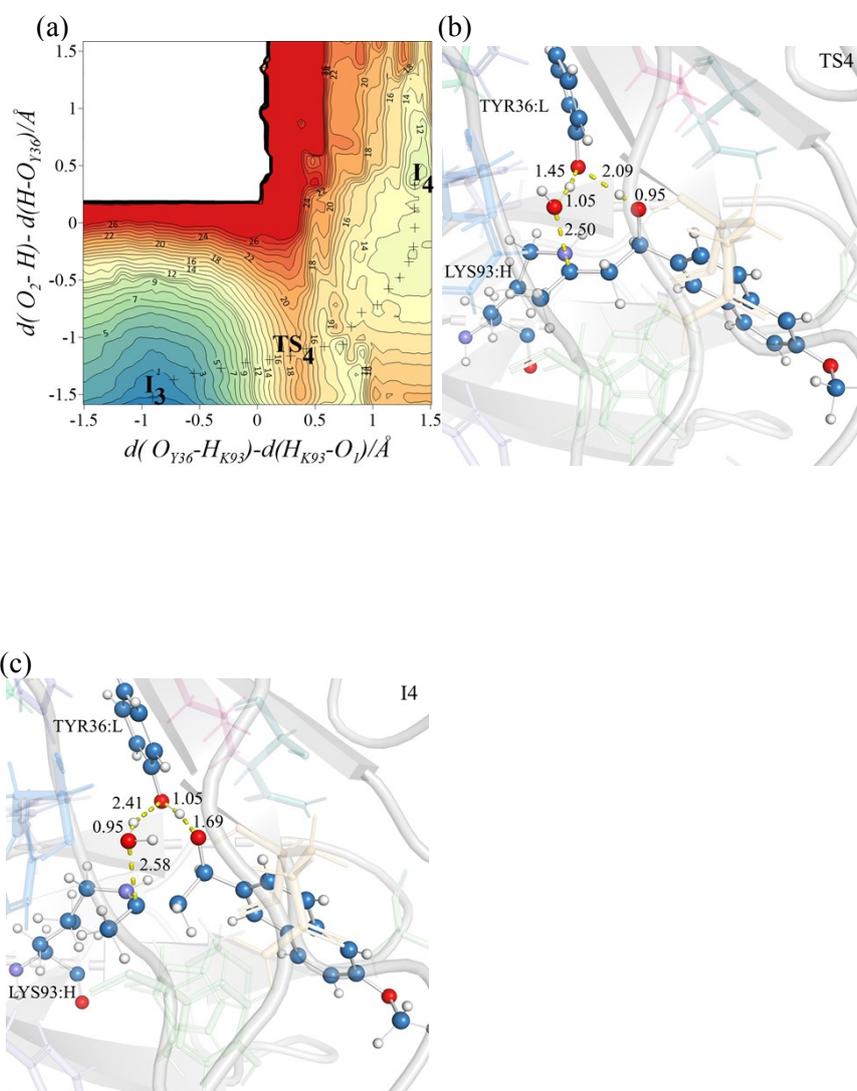
Figure S6. (a) M06-2X:AM1/MM FES for the fourth step of the reaction computed at 300 K. The path indicated as "+" corresponds to the minimum energy path through the TS$_4$. Representative snapshots of the (b) TS$_4$, (c) I$_4$. The values of energies are given in kcal mol^{-1} and distances in Å.
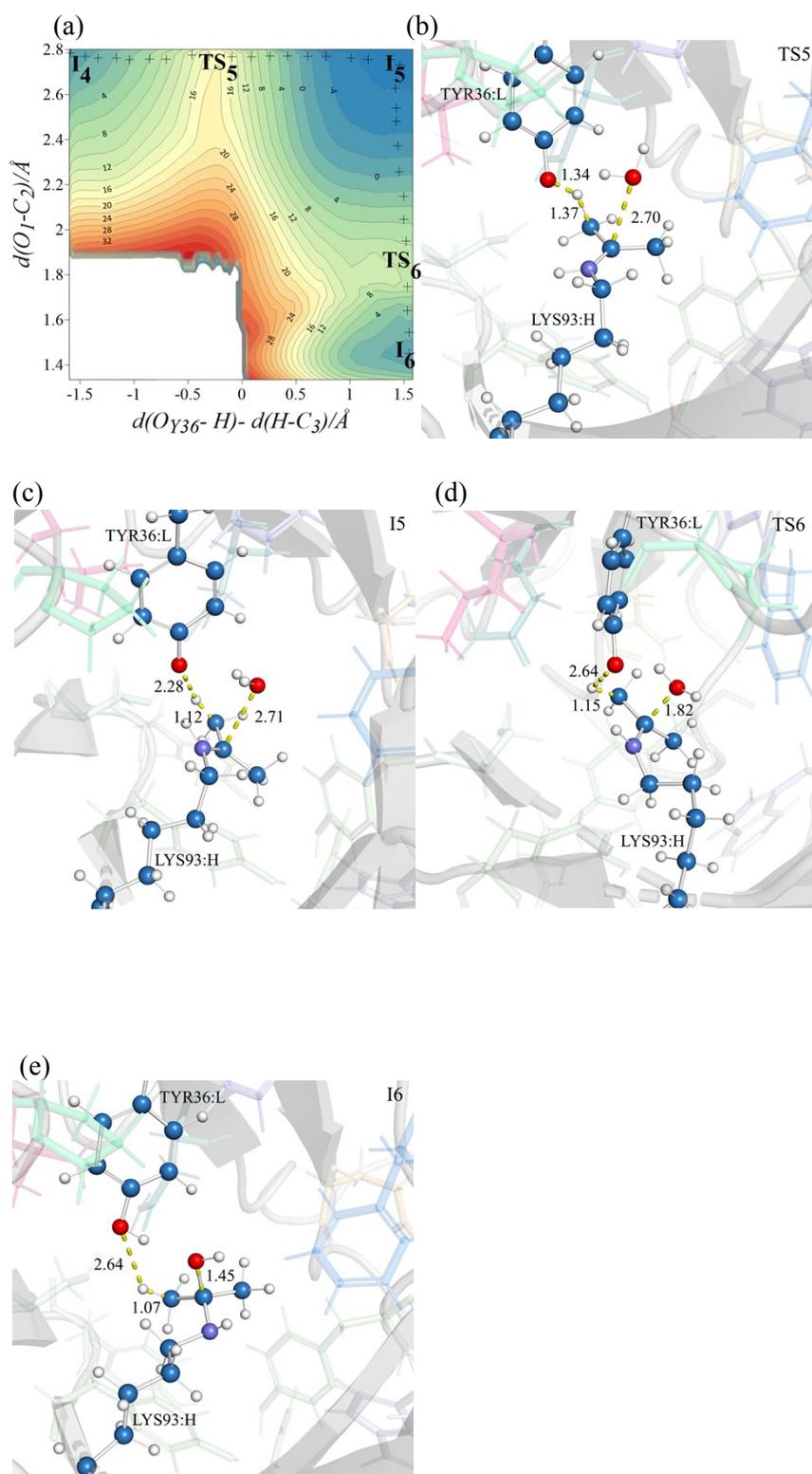
Figure S7. (a) M06-2X:AM1/MM FES for the fifth and sixth steps of the reaction computed at 300 K. The path indicated as "+" corresponds to the minimum energy path through the TS_5 and TS_6. Representative snapshots of (b) TS_5, (c) I_5, (d) TS_6, (e) I_6. The values of energies are given in kcal mol^{-1} and distances in Å.
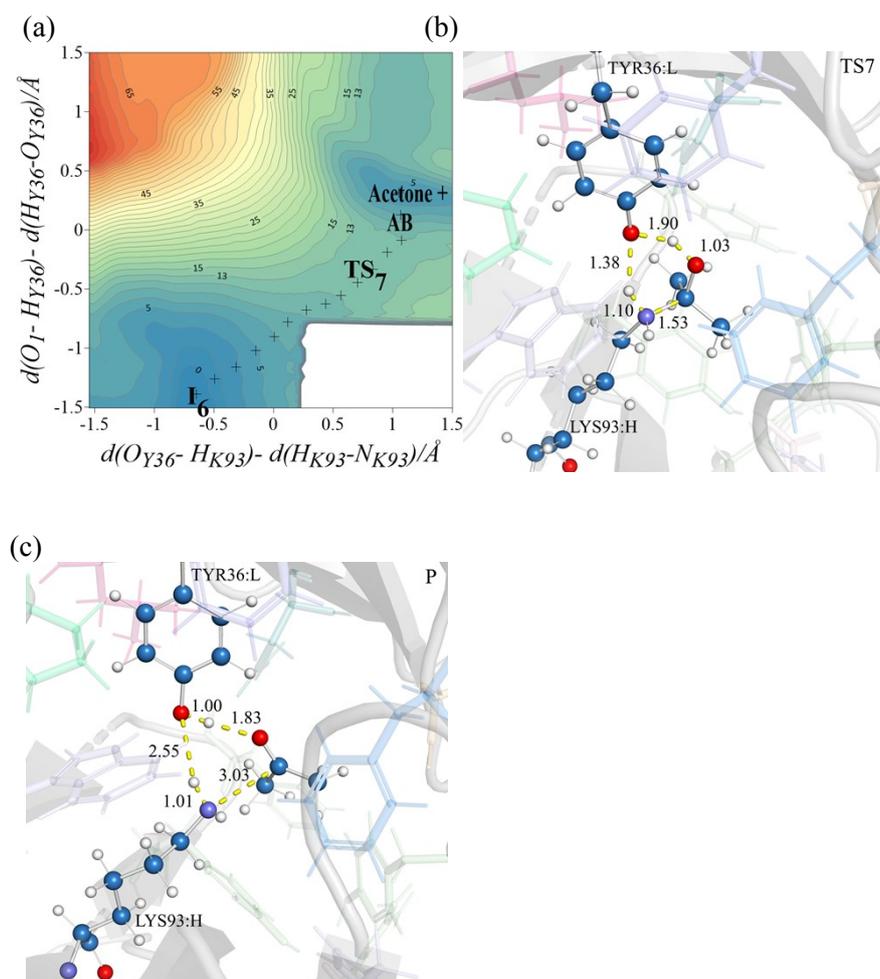
Figure S8. (a) M06-2X:AM1/MM FES for the seventh step of the reaction computed at 300 K. The path indicated as "+" corresponds to the minimum energy path through the TS_7. Representative snapshots of the (b) TS_7, (c) product. The values of energies are given in kcal mol^{-1} and distances in Å.

a) First step
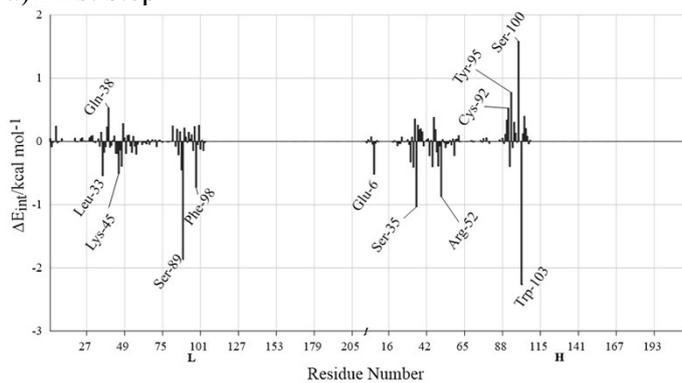
b) Second step
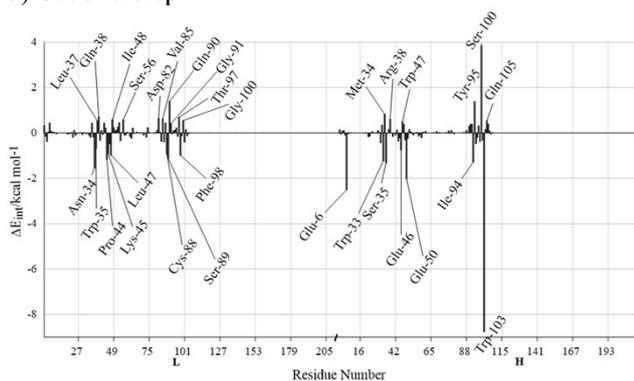
c) Fourth step
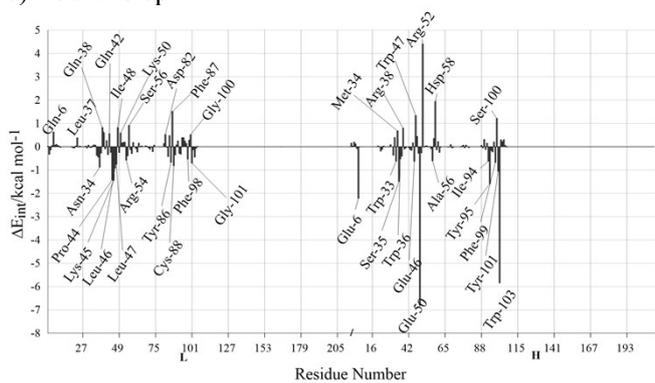
d) Fifth step
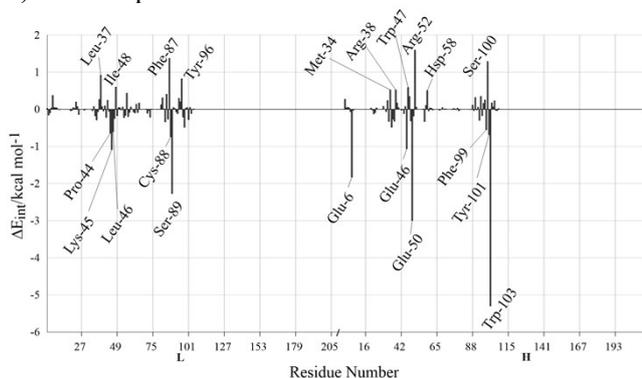
Figure S9. Averaged activation interaction energy of 33F12, for the (a) first, (b)second, (c) fourth and (d) fifth steps of retro-aldol reaction. Only residues showing interaction energies higher than 0.5 kcal mol^{-1}, in absolute value, are labelled in the panels.

a) First and second step
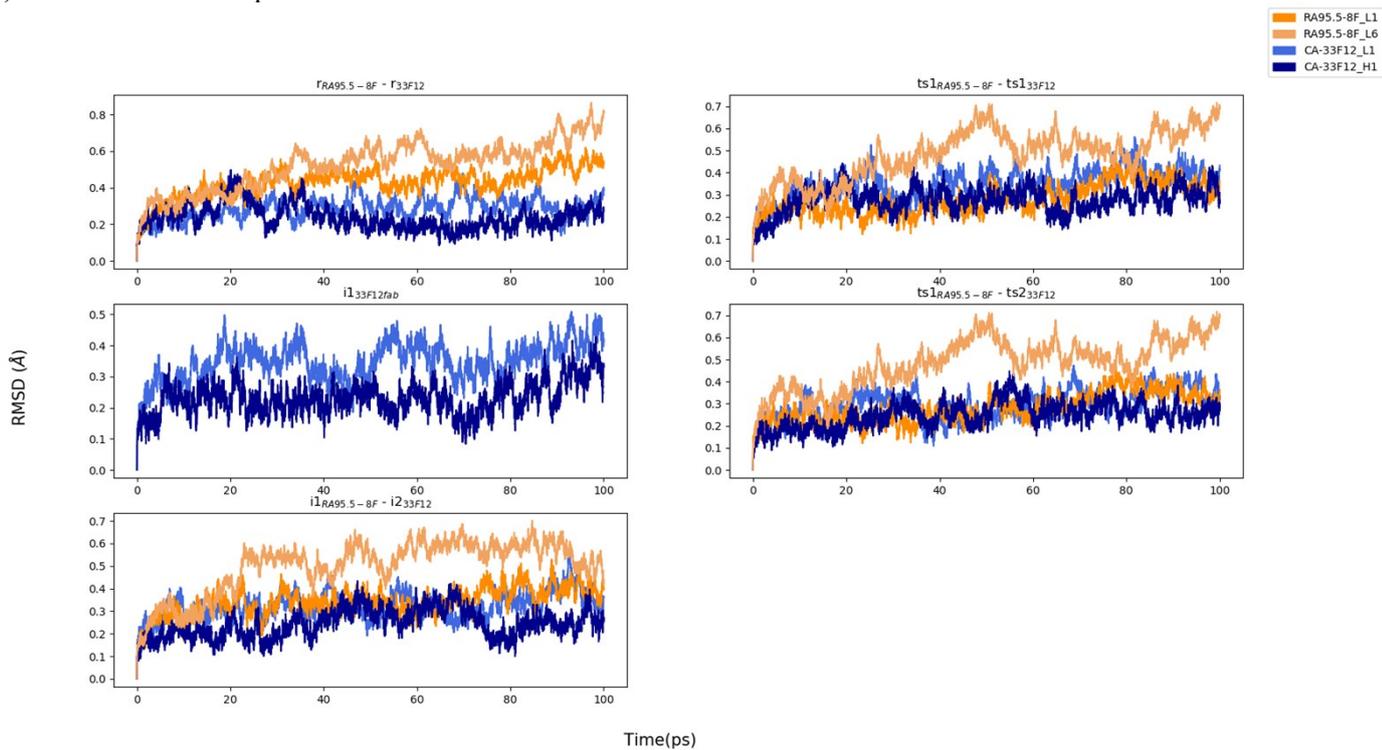
b) Fourth step
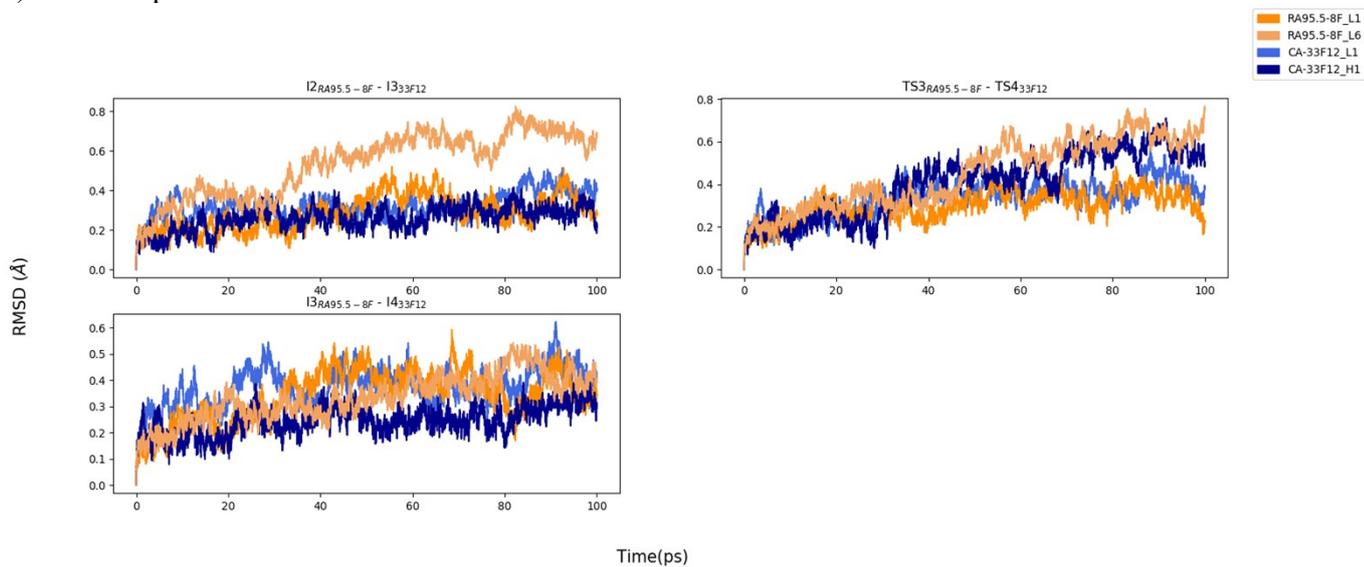
c) Fifth step
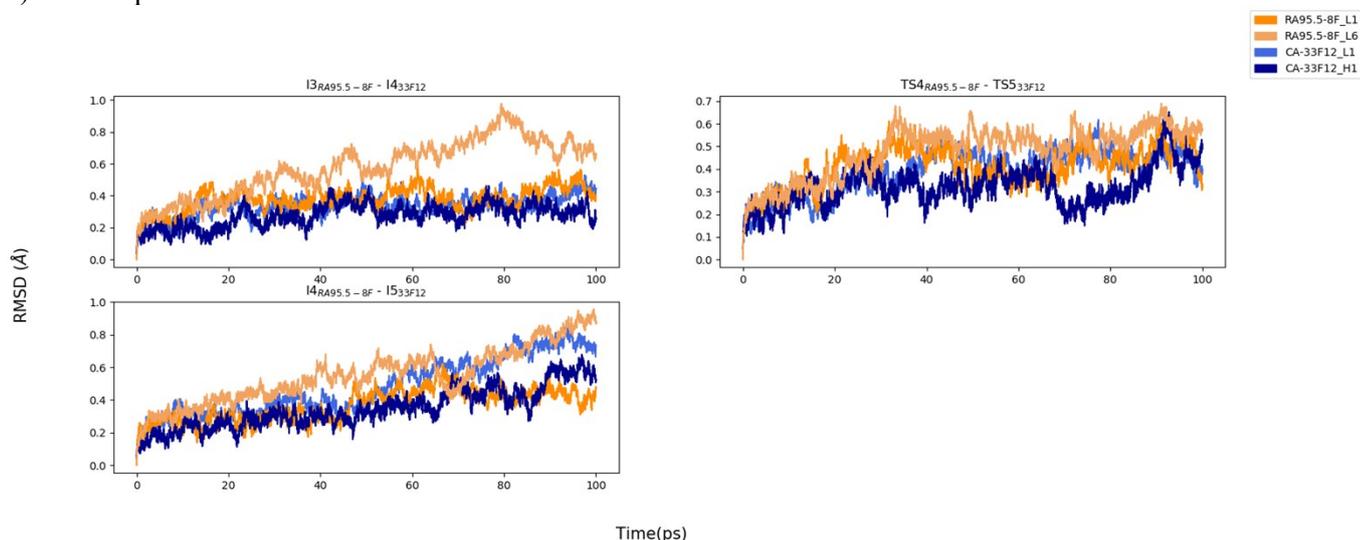
Figure S10. Time evolution of the root-mean-square deviation (RMSD) of the loops L1 (dark orange) and L6 (sandy brown) of *de novo* enzyme RA95.5-8F; and L1 (royal blue) and H1(dark blue) of catalytical antibody 33F12 along 100 ps of the QM/MM MD simulation.

REFERENCES

1   K. Humphrey, A. Dalke and K. Schulten, *J. Mol. Graph.*, 1996, **14**, 33–38.

2   D. C. Bas, D. M. Rogers and J. H. Jensen, *Proteins Struct. Funct. Genet.*, 2008, **73**, 765–783.

3   M. H. M. Olsson, M. Rostkowski and J. H. Jensen, *J. Chem. Theory Comput.*, 2011, **7**, 525–537.

4   L. Verlet, *Phys. Rev.*, 1967, **159**, 98–103.

5   J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kalé and K. Schulten, *J. Comput. Chem.*, 2005, **26**, 1781–1802.

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. Wiórkiewicz-Kuczera, D. Yin and M. Karplus, *J. Phys. Chem. B*,

1998, **102**, 3586–3616.

7   A. D. Vanommeslaeghe, K; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P; Vorobyov, I.; Mackerell JR, *J. Comput. Chem.*, 2009, **30**, 1545–1614.

8   A. D. Mackerell, M. Feig and C. L. Brooks, *J. Comput. Chem.*, 2004, **25**, 1400–1415.

9   M. J. Field, M. Albe, C. Bret, F. Proust-De Martin and A. Thomas, *J. Comput. Chem.*, 2000, **21**, 1088–1100.

10   S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen and P. A. Kollman, *J. Comput. Chem.*, 1992, **13**, 1011–1021.

11   E. Darve and A. Pohorille, *J. Chem. Phys.*, 2001, **115**, 9169–9183.

12   G. M. Torrie and J. P. Valleau, *J. Comput. Phys.*, 1977, **23**, 187–199.

13   W. L. Jorgensen, D. S. Maxwell and J. Tirado-rives, *Am. Chem. Soc.*, 1996, **7863**, 11225–11236.

14   W. L. Jorgensen and J. Tirado-Rives, *J. Am. Chem. Soc.*, 1988, **110**, 1657–1666.

15   J. Pranata, S. G. Wierschke and W. L. Jorgensen, *J. Am. Chem. Soc.*, 1991, **113**, 2810–2819.

16   W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J. Chem. Phys.*, 1983, **79**, 926–935.

17   M. J. Field, P. A. Bash and M. Karplus, *J. Comput. Chem.*, 1990, **11**, 700–733.

18   M. J. S. Dewar, E. G. Zoebisch, E. F. Healy and J. J. P. Stewart, *J. Am. Chem. Soc.*, 1985, **107**, 3902–3909.

19   J. J. Ruiz-Pernía, E. Silla, I. Tuñón and S. Martí, *J. Phys. Chem. B*, 2006, **110**, 17663–17670.

20   M. Roca, V. Moliner, J. J. Ruiz-Pernía, E. Silla and I. Tuñón, *J. Phys. Chem. A*, 2006, **110**, 503–509.

21   J. J. Ruiz-Pernía, E. Silla, I. Tuñón, S. Martí and V. Moliner, *J. Phys. Chem. B*, 2004, **108**, 8427–8433.

22   Y. Y. Chuang, J. C. Corchado and D. G. Truhlar, *J. Phys. Chem. A*, 1999, **103**, 1140–1149.

23   J. C. Corchado, E. L. Coitiño, Y. Y. Chuang, P. L. Fast and D. G. Truhlar, *J. Phys. Chem. A*, 1998, **102**, 2424–2438.

24   K. A. Nguyen, I. Rossi and D. G. Truhlar, *J. Chem. Phys.*, 1995, **103**, 5522–5530.

25   B. J. Lynch, Y. Zhao and D. G. Truhlar, *J. Phys. Chem. A*, 2003, **107**, 1384–1388.

26   J. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, M. . 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.; Sonnenb, D. J. Hada, Y. . Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, F. . B. Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, J. . M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, N. . Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, J. . Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, C. . Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, P. . Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, J. V. . Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, O.; Foresman, J. B.; Ortiz and D. J. Cioslowski, J.; Fox, 2009.