Legends of Supplementary Figures

Figure S1: Amino acid sequence alignment of *Arabidopsis thaliana* HPPD corresponding to the crystal structure 1SQD (A.tha) and *Daucus carota* HPPD (D. car) HPPD. Note that the first 21 amino acids of the Arabidopsis sequence are absent. They were removed in the recombinant Arabidopsis HPPD used for 1SQD crystal structure determination (S1). The three amino acid residues coordinating the catalytic iron are identified by a star. The conserved amino acids mutated in the present study are shown in yellow and are numbered according to the Arabidopsis HPPD (1SQD). The conserved residues S260, N275, Q286, Q300 and Q372 of the Carrot HPPD mutated in this study thus correspond to S246, N261, Q272, Q286 and Q358 of the Arabidopsis HPPD (1SQD), respectively.

Figure S2: Determination of WT and mutant Carrot HPPD kinetic properties**.** The activity of the purified recombinant WT or mutant Carrot HPPD was monitored by measuring the consumption of O_2 during the formation of HGA from HPP (S2). For the clarity of the manuscript mutated residues are numbered according to their corresponding position in Arabidopsis crystal structure 1SQD (S1)
Legends: HPP concentration (μM)

Figure S3: Artist movie of the HPPD reaction mechanism.

Supplementary information on Quantum mechanical/molecular mechanical (QM/MM) calculations:

Program and methods used-All the calculations were done within the QM/MM (Quantum Mechanics/Molecular Mechanics) framework (S3, S4) with a combination of Gaussian03 (S5) for the QM parts and Tinker 4.2 (S6) for the MM part. The QM/MM scheme was similar to the one used in ref (S7) in which dangling bonds at the QM/MM boundary were saturated by hydrogen atoms using the link-atom method (S4). The hybrid B3LYP density functional (S8) was used for the QM part while the CHARMM27 force field (S9) as implemented in Tinker was used for the MM part. Interactions between QM and MM parts were described via both mechanical (van der Waals and London forces) and electrostatic embeddings with direct interaction of MM charge with the electron density of the QM part. This ensured that the influence of the enzyme structure on the active site was taken into account.
QM/MM geometry optimizations were done with Gaussian/Tinker using a double-zeta basis set for all the QM atoms. We used the same LACVP (S10) basis set as Borowski *et al*. in their study of a QM-only reduced model of the same system (S11). “Loose” convergence criteria were used to speed up geometry optimizations but tight ones were kept for the convergence of the wave functions, which were computed using a fine integration grid in Gaussian.

Setup of the systems-Initial coordinates for 4-hydroxyphenylpyruvate dioxygenase used in QM/MM calculations were taken from the X-ray structure found in the Protein Data Bank under code 1SQD (S1). Since we were primarily interested in exploring the potential interactions of substrates with surrounding residues in a qualitative way, calculations were started directly from the PDB structure to which missing hydrogen atoms were added using the CHARMM Program (S13) with the charmm27 force field (S9). All histidine residues were singly protonated on their Nδ atom. Aspartate and glutamate residues were kept unprotonated (negative charge) while arginine and lysine residues were positively charged.

Since the PDB structure doesn't contain the substrate, the first task was to add it in the active site. This was done in two ways: i) Preliminary QM calculations were performed to obtain the structures of a reduced active site corresponding to those published by Borowski *et al.* Then, a superimposition (by least square minimization) of these active site atoms (iron and its protein ligands, without substrate), with their coordinates given by the QM calculation, on the same atoms in the PDB structure, allowed us to place the substrate in this latter structure. ii) The orientation of the substrate after superimposition of the active site atoms was modified by rigid body translation and rotation to better interact with the protein residues proved experimentally to be involved in the catalytic process. To

avoid being trapped in local energy minima, different rotamers of these protein residues were also considered.
The initial structures obtained after i) or ii) were then energy minimized using the QM/MM methodology described above. Among the different structures thus optimized, the ones with the lowest internal QM/MM energy are presented and discussed in this work.

QM/MM partition- The QM part of the system contained the active site with the side chains of H205, H287, E373, the Iron-oxo group and the substrate (HPP, HAP or INT – Ref. Fig.5). In addition and since we wished to focus on possible interactions of the substrate with particular residues of the protein, side chains of S246, N261, Q272, Q286 and Q358 were also included in the QM part. This made a total of 85 QM atoms, including the eight link atoms added to saturate dangling bonds.
Finally, the so-called active part, including both QM (all of them) and MM atoms which were allowed to relax during geometry optimization, contained the following residues: H205, V207, N209, F232, G243, L244, S246, A247, P259, I260, N261, Q272, E262, I273, Q286, H287, A289, L347, L356, L357, Q358, I359, F360, F371, E373, I375, F398, G399, K400, N402 and F403. Of course, substrate, Iron and distal oxygen were also part of this region.

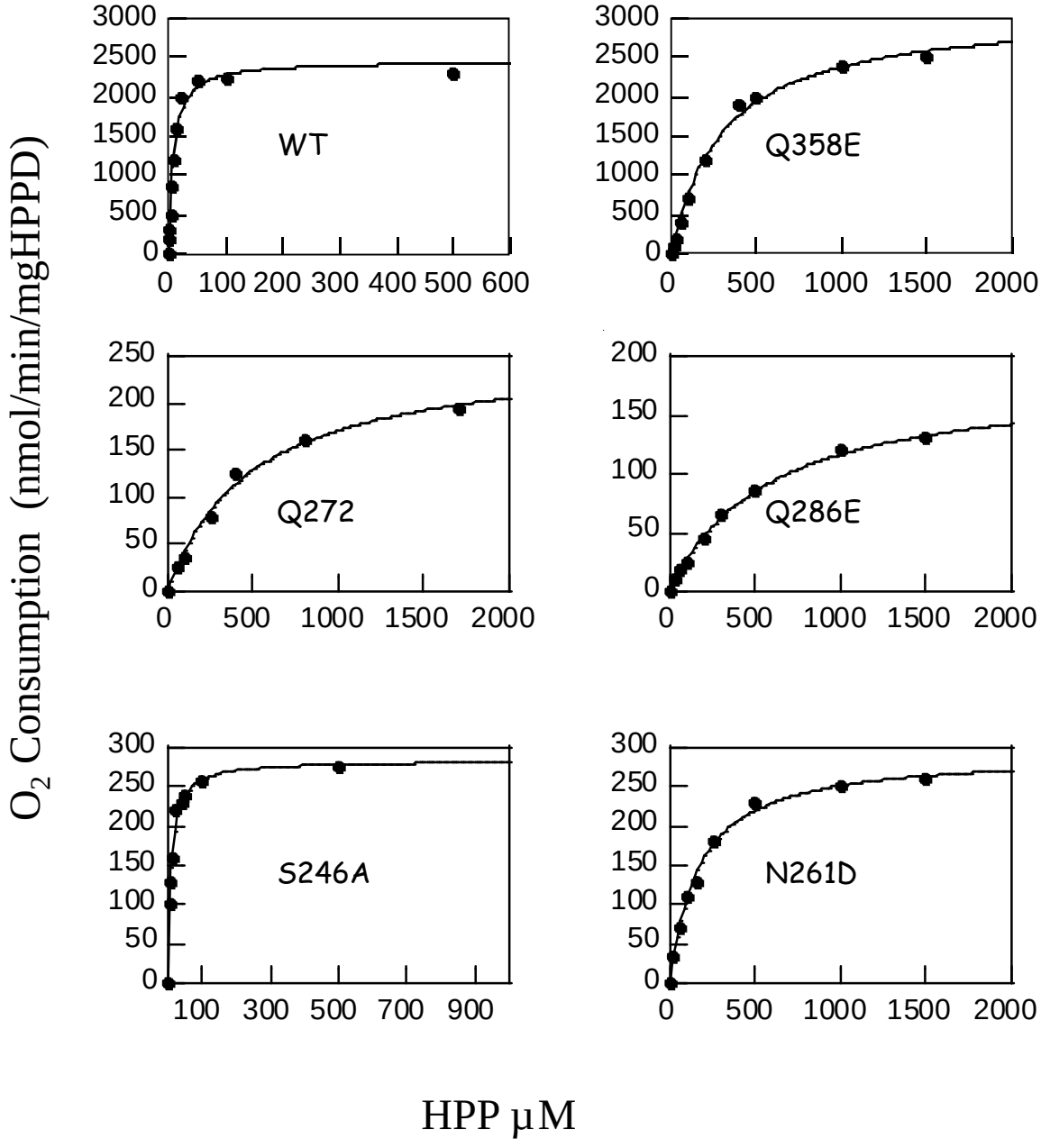
References for Supplementary Materials:

(S1). Yang, C., Pflugrath, J.W., Camper, D.L., Foster, M. L., Pernich, D. J. and Walsh, T.A. (2004) *Biochemistry* **43**, 10414-10423
(S2). Garcia, I., Job, D., Matringe, M. (2000) *Biochemistry* **39**, 25, 7501-7507
(S3) Warshell, A., and Levitt, M. (1976) *J. Mol. Biol.* **103**,227-249
(S4) Field, M. J., Bash, P. A., and Karplus, M. (1990) *J. Comput. Chem.* **11**, 700-733
(S5) Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., Montgomery, J. A., Jr., Vreven, T., Kudin, K. N., Burant, J. C., Millam, J. M., Iyengar, S. S., Tomasi, J., Barone, V., Mennucci, B., Cossi, M., Scalmani, G., Rega, N., Petersson, G. A., Nakatsuji, H., Hada, M., Ehara, M., Toyota, K., Fukuda, R., Hasegawa, J., Ishida, M., Nakajima, T., Honda, Y., Kitao, O., Nakai, H., Klene, M., Li, X., Knox, J. E., Hratchian, H. P., 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., Ayala, P. Y., Morokuma, K., Voth, G. A., Salvador, P., Dannenberg, J. J., Zakrzewski, V. G., Dapprich, S., Daniels, A. D., Strain, M. C., Farkas, O., Malick, D. K., Rabuck, A. D., Raghavachari, K., Foresman, J. B., Ortiz, J. V., Cui, Q., Baboul, A. G., Clifford, S., Cioslowski, J., Stefanov, B. B., Liu, G., Liashenko, A., Piskorz, P., Komaromi, I., Martin, R. L., Fox, D. J., Keith, T., Al-Laham, M. A., Peng, C. Y., Nanayakkara, A., Challacombe, M., Gill, P. M. W., Johnson, B., Chen, W., Wong, M. W., Gonzalez, C., Pople, J. A. (2004), Gaussian 03, revision B.05, Gaussian, Inc.: Wallingford, CT.
(S6) Ponder, J. W. (2001) *TINKER: Software Tools for Molecular Design*, 4.2 ed.; Washington University School of Medicine: Saint Louis, MO.
(S7) Ferré, N., and Olivucci, M. (2003) *J. Am. Chem. Soc.*, **125**, 6868–6869.
(S8a) Becke, A. D.(1988) *Phys. Rev. A,* **38**, 3098-3100.
(S8b) Becke, A. D., (1993) *Chem. Phys.* **98**, 1372-1377.
(S8c) Becke, A. D. (1993) *Chem. Phys.* **98**, 5648-5652.
(S8d) Lee, C., Yang, W., Parr, R. G. (1988) Phys. Rev. B, **37**, 785-789.
(S9) Foloppe, N., and MacKerell, A. D., (2000) *J. Comput. Chem.*, **21**, 86-104
(S10a) Hay, J. P., Wadt, W. R. (1985) *J. Chem. Phys.* **82**, 299.
(S10b) Friesner, R. A., Murphy, R. B., Beachy, M. D., Ringlanda, M. N., Pollard, W. T., Dunietz, B. D., and Cao, Y. X. (1999) *J. Phys. Chem. A* **103**, 1913.
(S11) Borowski, T., Bassan, A., and Siegbahn, P. E. M., (2004) *Biochemistry*, **43**, 12331-12342
(S12) Brooks, B.R., Bruccoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S., and M. Karplus, M. (1983) *J. Comput. Chem.*, G, 187-217

Figure S1


1
A. tha ----------PGFKLVGFSKFVRKNPKSDKFKVKRFHHIEFWCGDATNVA
D. car MGKKQSEAEILSSNSSNTSPATFKLVGFNNFVRANPKSDHFAVKRFHHIEFWCGDATNTS
A. tha RRFSWGLGMRFSAKSDLSTGNMVHASYLLTSGDLRFLFTAPYSPSLSAGEIKPTTTASIP
D. car RRFSWGLGMPLVAKSDLSTGNSVHASYLVRSANLSFVFTAPYSPSTTTS----SGSAAIP
A. tha SFDHGSCRSFFSSHGLGVRAVAIEVEDAESAFSISVANGAIPSSPPIVLNEAVTIAEVKL
D. car SFSASGFHSFAAKHGLAVRAIALEVADVAAAFEASVARGAR PASAPVELDDQAWLAEVEL
205
A. tha YGDVVLRYVSYKAEDTEKSEFLPGFERVEDASSFP-LDYGIRRLDHAVGNVPELGPALTY
D. car YGDVVLRFVSFGREE---GLFLPGFEAVEGTASFPDLDYGIRRLDHAVGNVTELGPVVEY
246 261 * 272
A. tha VAGFTGFHQFAEFTADDVGTAESGLNSAVLASNDEMVLLPINEPVHGTKRKSQIQTYLEH
D. car IKGFTGFHEFAEFTAEDVGTLESGLNSVVLANNEEMVLLPLNEPVYGTKRKSQIQTYLEH
286
A. tha NEGAGLQHLALMSEDIFRTLREMRKRSSIGGFDFMPSPPPTYYQNLKKRVGDVLSDDQIK
D. car NEGAGVQHLALVSEDIFRTLREMRKRSCLGGFEFMPSPPPTYYKNLKNRVGDVLSDEQIK
*
358 373
A. tha ECEELGILVDRDDQGTLLQIFTKPLGDRPTIFIEIIQRVGCMMKDEEGKAYQSGGCGGFG
D. car ECEDLGILVDRDDQGTLLQIFTKPVGDRPTLFIEIIQRVGCMLKDDAGQMYQKGGCGGFG
*
A. tha KGNFSELFKSIEEYEKTLEAKQLVG----
D. car KGNFSELFKSIEEYEKTLEAKQITGSAAA

Figure S2


WT
Q358E
Q272
Q286E
S246A
N261D
O_2 Consumption (nmol/min/mgHPPD)
HPP µM