Supporting Information

Rationally Guided Improvement of NOV1 Dioxygenase for the Conversion of Lignin-Derived Isoeugenol to Vanillin

Mario De Simone[1], Laura Alvigini[2], Lur Alonso-Cotchico[3], Vânia Brissos[1], Jonatan Caroli[2], Maria Fátima Lucas[3], Emanuele Monza[3], Eduardo Pinho Melo[4], Andrea Mattevi[2*] and Lígia O Martins[1*]

[1]*Instituto de Tecnologia Química e Biológica António Xavier, Universidade NOVA de Lisboa, Av. da República, 2780-157 Oeiras, Portugal*
[2]*Dept. Biology and Biotechnology, University of Pavia, Via Ferrata 9, 27100 Pavia, Italy*
[3]*Zymvol Biomodeling, Carrer Roc Boronat, 117, 08018 Barcelona, Spain*
[4]*Centro de Ciências do Mar, Universidade do Algarve, 8005-139 Faro, Portugal*

*****Correspondence: andrea.mattevi@unipv.it, lmartins@itqb.unl.pt

Computational methods

Structure preparation. The X-ray structures of the wild-type (PDB code: 5j55)[1] and S283F NOV1 variant (this work) were used as model systems. Both structures were cleaned with Yasara, the hydrogen atoms were added, and the hydrogen bond networks were optimized.[2] The loops composed of the residues 383-389 for the wild-type and 383-390 for the variant, missed in both X-ray structures, were built using BuildLoop.[3] The best-fit model was selected, and its structure was minimized using the AMBER14 force field.[4] From these templates, both the *apo* and *holo* forms of the NOV1 wild-type and S283F variant were modeled. For consistency, the *holo* forms of the proteins were built like Fe(II)-O_2 complexes (which represents a pre-catalytic center) and in their ligand-free form. In these structures, the vanillin molecule was manually removed. For the *holo* variant, an oxygen molecule was added and coordinated to the metal center (already present in the WT X-ray). The resulting Fe(II)-O_2complex was minimized with the four histidine residues (167, 218, 284, 476) coordinating the metal; for the *apo* forms, the ligand, metal, and molecular oxygen were manually removed. The AMBER14 force field was used in all the energy and gradient calculations. The resulting systems were used as starting point for Molecular Dynamics (MD) simulations.

Molecular Dynamics Simulations. Four independent trajectories were run for four different systems: 1) wild-type in its Fe(II)-O_2 substrate-free form, 2) S283F in its Fe(II)-O_2 substrate-free form, 3) wild-type in its *apo* form, and 4) S283F variant in its *apo* form. The preparation of the NOV1 structures and the MD simulations were carried out using Yasara.[5] The setup included optimization of the hydrogen bonds network, pKa prediction at pH 7.4, solvation using a water buffer 9 Å around the protein to form a cubic box and charge neutralization using Na^+ ions. The system was then minimized with steepest descent and simulated annealing, followed by 400 ps of equilibration at constant volume, using a time step of 1 fs. The temperature was gradually increased to 300K, along the first half of the equilibration stage, and then kept fixed. Each production run, 16 in total, consisted of 200 ns at the same temperature at a fixed volume. The bonded and non-bonded forces were updated every 2.5 fs and 5 fs, respectively, and a Berendsen thermostat[6] was used to control the temperature. The protein was treated with the AMBER14 force field [4] and the waters with the TIP3P.[7] The van der Waals forces cutoff was set at 8 Å and the long-range electrostatic forces were represented using the Particle Mesh Ewald algorithm.[8] The LINCS[9] and SETTLE[10] algorithms were used to restrain stretching and bending terms involving the hydrogen atoms and the water molecules of the system.

Free volume analysis. MDpocket, part of the Fpocket suite[11, 12], was used to assess the free volume available inside the core of the protein and active site along with the MD simulation timescale.

Molecular graphics. The molecular graphics have been created with YASARA (www.yasara.org), POVRay (www.povray.org)[13], and PyMol.[14]

Plots. The treatment of numerical data and analytical plots have been performed using the python modules pandas,[15, 16] NumPy,[17] Matplotlib,[17] and Seaborn.[18]

Table S1 - Primers used for the construction of the thirty-five variants used in this work.

#	Primers	Sequence
1	N120L Fwd	ACTGCACTCACAAATGCGTTTGTTTTCGGAGGTAAG
	N120L Rev	ATTTGTGAGTGCAGTCGAACGGATCTCTCCTTTAACAG
2	N120S Fwd	GATCCGTTCGACTGCAAGCACAAATGCGTTTG
	N120S Rev	CAAACGCATTTGTGCTTGCAGTCGAACGGATC
3	T121P Fwd	GCAAACCCAAATGCGTTTGTTTTCGGAGGTAAGTTGTG
	T121P Rev	CGCATTTGGGTTTGCAGTCGAACGGATCTCTC
4	T121A Fwd	CCGTTCGACTGCAAACGCGAATGCGTTTGTTTTC
	T121A Rev	GAAAACAAACGCATTCGCGTTTGCAGTCGAACGG
5	F281W Fwd	AATTGTTGGGCGTCTCACGTTTTGAACGCATG
	F281W Rev	AGACGCCCAACAATTGTCGCGGGTAAAC
6	S283F Fwd	GCGTTTCACGTTTTGAACGCATGGCAGGAG
	S283F Rev	CAAAACGTGAAACGCAAAACAATTGTCGCGGGTAAAC
7	S283Q Fwd	GCGCAACACGTTTTGAACGCATGGCAGGAGGGAACAAAGATC
	S283Q Rev	CAAAACGTGTTGCGCAAAACAATTGTCGCGGGTAAACCAACGAATAT
8	S283I Fwd	GACAATTGTTTTGCGATTCACGTTTTGAACGC
	S283I Rev	GCGTTCAAAACGTGAATCGCAAAACAATTGTC
9	S283E Fwd	GACAATTGTTTTGCGGTGCACGTTTTGAACGC
	S283E Rev	GCGTTCAAAACGTGCACCGCAAAACAATTGTC
10	S283N Fwd	GACAATTGTTTTGCGAACCACGTTTTGAACGC
	S283N Rev	GCGTTCAAAACGTGGTTCGCAAAACAA TTGTC
11	S283V Fwd	GACAATTGTTTTGCGGTTCACGTTTTGAACGC
	S283V Rev	GCGTTCAAAACGTGAACCGCAAAACAATTGTC
12	S283C Fwd	GACAATTGTTTTGCGTGCCACGTTTTGAACGC
	S283C Rev	GCGTTCAAAACGTGGCACGCAAAACAATTGTC
13	S283T Fwd	GACAATTGTTTTGCGACCCACGTTTTGAACGC
	S283T Rev	GCGTTCAAAACGTGGGTCGCAAAACAATTGTC
14	F307H Fwd	GGCCAAAAACAATATGTGGCCTTTTTTTCCCGACG
	F307H Rev	CGTCGGGAAAAAAAGGCCACATATTGTTTTTGGCC
15	F307W Fwd	AATATGTGGCCTTTTTTTCCCGACGTTCATGGCGCACCATTCAACG
	F307W Rev	AAAAGGCCACATATTGTTTTTGGCCTCACAAGTAACAAAGTGGA
16	F309W Fwd	CAATATGTTCCCTTGGTTTCCCGACGTTCATG
	F309W Rev	CATGAACGTCGGGAAACCAAGGGAACATATTG
17	F473W Fwd	CAATCCGTCTGCGCTGGGGACTTCACGGGAAC
	F473W Rev	GTTCCCGTGAAGTCCCCAGCGCAGACGGATTG
18	F473E Fwd	CAATCCGTCTGCGCGAAGGACTTCACGGGAAC
	F473E Rev	GTTCCCGTGAAGTCCTTCGCGCAGACGGATTG
19	F473Q Fwd	CAATCCGTCTGCGCCAGGGACTTCACGGGAAC
	F473Q Rev	GTTCCCGTGAAGTCCCTGGCGCAGACGGATTG
20	L475S Fwd	CGCTTCGGATCTCACGGGAACTGGGCAAATGCCGATGAAATTG
	L475S Rev	GTGAGATCCGAAGCGCAGACGGATTGGGATATTAACCGT
21	L475G Fwd	CAATCCGTCTGCGCTTCGGAGGCCACGGGAACTGGGCAAATG
	L475G Rev	CATTTGCCCAGTTCCCGTGGCCTCCGAAGCGCAGACGGATTG
22	L475T Fwd	CAATCCGTCTGCGCTTCGGAACCCACGGGAACTGGGCAAATG
	L475T Rev	CATTTGCCCAGTTCCCGTGGGTTCCGAAGCGCAGACGGATTG
23	E353D Fwd	GTGACACCGCTGCGGATTTTCCTCGCATCGAC
	E353D Rev	GTCGATGCGAGGAAAATCCGCAGCGGTGTCAC
24	F281C-S283I Fwd	GGTTTACCCGCGACAATTGTTGCGCGATTCACGTTTTGAACGCATGGC
	F281C-S283I Rev	GCCATGCGTTACAAACGTGAATCGCGCAACAATTGTCGCGGGTAAACC
25	F281C-S283T Fwd	GGTTTACCCGCGACAATTGTTGCGCGACCCACGTTTTGAACGCATGGC
	F281C-S283T Rev	GCCATGCGTTCAAAACGTGGGTCGCGCAACAATTGTCGCGGGTAAACC
26	F281I-S283V Fwd	GGTTTACCCGCGACAATTGTATTGCGGTGCACGTTTTGAACGCATGGC
	F281I-S283V Rev	GCCATGTCAAAACGTGCACCGCCATACAATTGTCGCGGGTAAACC
27	F281H-S283I Fwd	GGTTTACCCGCGACAATTGTCATGCGATTCACGTTTTGAACGCATGGC
	F281H-S283I Rev	GCCATGCGTTCAAAACGTGAATCGCATGACAATTGTCGCGGGTAAACC
28	F281H-S283V Fwd	GGTTTACCCGCGACAATTGTCATGCGGTGCACGTTTTGAACGCATGGC
	F281H-S283V Rev	GCCATGCGTTCAAAACGTGCACCGCATGACAATTGTCGCGGGTAAACC

29	F281M-S283I Fwd	GGTTTACCCGCGACAATTGTATGGCGATTCACGTTTTGAACGCATGGC
	F281M-S283I Rev	GCCATGTCAAAACGTGAATCGCCATACAATTGTCGCGGGTAAACC
30	F281M-S283T Fwd	GGTTTACCCGCGACAATTGTATGGCGACCCACGTTTTGAACGCATGGC
	F281M-S283T Rev	GCCATGCGTTCAAAACGTGGGTCGCCATACAATTGTCGCGGGTAAACC
31	F281M-S283V Fwd	GGTTTACCCGCGACAATTGTATGGCGGTGCACGTTTTGAACGCATGGC
	F281M-S283V Rev	GCCATGCGTTCAAAACGTGCACCGCCATACAATTGTCGCGGGTAAACC
32	F473W-L475G Fwd	CCCAATCCGTCTGCGCTGGGGAGGCCACGGGAACTGGGCAAATGCC
	F473W-L475G Rev	GGCATTTGCCCAGTTCCCGTGGCCTCCCCAGCGCAGACGGATTGGG
33	F473W-L475T Fwd	CCCAATCCGTCTGCGCTGGGGAACCCACGGGAACTGGGCAAATGCC
	F473W-L475T Rev	GGCATTTGCCCAGTTCCCGTGGGTTCCCCAGCGCAGACGGATTGGG
34	F281M S283I F307H[1] Fwd	GGCCAAAAACAATATGTGGCCTTTTTTTCCCGACG
	F281M S283I F307H[1] Rev	CGTCGGGAAAAAAAGGCCACATATTGTTTTTGGCC
35	F281M S283T F307H[2] Fwd	GGCCAAAAACAATATGTGGCCTTTTTTTCCCGACG
	F281M S283T F307H[2] Rev	CGTCGGGAAAAAAAGGCCACATATTGTTTTTGGCC

[1] F307H mutation was introduced using the gene containing F281M_S283I mutations
[2] F307H mutation was introduced using the gene containing F281M_S283T mutations

Table S2. Data collection and refinement statistics for the crystal structure of S283F.

Space group	C222
Unit cell axes (Å)	*a* = 178.471
	b = 187.984
	c = 105.867
Resolution (Å)	2.9
PDB code	7QR6
R_{sym} [a,b] (%)	9.2 (73.1)
$CC_{1/2}$ (%)	98.1 (30.3)
Completeness (%)	98.9 (97.5)
Unique reflections	36,348
Redundancy	3.5 (3.5)
I/σ [b]	6.1 (1.6)
N° of non-hydrogen atoms [c]	
protein/Fe	11370/3
water	16
Average B value for protein atoms (Å^2)	71.0
R_{cryst} [b,d] (%)	21.8
R_{free} [b,d] (%)	29.0
Rms bond length (Å)	0.008
Rms bond angles (°)	1.48

[a] $R_{sym}=\sum|I_i-<I>|/\sum I_i$, where I_i is the intensity of i^{th} observation and <I> is the mean intensity of the reflection.

[b] Values in parentheses are for reflections in the highest resolution shell.

[c] The asymmetric unit contains three protein chains. The final model comprises residues 10-382 and 391-488 of subunit A, 5-382 and 391-488 of subunit B, 3-382, and 391-488 of subunit C.

[d] $R_{cryst}=\sum|F_{obs}-F_{calc}|/\sum|F_{obs}|$ where F_{obs} and F_{calc} are the observed and calculated structure factor amplitudes, respectively. R_{cryst} and R_{free} were calculated using the working and test sets, respectively.

Table S3. Mutant design and variants' enzymatic activity. Reactions were performed using cell-crude extracts in 0.1 M Tris-HCl, pH 9, in the presence of 1 mM isoeugenol at room temperature.

#	Enzyme	Rational design	Initial Activity (iA) (mU/mL)	Relative activity to the wild-type ($iA_{Variant}/iA_{WT}$)
	Wild-type		0.252 ± 0.027	-
1	N120L	[19]	nd	nd
2	N120S	Rosetta + docking	0.019 ± 0.006	0.07 ± 0.03
3	T121P	Rosetta	nd	nd
4	T121A	Rosetta + docking	0.052 ± 0.009	0.18 ± 0.03
5	F281W	[18]	0.047 ± 0.011	0.23 ± 0.05
6	S283F	[16]	0.433 ± 0.082	1.50 ± 0.29
7	S283Q	[19]	0.031 ± 0.005	0.15 ± 0.02
8	S283I	Rosetta + docking	0.044 ± 0.019	0.18 ± 0.08
9	S283E	Rosetta + docking	0.040 ± 0.010	0.17 ± 0.04
10	S283N	Rosetta + docking	nd	nd
11	S283V	Rosetta + docking	0.118 ± 0.027	0.61 ± 0.14
12	S283C	Rosetta + docking	0.012 ± 0.004	0.05 ± 0.02
13	S283T	Rosetta + docking	0.063 ± 0.018	0.26 ± 0.02
14	F307H	Visual inspection	0.017 ± 0.010	0.07 ± 0.04
15	F307W	[14]	0.030 ± 0.006	0.10 ± 0.02
16	F309W	Rosetta + docking	0.016 ± 0.003	0.05 ± 0.01
17	F473W	Rosetta + docking	0.020 ± 0.005	0.07 ± 0.02
18	F473E	Rosetta + docking	0.018 ± 0.004	0.06 ± 0.01
19	F473Q	Rosetta + docking	0.014 ± 0.005	0.06 ± 0.02
20	L475S	[19]	0.011 ± 0.004	0.04 ± 0.02
21	L475G	Rosetta	0.017 ± 0.004	0.07 ± 0.02
22	L475T	Rosetta + docking	0.016 ± 0.003	0.05 ± 0.01
23	E353D	Rosetta + docking	0.013 ± 0.006	0.05 ± 0.03
24	F281C S283I	Rosetta + docking	0.283 ± 0.055	1.02 ± 0.18
25	F281C S283T	Rosetta + docking	0.015 ±0.005	0.05 ± 0.02
26	F281I S283V	Rosetta + docking	0.012 ± 0.004	0.05 ± 0.02
27	F281H S283I	Rosetta + docking	0.193 ± 0.043	0.70 ± 0.14
28	F281H S283V	Rosetta + docking	0.011 ± 0.003	0.04 ± 0.02
29	F281M S283I	Rosetta + docking	0.273 ± 0.048	0.98 ± 0.16
30	F281M S283T	Rosetta	0.282 ± 0.061	1.02 ± 0.20
31	F281M S283V	Rosetta + docking	0.252 ± 0.023	0.90 ± 0.09
32	F473W L475G	Rosetta + docking	0.022 ± 0.004	0.09 ± 0.02
33	F473W L475T	Rosetta + docking	0.017 ± 0.003	0.07 ± 0.03
34	F281M S283I F307H	Rosetta + docking	0.038 ± 0.015	0.20 ± 0.10
35	F281M S283T F307H	Rosetta + docking	0.014 ± 0.007	0.04 ± 0.03

nd – not detected

Table S4. Comparison of performance of whole-cell reactions in the conversion of isoeugenol in vanillin.

Enzyme	Origin	Cell dry weight (mg/mL)	Organic solvent % (v/v)	Substrate (mM)	Molar conversion yields (%)	Ref
NOV1-S283F	*Novosphingobium aromatocivorans*	1.3	3.5% Ethanol	100	>99	This study
IE27	*Pseudomonas putida*	15	10% DMSO	230	81	[19]
IEM720-F281Q	metagenome	34	10% DMSO	100	75	[20]
*Hs*CCO	*Herbaspirillum seropedicae*	18	10% glycerol	10	80	[21]
*Rb*CCO	*Rhodobacteraceae bacterium*	18	10% glycerol	10	75	[21]

Table S5. Binding energy (ΔG) and dihedral energy (ΔE) values in kcal/mol found for isoeugenol and resveratrol at the binding site of wild-type and S283F NOV1variant in the ensemble docking experiments.

	Isoeugenol					Resveratrol				
	$\Delta G_{binding}$		**$\Delta E_{dihedral}$**		**$\Delta\Delta E_{dihedral}$**	**$\Delta G_{binding}$**		**$\Delta E_{dihedral}$**		**$\Delta\Delta E_{dihedral}$**
variant	**WT**	**S283F**	**WT**	**S283F**	**(S283F-WT)**	**WT**	**S283F**	**WT**	**S283F**	**(S283F-WT)**
mean	-5.97	-5.96	9.24	10.25	1.01	-6.80	-6.98	14.55	18.87	4.32
std	0.62	0.50	3.69	4.16	0.48	1.34	0.38	5.21	5.54	0.33
min	-7.10	-6.84	0.92	2.27	1.35	-8.64	-7.52	3.92	8.77	4.85
25%	-6.37	-6.34	6.84	7.30	**0.46**	-7.64	-7.30	11.39	15.53	**4.14**
50%	-6.07	-5.93	9.43	9.97	**0.54**	-7.20	-7.03	14.10	19.69	**5.59**
75%	-5.64	-5.67	11.48	13.22	**1.74**	-6.30	-6.81	16.94	22.65	**5.71**
max	-2.77	-4.89	18.38	18.42	0.04	0.80	-6.24	31.35	28.45	-2.90

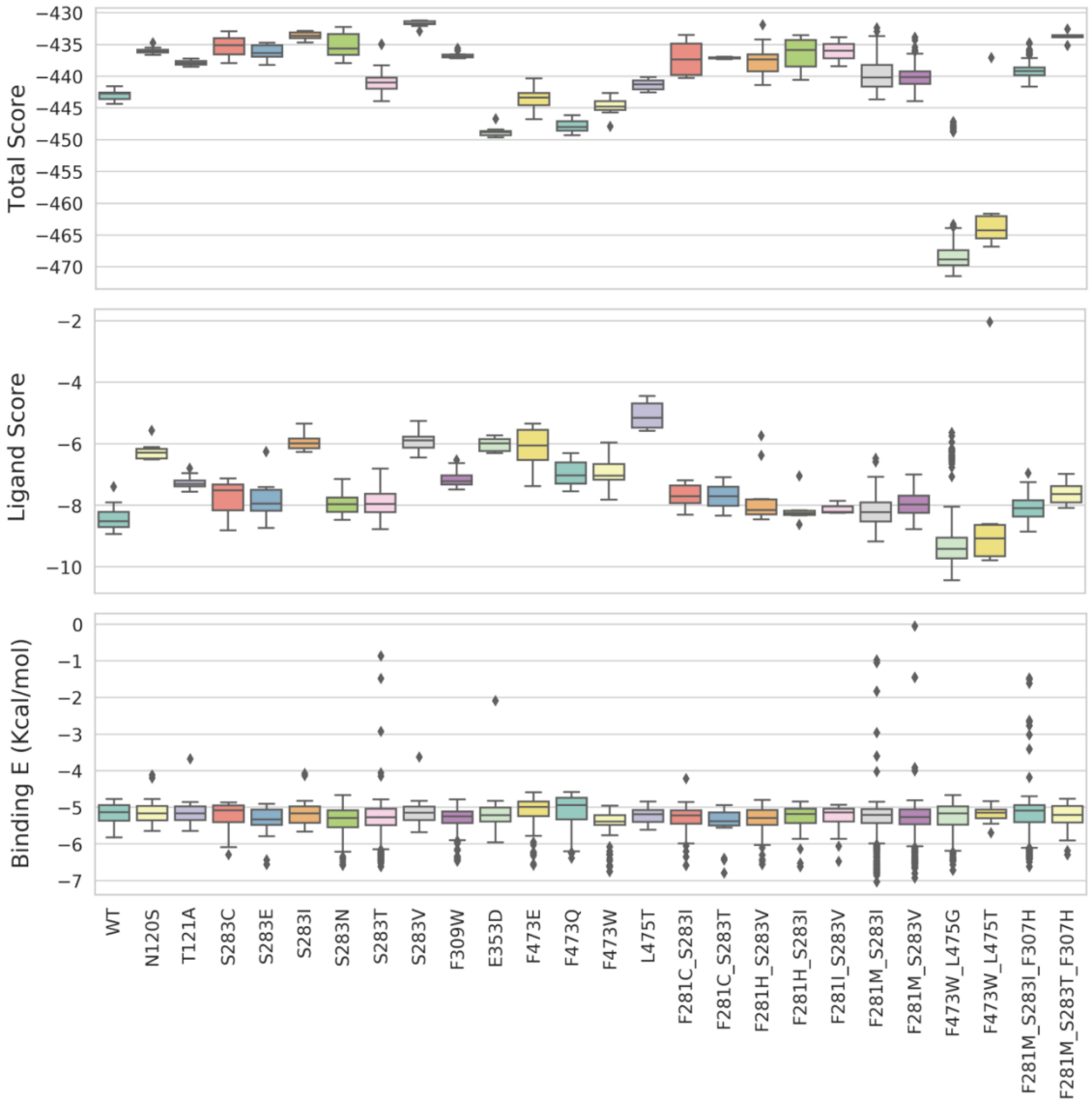

Fig. S1. Rosetta Ligand Score, Rosetta Total Score, and VINA Autodock binding energy values for the *in silico* variants selected for experimental validation.

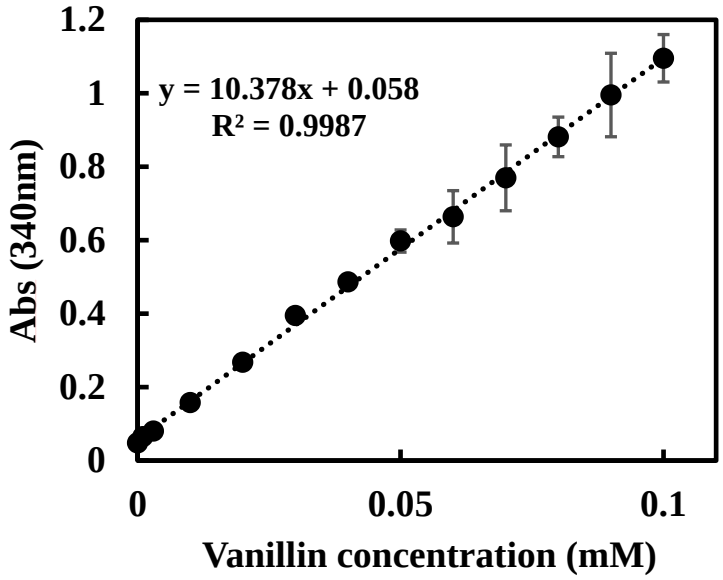

Fig. S2. Calibration curve for vanillin quantification. The Abs_{340nm} was measured for vanillin concentrations from 0 to 0.1 mM and the epsilon for vanillin ($\varepsilon_{vanillin}$) was determined by resorting to the Lambert-Beer law equation, $A = \varepsilon.l.c.$ $\varepsilon_{vanillin}$= 15.97 mL μmol^{-1} cm^{-1} (l = 0.65 cm).

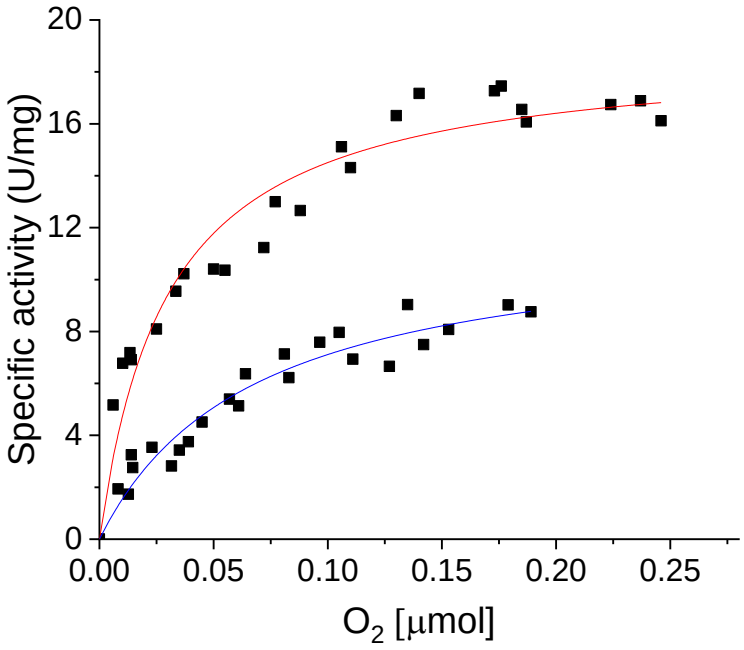

Fig. S3. Steady-state kinetic analyses of reactions of isoeugenol and O_2 catalyzed by wild-type (blue line) and S283F variant (red line). Reactions were started by adding the enzyme to a nitrogen-purged mixture containing 4 mM isoeugenol in 100 mM Tris-HCl, pH 9. Kinetic data were fitted directly to the Michaelis-Menten equation using the Origin® software.

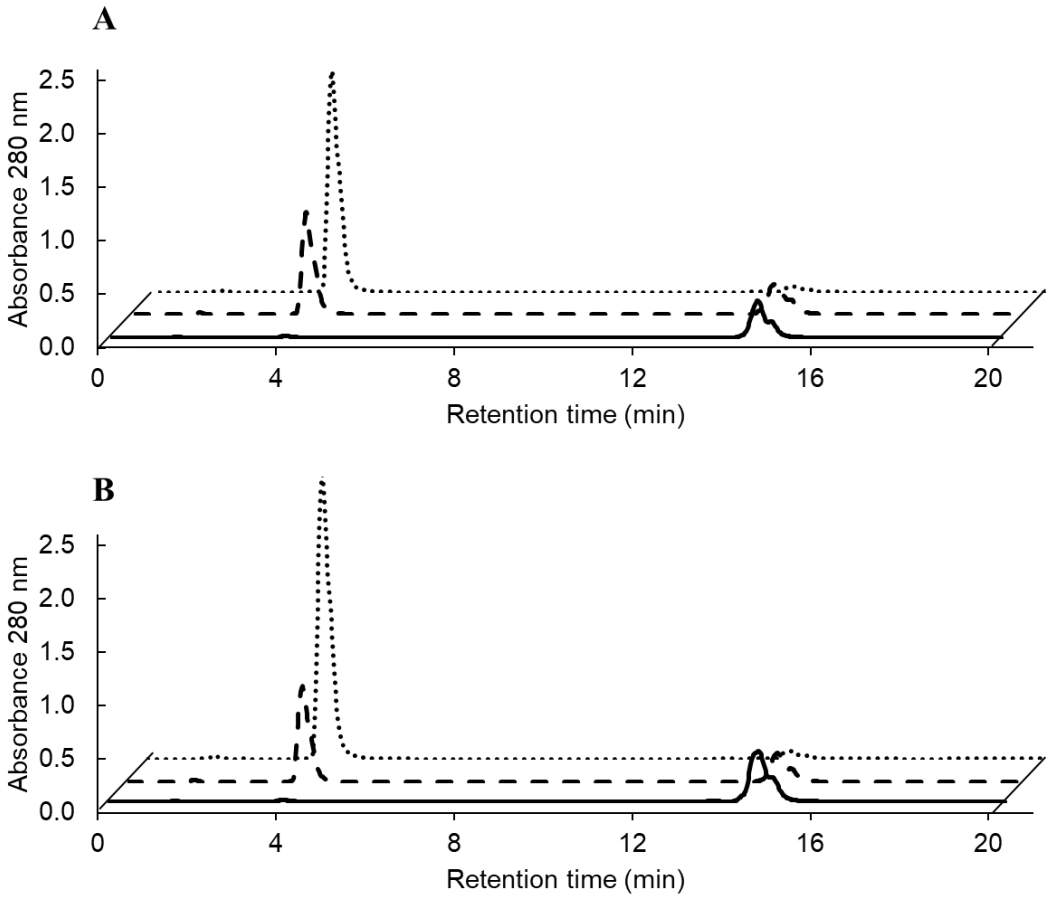

Fig. S4. HPLC-elution profile (280 nm) of reactions performed with purified wild-type (**A**) and S283F variant (**B**) preparations. Aliquots were taken at initial time (solid line) and after 15 min (dash line) and 90 min (dot line) of reaction. Retention time for vanillin and isoeugenol are 3.9 and 14.5 min, respectively.

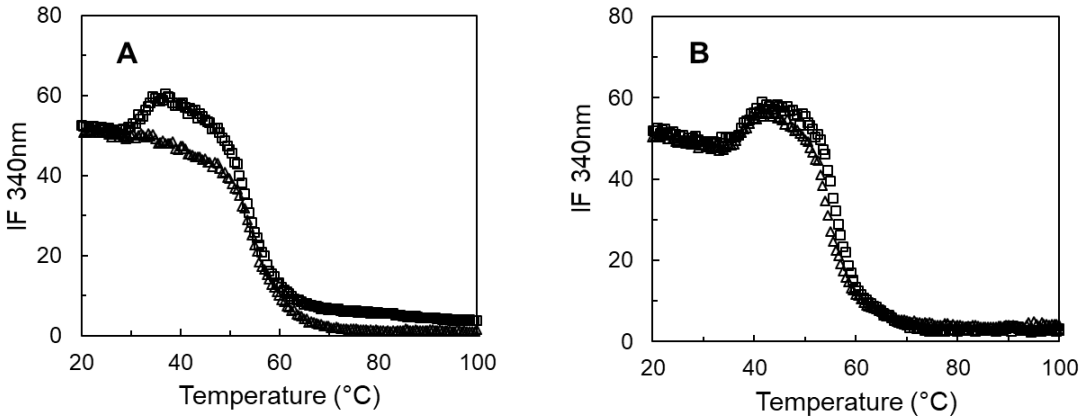

Fig. S5. Tryptophan fluorescence emission at 340 nm at increasing temperatures of wild-type **(A)** and S283F variant **(B)** in the absence (squares) and after incubation with 2000 equivalents of EDTA (triangles).

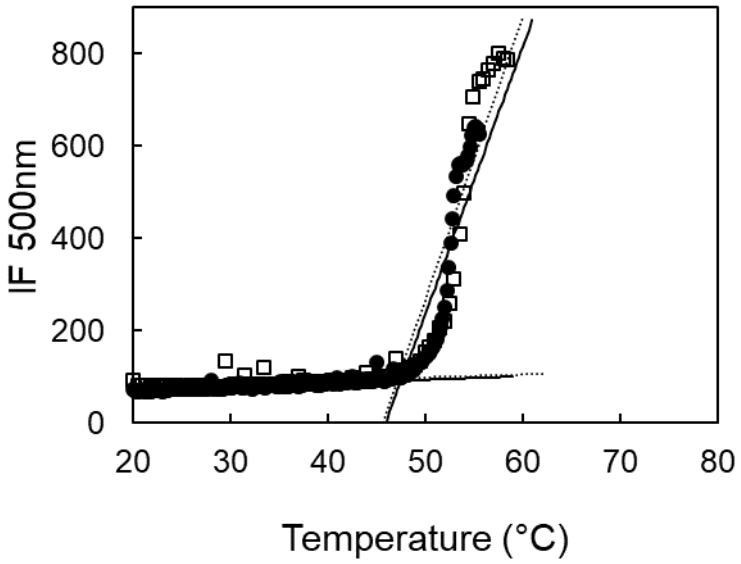

Fig. S6. Static light scattering at 500 nm (T_{agg} = 47°C) for wild-type (filled circles) and S283F variant (empty squares).

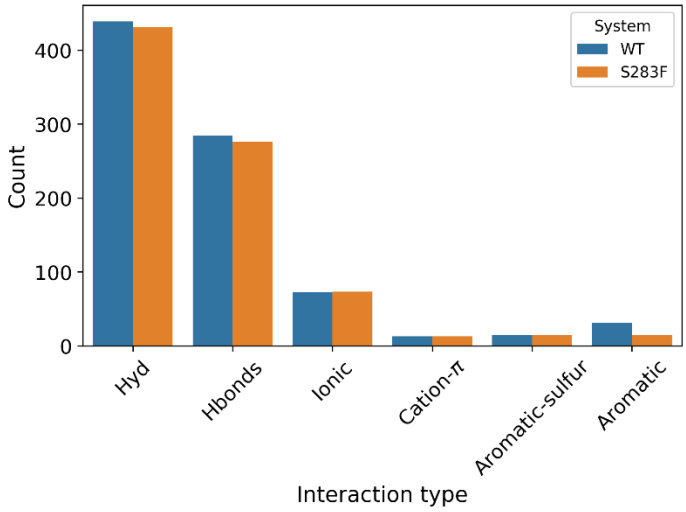

Fig. S7. Inter-residue interaction analysis was performed with PIC (Protein Interaction Calculator) using the X-ray structures of wild-type NOV1 (PDB: 55j5) and S283F variant (this work). Overall counting of interactions depicted by PIC in both wild-type NOV1and variant. It includes sidechain-sidechain, backbone-backbone, and sidechain-backbone interactions.

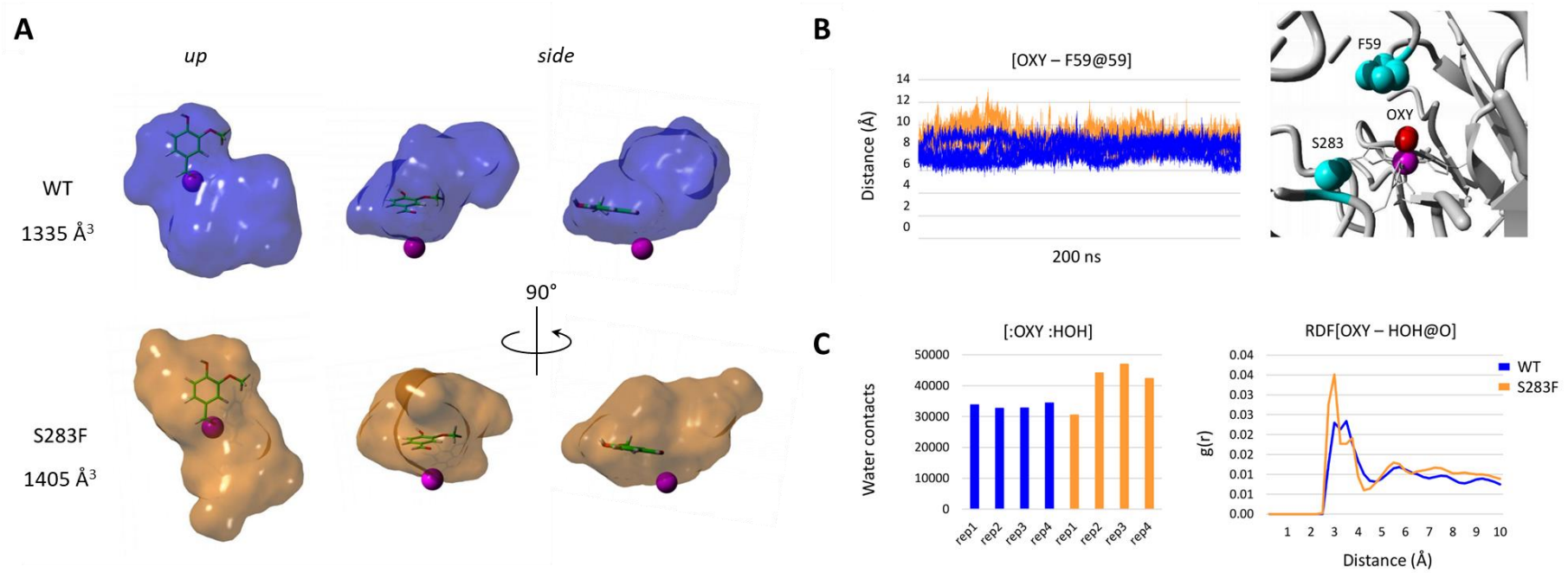

Fig. S8. (A) Empty volumes of the active site of the wild-type NOV1 (blue) and S283F variant (orange) calculated with MDpocket over 800 ns of MD simulation. The vanillin ligand found in the X-ray 5J55 is superposed. **(B)** Distance between the molecular oxygen bound to Fe and the center of mass of the ring of Phe59 along with the four replicas of 200 ns MD simulation. **(C)** Water counting inside the active site (left) and Radial Distribution Function (RDF) (right) of water molecules concerning the molecular oxygen during 800 ns of MD simulation.

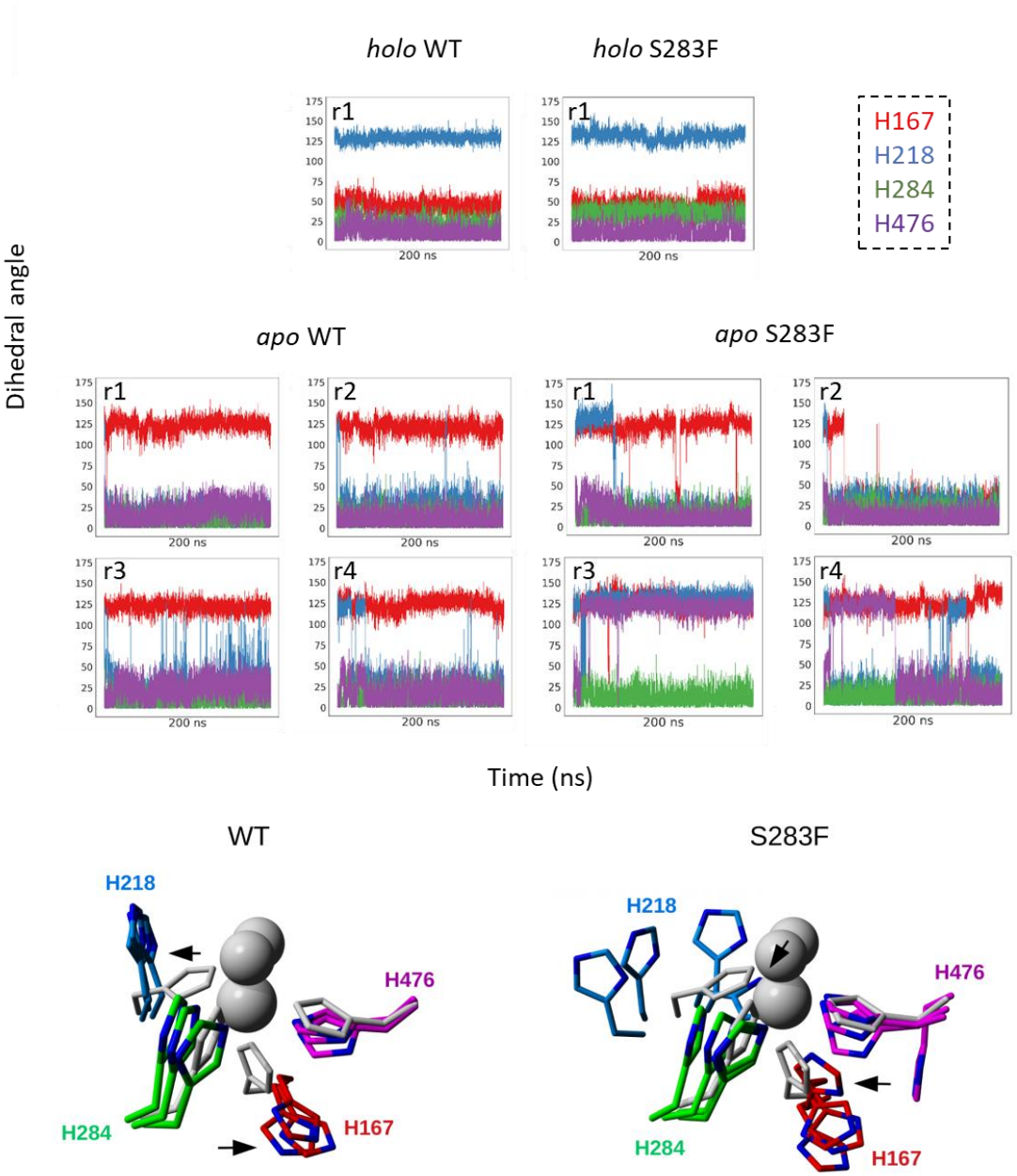

Fig. S9. On the top, dihedral angles [C-CA-CB-CG] define the orientation of the histidine 167 (red), 218 (blue), 284 (green), and 476 (purple) side chains, measured on the four replicas of 200 ns MD simulation. The values for the wild-type and mutant in their *holo* (Fe-bound) and *apo* (Fe-free) forms are represented. For simplicity, only one replica is shown for the *holo* form (which is consistent for the four replicas). The bottom illustrates the most representative clusters concerning the histidines conformations along with the MD simulation of the iron-free form of both WT and mutant. For comparison, these appear superposed with the catalytic center of the Fe(II)-O_2 bound state as it appears in the 5J55 X-ray structure.

REFERENCES

1. McAndrew, R. P.; Sathitsuksanoh, N.; Mbughuni, M. M.; Heins, R. A.; Pereira, J. H.; George, A.; Sale, K. L.; Fox, B. G.; Simmons, B. A.; Adams, P. D., Structure and Mechanism of NOV1, A Resveratrol-Cleaving Dioxygenase. *Proc Nat Acad Sci USA* **2016,** *113*, 14324-14329.
2. Krieger, E.; Dunbrack, R. L., Jr.; Hooft, R. W.; Krieger, B., Assignment Of Protonation States In Proteins And Ligands: Combining Pka Prediction With Hydrogen Bonding Network Optimization. *Methods Mol Biol* **2012,** *819*, 405-21.
3. Canutescu, A. A.; Dunbrack, R. L., Jr., Cyclic Coordinate Descent: A Robotics Algorithm For Protein Loop Closure. *Protein Sci* **2003,** *12*, 963-72.
4. Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C., Comparison Of Multiple Amber Force Fields And Development Of Improved Protein Backbone Parameters. *Proteins* **2006,** *65*, 712-725.
5. Krieger, E.; Vriend, G., New Ways To Boost Molecular Dynamics Simulations. *J Comput Chem* **2015,** *36*, 996-1007.
6. Berendsen, H. J. C.; Postma, J. P. M.; Vangunsteren, W. F.; Dinola, A.; Haak, J. R., Molecular-Dynamics with Coupling to an External Bath. *J. Chem. Phys.* **1984,** *81*, 3684-3690.
7. Mark, P.; Nilsson, L., Structure And Dynamics Of The TIP3P, SPC, And SPC/E Water Models at 298 K. *J Phys Chem A* **2001,** *105* (43), 9954-9960.
8. Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G., A Smooth Particle Mesh Ewald Method. *J Chem Phys* **1995,** *103*, 8577-8593.
9. Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M., LINCS: A Linear Constraint Solver For Molecular Simulations. *J Comput Chem* **1997,** *18*, 1463-1472.
10. Miyamoto, S.; Kollman, P. A., Settle - an Analytical Version of the Shake and Rattle Algorithm for Rigid Water Models. *J. Comput. Chem.* **1992,** *13*, 952-962.
11. Schmidtke, P.; Le Guilloux, V.; Maupetit, J.; Tuffery, P., fpocket: Online Tools For Protein Ensemble Pocket Detection And Tracking. *Nucleic Acids Res* **2010,** *38*, W582-9.
12. Le Guilloux, V.; Schmidtke, P.; Tuffery, P., Fpocket: an Open Source Platform For Ligand Pocket Detection. *BMC Bioinformatics* **2009,** *10*, 168.
13. Krieger, E.; Vriend, G., YASARA View-Molecular Graphics For All Devices-From Smartphones To Workstations. *Bioinformatics* **2014,** *30*, 2981-2982.
14. Schrodinger, L., The PyMOL Molecular Graphics System. **2010**.
15. Reback J., j., Wes McKinney W., Van den Bossche J., Augspurger T., Cloud P., Dong K., pandas-dev/pandas: Pandas 1.3.0 (Version v1.3.0). Zenodo. **2021**.
16. McKinney, W., Data Structures for Statistical Computing in Python. *Proc. Of The 9th Python In Science Conf* **2010**.
17. Harris, C. R.; Millman, K. J.; van der Walt, S. J.; Gommers, R.; Virtanen, P.; Cournapeau, D.; Wieser, E.; Taylor, J.; Berg, S.; Smith, N. J.; Kern, R.; Picus, M.; Hoyer, S.; van Kerkwijk, M. H.; Brett, M.; Haldane, A.; Del Rio, J. F.; Wiebe, M.; Peterson, P.; Gerard-Marchant, P.; Sheppard, K.; Reddy, T.; Weckesser, W.; Abbasi, H.; Gohlke, C.; Oliphant, T. E., Array Programming with NumPy. *Nature* **2020,** *585*, 357-362.
18. Waskom, M. L., seaborn: statistical data visualization. *J Open Source Software* **2021,** *6, 3021*.
19. Yamada, M.; Okada, Y.; Yoshida, T.; Nagasawa, T., Vanillin Production Using Escherichia Coli Cells Over-Expressing Isoeugenol Monooxygenase of *Pseudomonas putida*. *Biotechnol Lett* **2008,** *30*, 665-670.

20. Zhao, L.; Jiang, Y.; Fang, H.; Zhang, H.; Cheng, S.; Rajoka, M. S. R.; Wu, Y., Biotransformation of Isoeugenol into Vanillin Using Immobilized Recombinant Cells Containing Isoeugenol Monooxygenase Active Aggregates. *Appl Biochem Biotechnol* **2019,** *189*, 448-458.
21. Han, Z.; Long, L.; Ding, S., Expression and Characterization of Carotenoid Cleavage Oxygenases From Herbaspirillum seropedicae and Rhodobacteraceae bacterium Capable of Biotransforming Isoeugenol and 4-Vinylguaiacol to Vanillin. *Front Microbiol* **2019,** *10*, 1869.