Electronic Supplementary Information

Antioxidant canolol production from a renewable feedstock via an engineered decarboxylase

Krista L. Morley,^a Stephan Grosse,^a Hannes Leisch,^a and Peter C.K. Lau*^{ab}

^aAquatic and Crop Resource Development, National Research Council Canada, 6100 Royalmount Ave., Montreal, Quebec H4P 2R2, Canada;

^{a*}National Research Council Canada, 6100 Royalmount Avenue, Montreal, Quebec H4P 2R2, Canada. E-mail: <u>peter.lau@cnrc-nrc.gc.ca</u>; Fax: (514) 496-6265; Tel: (514) 496-6325 ^bMcGill University, Departments of Chemistry and Microbiology & Immunology, McGill University, Montreal, Quebec H3A 2B4, Canada, and FQRNT Centre in Green Chemistry and Catalysis, Montreal, Quebec, Canada

1. Protein Engineering of PAD

The X-ray crystal structure of PAD was solved recently.¹ A model of PAD containing FA (described in the experimental section) in the active site was used to select five residues for saturation mutagenesis: Tyr19, Val38, Val70, Ile85, and Phe87. Using NNK degeneracy (32 codons), 20 amino acids were encoded at each position. To ensure complete coverage of each library, 192 colonies from each transformation (960 colonies total) were picked into 96-well plates and screened for decarboxylation of SA. Enzyme activity was determined spectrophotometrically by monitoring the decrease in the level of

SA over time at 30 °C and 310 nm. Wild-type PAD was inactive towards SA (<1% conversion). No hits were obtained from the Tyr19, Val38, Val70, Phe87 libraries. However, screening of the Ile85 library identified 8 clones with activity towards SA. DNA sequencing of these clones revealed Ile85Ala (3 clones), Ile85Gly (2 clones), Ile85Ser (2 clones), Ile85Thr (1 clone) mutations. In 1-mL decarboxylation reactions of SA with the four mutants, Ile85Ala (SAD) showed the highest activity (Table S1) and was selected for purification and further characterization.

| Enzyme | SA ($U \text{ mg}^{-1}$) | FA (U mg ⁻¹) |
|----------|----------------------------|--------------------------|
| WT-PAD | n. d. | 5.8 |
| Ile85Ala | 1.4 | 2.3 |
| Ile85Gly | 0.9 | 0.9 |
| Ile85Ser | 0.4 | 2.9 |
| Ile85Thr | 0.5 | 4.4 |

Table S1. Crude Lysate Activities of Ile85 mutants and wild-type PAD

2. Kinetic Analysis of SAD and Wild-type PAD

The kinetic parameters (K_m and k_{cat}) of wild-type PAD and SAD for FA and SA were determined by measuring initial rates at 30 °C with varying substrate concentrations (1- 20 mM) using the double-reciprocal transformation (Lineweaver-Burk plot) of the Michaelis-Menten equation under steady-state conditions. Purified protein (see experimental section) was used at a concentration of 0.5 µg/mL. Due to the low solubility

of SA in aqueous solutions, dimethylsulfoxide (DMSO) was used as a cosolvent (5 vol%). Substrate and product were quantified by HPLC as described above.

Less than 1% conversion of SA was obtained with wild-type PAD and so kinetic parameters could not be obtained.

Experimental

1. Molecular modeling

Structure preparation. The crystal structures of three PADs, from *B. pumilus* (PDB code 3NAD), *B. subtilus* (PDB code 2P8G) and *L. plantarum* (PDB code 2GC9), were prepared for substrate docking and MD simulations. All structural manipulations were carried out in SYBYL 8.1.1 (Tripos, Inc., St. Louis, MO). Crystallographic water molecules, ions and co-solvents (if present) were removed, and selenomethionine residues (if present) were replaced by methionine residues. Scanning of a rotamer library repaired missing side chains, and hydrogen atoms were added according to protonation states at physiological pH. The protonation and tautomeric states of His residues, the rotameric state of Asn and Gln residues, and the position of rotatable polar hydrogen atoms were determined by visual examination of their respective structural context. All hydrogen atoms were energy-minimized with the AMBER FF94 force field² by keeping the nonhydrogen atoms at their crystallographic position. The 3D structure of FA, was generated and then energy-minimized with the MMFF94 force-field.³ Substrate docking. We applied a consensus docking paradigm, consisting of three different protein-ligand docking programs: SURFLEX^{4,5} as implemented in SYBYL 8.1.1, FITTED

3.0,^{6,7,8} and our in-house exhaustive docking program (H. Hogues, unpublished parametrization). Using the three available crystal structures of homologous PADs, docking in the rigid protein context was carried out in order to arrive at plausible docking solutions for three PAD substrates: p-coumaric acid, caffeic acid and FA. Putative binding sites in all available subunits were defined by manually placing one of the substrates (*p*-coumaric acid) in the β -barrel cavity. Default parameters were used for each docking program, except for those parameters that allow the sampled space to overflow outside of the cavity: threshold and bloat parameters of SURFLEX increased to 1.0 and 10, respectively, the Grid Sphere Size ProCESS parameter of FITTED was set to 20, and a grid extent definition of our in-house docking program to match closely the space searched with the other two programs. Four poses where consistently found by all docking programs. MD simulations. The four consensus binding modes generated in the protein-ligand docking stage, for all three substrates, were subjected to MD simulations in the dimeric structure of *B. pumilus* PAD (12 simulations total). Several crystallographic water O atoms where retained, most of them at the dimer interface, but also a few well-defined in the β -barrel cavity that are compatible with docked ligand for each binding mode and each substrate. The AMBER10 suite of programs^{9,10} was used to perform the MD simulations and trajectory analysis. The AMBER FF99SB force field^{11,12} was used for the protein and GAFF force field¹³ was used for the substrates. Solute atomic partial charges were calculated with the AM1BCC method^{14,15} implemented within QuACPAC (OpenEye, Inc., Santa Fe, NM). After 1000 iterations of energy minimization of the solute in vacuo, each dimeric-substrate complex was solvated in a truncated octahedron with TIP3P explicit water¹⁶ extending up to 12 Å around the solute, and 23 Na+ ions were added to maintain electroneutrality of the system. Applying harmonic restraints with force constants of 10 kcal mol-1 Å-2 to all solute atoms, the system was minimized first, followed by heating from 100 K to 300 K over 25 ps in the canonical ensemble (NVT), and by equilibrating to adjust the solvent density under 1 atm pressure over 25 ps in the isothermal isobaric ensemble (NPT) simulation. The harmonic restraints were then gradually reduced to zero with four rounds of 25 ps simulations. Each complex system was then simulated for 5 ns of NPT production run, with snapshots collected every 1 ps. A 2 fs time-step and a 9 Å non-bonded cutoff were used. The Particle Mesh Ewald (PME) method¹⁷ was used to treat long-range electrostatic interactions, and bond lengths involving bonds to hydrogen atoms were constrained by SHAKE.¹⁸ The final modeled complex structures were obtained by averaging over the last 3 ns MD trajectories, followed by 1000 steps of energy minimization in vacuo. Standard analyses of MD 4 trajectories were carried out with PTRAJ in AMBER 10. Protein-ligand binding affinities were estimated with SIETRAJ¹⁹ by calculating the average solvated interaction energy (SIE)²⁰ over 75 snapshots at 40 ps intervals over the last 3 ns of the MD trajectories

2. Saturation Mutagenesis

Saturation mutagenesis libraries were generated using the QuikChange II Site-Directed Mutagenesis Kit (Agilent Technologies,) with the pKK223-3 vector containing the cloned PAD-encoding gene [pKFAD]²¹ from *B. pumilus* UI-670 as a template. The five sets of NNK degenerate primers were:

Y19 (5'-gatttatacgtatgaaaatggctgggaannkgagatttacattaaaaacgatcacacca-3');

cY19 (5'-tggtgtgatcgtttttaatgtaaatctcmnnttcccagccattttcatacgtataaatc-3');

V38 (5'-gcattcatagcggaatgnnkggaggccgctgggttcg-3');

cV38 (5'-cgaacccagcggcctccmnncattccgctatgaatgc-3');

V70 (5'-acggaaccgactgggacagacnnktctttgaactttatgcctgag-3');

cV70 (5'-ctcaggcataaagttcaaagamnngtctgtcccagtcggttccgt-3');

I85 (5'-gagaaacgcatgcatggcgtcnnkttcttcccaaaatgggtgcat-3');

cI85 (5'-atgcacccattttgggaagaamnngacgccatgcatgcgtttctc-3');

F87 (5'-cgcatgcatggcgtcattttcnnkccaaaatgggtgcatgaacgt-3');

cF87 (5'-acgttcatgcacccattttggmnngaaaatgacgccatgcatgcg-3').

The plasmid library of saturation mutants were transformed into *E. coli* JM109, and plated on LB agar plates containing ampicillin (100 μ g/mL). 192 colonies from each transformation were picked into two 96-well plates containing 100 μ L of LB for overnight growth at 30 °C.

3. Enzyme Production and Screening in 96-Well Format

10 μ L of overnight culture described above was added to 1 mL of LB medium (100 μ g/mL ampicillin) in 96-well deep-well masterblocks (2 mL, Greiner Bio-One). The blocks were incubated for 6 hours at 30 °C with shaking at 225 rpm. Following centrifugation for 30 min (4 °C, 3600 rpm) the supernatant was removed and the cells were resuspended in 100 μ L of 10 mM sodium phosphate buffer (pH 7.0). The resuspended cells were centrifuged (10 min, 4 °C, 3600 rpm) and 5 μ L of the supernatant was transferred to each well of a UV-microplate for screening. To each well of the microplate was added 100 μ L of 0.5 mM SA (0.5% THF, 10 mM sodium phosphate, pH

7.0). Enzymatic decarboxylation was monitored spectrophotometrically by measuring the decrease in the level of SA over time at 30 °C and 310 nm. Cells expressing the wild-type PAD and also *E. coli* JM109 cells (without recombinant plasmid) were used as negative controls in the screening.

4.1 mL Decarboxylations (Confirmation of screening results)

Screening hits were expressed on 20 mL scale and the cells were resuspended in 2 mL of 10 mM sodium phosphate buffer (pH 7.0) for sonication. Total protein concentration of the cell lysates was determined by the Bradford method. 1 mL decarboxylation reactions were performed in 2 mL HPLC vials with 0.1 mg/mL of crude lysate, 4.5 mg/mL of SA (20 mM) in 100 mM sodium phosphate buffer (pH 7.0) containing 5 vol% THF. The reactions were incubated at 30 °C with shaking at 225 rpm. The reactions were stopped by the addition of 0.5 mL acetonitrile and precipitated protein was removed by centrifugation. Substrate and product were quantified by HPLC using the following system setup: Waters Millenium System 4.0, Luna 5µm C18(2) column (150 x13 4.6mm), PDA-detector (258 and 280 nm channels extracted), 0.6 ml/min flow rate at 40° C, mobile phase: CH3CN (A) and 0.05% (w/v) trifluoroacetic acid (B) using the following elution profile (A/B): 3 min at 10/90, 12 min linear gradient to 50/50, 6 min linear gradient to 100/0, 3 min at 100/0, 6 min linear gradient to 10/90 and reequilibration for 10 min at 10/90; retention time of SA: 16.0 min; retention time of VS: 21.4 min. One unit of enzyme activity was defined as 1 µmole product formation per minute (U) per mg of protein (U/mg) or mL of crude enzyme lysate (U/mL).

5. Protein Purification

All purification procedures were performed at 277 K on an ÄKTA Explorer 100 Air chromatography system (GE Healthcare, Baie d'Urfe, QC). The crude extract was loaded on a HiPrep DEAE-Sepharose FF column (16/10) equilibrated with 20 mM sodium phosphate buffer (pH 7.0) at a flow rate of 1.5 ml/min. The column was washed with the same buffer containing 0.15 M NaCl until no protein could be detected (A_{280}) nm) in the flow through, and the enzyme subsequently eluted with a linear gradient of 0.15 - 0.6 M NaCl. Active fractions were pooled, and solid ammonium sulfate was slowly added to give a final concentration of 30% (w/v). The enzyme-containing solution was then loaded on a Butyl-S-Sepharose 6 FF (16/10) column which had been previously equilibrated with 30% (w/v) ammonium sulfate in 20 mM sodium phosphate buffer (pH 7.0). PAD was eluted using a linear gradient of 30% to 0% (w/v) ammonium sulfate. Active fractions were collected, pooled and concentrated by ultrafiltration with a membrane exclusion size of 10 kDa in a 50 ml Amicon stir cell (Millipore, Billerica, MA) and applied to a HiLOAD Superdex 200 prep grade size exclusion column (16/60) which was previously equilibrated with 20 mM sodium phosphate buffer (pH 7.0) containing 150 mM NaCl. Protein was eluted with the same buffer (flow rate of 1.5 ml/min) and collected in 2 ml fractions. During each chromatographic step the protein profile was monitored by its absorbance at 280 nm. Protein concentration was determined as described by Bradford²² and alternatively by a Micro-Biuret method.²³ Bovine serum albumin was used as a protein standard.

6. GC/MS Analysis

VS was analyzed on an Agilent Technologies 7890A gas chromatograph coupled to a 5975C quadrupole mass spectrometer and a CTC Analytics Combipal autosampler (Figure S3). One microliter was injected under pulse split (1:10) conditions on a 30-m by 0.25-mm by 0.25-µm HP-5MS capillary column (Agilent). The temperature program was as follows: 70°C for 1.5 min, 65°C/min to 135°C, 5°C/ min to 280°C, and 15 min at 280°C. Helium was used as the carrier gas. The injector and the detector were adjusted to 250°C and 280°C. The mass selective detector (MSD) was run in the electron ionization (EI; 70-eV) scan mode between 45 and 500 atomic mass units (amu).



Figure S1. (A) ¹H-NMR spectrum of VS from enzymatic decarboxylation of commercial SA. (B) ¹H-NMR spectrum of VS from enzymatic decarboxylation of SA from canola meal.



Figure S2. GC/MS spectrum of VS from enzymatic decarboxylation of commercial SA.



Figure S3. HPLC chromatogram of VS from enzymatic decarboxylation of SA from canola meal.

Inhibition Studies of SAD by VS.

For the enzymatic decarboxylation with SAD crude lysate in aqueous buffer, higher SA concentrations than 5 g/L led to incomplete reactions. Previous experiments with wild-type PAD indicated that VG, the decarboxylation product of FA, inhibited the enzyme even at low concentrations (0.25 g/L). To determine if VS was inhibiting SAD, VS was added in different concentrations (0.1, 0.25, 0.5, 1 g/L) to the enzyme assay and the conversion of SA (5 g/L) was determined by HPLC. Similarly, we observed a significant decrease in activity by adding only 0.25 g/L VS. At a VS concentration of 1 g/L, only 34% of residual activity was detected (Figure S4).



Figure S4. The relative activity of SAD crude lysate and the inhibitory effect of added VS.



Figure S5. DSC analysis of PVS polymer



Figure S6. Thermogravimetric analysis (TGA) of PVS polymer





| # indent | E (GPa) | H (GPa) |
|----------|---------------|-----------------|
| 1 | 5.23 | 0.283 |
| 2 | 5.26 | 0.286 |
| 3 | 5.18 | 0.28 |
| 4 | 5.21 | 0.285 |
| 5 | 5.12 | 0.277 |
| 6 | 5.28 | 0.288 |
| 7 | 5.21 | 0.282 |
| 8 | 5.21 | 0.287 |
| 9 | 5.27 | 0.284 |
| 10 | 5.25 | 0.284 |
| Average | 5.22 +/- 0.05 | 0.283 +/- 0.003 |

Figure S7. Nanoindentation characterization of PVS polymer. H and E are calculated throughout the displacement into the sample (at each unload of the oscillating tip). The value reported for each test is an average of the values in the range of 1200 and 1800 nm. Poisson coefficient used for calculation: 0.3

⁸ C. R. Corbeil and N. Moitessier, J. Chem. Inf. Model., 2009, 49, 997.

¹⁰ D. A. Case, T. Darden, T. E. Cheatham, C. Simmerling, J. Wang, R. E. Duke, R. Luo,

- M. Crowley, R.C. Walker, W. Zhang, K. M. Merz, B. Wang, S. Hayik, A. Roitberg, G.
- Seabra, I. Kolossvary, K. F. Wong, F. Paesani, J. Vanicek, X. Wu, S. R. Brozell, T.

¹¹ J. Wang, P. Cieplak and P.A. Kollman, J. Comput. Chem., 2000, **21**, 1049.

¹³ J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman and D. A. Case, *J. Comput. Chem.*, 2004, **25**, 1157.

¹⁶ W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, J.

Chem. Phys., 1983, **79**, 926.

¹⁷ T. Darden, D. York and L. Pedersen, J. Chem. Phys., 1993, **98**, 10089.

¹⁸ J. P. Ryckaert, G. Ciccotti and H. J. C. Berendsen, J. Comput. Phys., 1977, 23, 327.

¹⁹ Q. Cui, T. Sulea, J. D. Schrag, C. Munger, M. N. Hung, M. Naim, M. Cygler and E. O. Purisima, *J. Mol. Biol.*, 2008, **379**, 787.

²⁰ M. Naim, S. Bhat, K. N. Rankin, S. Dennis, S. F. Chowdhury, I. Siddiqi, P. Drabik, T. Sulea, C. I. Bayly, A. Jakalian and E.O. Purisima, *J. Chem. Inf. Model.*, 2007, **47**, 122.

²¹ J. Yang, S. Wang, M.-J. Lorrain, D. Rho, K. Abokitse and P. C. K. Lau, *Appl.*

Microbiol. Biotechnol., 2009, 84, 867.

²² M. M. Bradford, *Anal. Biochem.*, 1976, **72**, 248.

¹ A. Matte, S. Grosse, H. Bergeron, K. Abokitse and P. C. K. Lau, *Acta Cryst.*, 2010, **F66**, 1407.

² W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman, *J. Am. Chem. Soc.*, 1995, **117**, 5179.

³ T. A. Halgren, J. Comput. Chem., 1999, **20**, 730.

⁴ A. N. Jain, J. Med. Chem., 2003, 46, 499.

⁵ A. Jain, J. Comput. Aided Mol. Des., 2007, **21**, 281.

⁶ C. R. Corbeil, P. Englebienne and N. Moitessier, J. Chem. Inf. Model., 2007, 47, 435.

⁷ C. R. Corbeil, P. Englebienne, C. G. Yannopoulos, L. Chan, S. K. Das, D. Bilimoria, L. Heureux and N. Moitessier, *J. Chem. Inf. Model.*, 2008, **48**, 902.

⁹ D. A. Case, T. E. Cheatham, T. Darden, H. Gohlke, R. Luo, K. M. Merz, A. Onufriev, C. Simmerling, B. Wang and R. J. Woods, *J. Comput. Chem.*, 2005, **26**, 1668.

Steinbrecher, H. Gohlke, L. Yang, C. Tan, J. Mongan, V. Hornak, G. Cui, D. H.

Mathews, M. G. Seetin, C. Sagui, V. Babin and P.A. Kollman, *AMBER 10*, University of California, San Francisco, 2008.

¹² V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg and C. Simmerling, *Proteins*, 2006, **65**, 712.

 ¹⁴ A. Jakalian, B. L. Bush, D. B. Jack and C. I. Bayly, *J. Comput. Chem.*, 2000, 21, 132.
¹⁵ A. Jakalian, D. B. Jack and C. I. Bayly, *J. Comput. Chem.*, 2002, 23, 1623.

Electronic Supplementary Material (ESI) for Green Chemistry This journal is C The Royal Society of Chemistry 2013

²³ R. F. Itzhaki and D. M. Gill, *Anal. Biochem.*, 1964, **121**, 401.