Supporting Information

Structural evidence for the dopamine-first mechanism of norcoclaurine synthase

Benjamin R. Lichman,[○†#] Altin Sula,[○‡] Thomas Pesnot,[§] Helen C. Hailes,[§] John M. Ward,[†] and Nicholas H. Keep[‡*]

[†]Department of Biochemical Engineering, University College London, Gower Street, London WC1E 6BT, UK

[‡] Institute for Structural and Molecular Biology, Department of Biological Sciences, Birkbeck University of London, Malet Street, London WC1E 7HX, UK

[§]Department of Chemistry, University College London, Christopher Ingold Building, London WC1H 0AJ, UK

Present Address: [#]John Innes Centre, Norwich Research Park, Norwich NR4 7UH, UK.

[○]B.R.L. and A.S. contributed equally.

Contents

Experimental Procedures

Protein purification and expression

A construct containing a codon optimised, truncated *Thalictrum flavum* NCS[1] gene (ΔN33C196*Tf*NCS), with an N-terminal hexahistidine tag and a TEV protease cleavage site was synthesised and cloned into pD451-SR (ATUM, CA, USA)[2]. The plasmid was transformed into BL21 (DE3) cells and a single colony inoculated 100 ml of Terrific broth media (TB) for 16 hours. One litre of TB was inoculated with 4% v/v of overnight culture and grown for 2 hours at 37 °C, then 1 hour at 25 °C. The protein was overexpressed by addition of 0.5 mM isopropylthiogalactoside, incubated for 3 hours at 25 °C and then harvested by centrifugation.

Cell pellets were suspended in binding buffer (50 mM Hepes, 100 mM NaCl, 20 mM Imidazole pH 7.5) and 10% v/v BugBuster 10X (Merck Millipore, Germany) was used to break the cells. After centrifugation at 25,000 g for 1 hour, the lysate was loaded onto 1 ml of Ni-Sepharose HP resin (GE Healthcare). The protein was eluted from the resin with elution buffer (50 mM Hepes, 100 mM NaCl, 500 mM imidazole, pH 7.5) after washing with binding buffer and washing buffer (50 mM Hepes, 100 mM NaCl, 50 mM Imidazole pH 7.5) for 5 column volumes respectively. The eluted fractions were pooled and 0.1 mg of TEV protease (containing a N-terminal His-tag) was added to the sample and dialysed in 4 litres of dialysis buffer (20 mM Tris, 50 mM NaCl, pH 7.5) for 16 hours at 4 °C.

The sample was loaded to a 1 ml of Ni-Sepharose HP resin to bind uncut NCS and TEV protease. Cut NCS was washed off the resin with wash buffer (20 mM Tris, 50 mM NaCl, 50 mM Imidazole, pH 7.5). Size exclusion chromatography was used to purify the NCS protein further using Superdex 75 16/600 column (GE, Healthcare). The eluents were pooled and concentrated using a 10 kDa cut off Vivaspin concentrator (Sartorius, Germany) to 12 mg/ml. The protein sample was either used directly to set up crystallization trials or stored at -80 °C.

Protein crystallisation and data processing

The truncated NCS *apo* protein crystals were grown by the sitting-drop method in 96-well crystallisation plates (Molecular Dimensions) in 10% w/v polyethylene glycol (PEG) 1000 and 10% w/v PEG 8000. Larger crystals were obtained by hanging-drop method. The protein was incubated with 10 mM of mimic compound **6** and crystallised in the same condition as the *apo* protein. The crystals were cryo-protected in crystallisation buffer containing 20% ethylene glycol. Diffraction data for the *apo* structure were collected at Soleil beamline Proxima 1 whereas the final mimic-bound dataset was collected at Diamond beamline I02. The diffraction images were processed using xia2 and XDS[3] software packages, scaled and merged using Aimless in the CCP4 program suite[4]. The initial phases of the apo NCS models were solved by molecular replacement with the program Phaser[5] using the previous apo NCS structure (PDB: 2VNE[6]) as the search model. Model building was performed with COOT[7] and refinement was done with Refmac5[8] using TLS (one group per chain including the associated water molecules and ligand) and local non-crystallographic symmetry restraints. The positions of both aromatic rings of the mimic was clear in all three copies in the asymmetric unit of the mimic-bound structure from initial difference maps. There was a ring like density next to the dopamine ring despite there being no ring closure in the mimic. When the mimic is placed in the conformation proposed to be productive in the reaction mechanism refinement gave strong (>5 sigma) difference density where a 6th atom could make up a second ring. Conversely if the non-productive conformation where the dopamine is flipped and the rest of the molecule comes off the other side of the ring, is refined alone there is even stronger difference density where the C9 atom is in the first conformation. Neither of these positions correspond to water molecules in the *apo* structure probably ruling

out a mixed apo/ligand structure[9]. We propose that the structure is a mixture of these productive and unproductive ligand conformations.

Alternative ligands were tried: neither a five-membered ring oxidation product nor the product of the typical enzyme reaction gave plausible fits ruling out any structure with the R group coming from an atom adjacent to the dopamine. A tertiary amine fills the density but gives poorer R factors than the two-conformation fit, and such a compound is also chemically implausible in the conditions used. Placing a water in the difference density gives a lower Rfree than the two-conformation model and no difference density. However, the water is too close to the Nitrogen (1.6 Å) and the ring (1.8 Å) and is only on the very edge of the density.

The final deposited model used the 'complete' occupancy refinement in Refmac5 such that the combined occupancy of the two ligands are constrained to 1.0 in each copy. This final occupancy is not particularly stable and depends on slight drift apart of the two ligands during refinement. The unproductive conformation often ends up with a lower occupancy and a higher B factor and can drift to very low occupancy and high B factor and move quite far out of the density resulting in a return of the difference density peak. Conversely more even occupancy results when the dopamine ring of the unproductive conformation moves away from the optimum individual fit to the ring allowing the ring linking atoms in minor conformation to be closer to the position of the major conformation. This results in less difference density and has been deposited. Other refinement packages did not give better results for the two ligand model in our hands.

Data collection and refinement statistics are summarized in Table S1. Figures and RMSD comparisons were performed using UCSF-Chimera (http://www.rbvi.ucsf.edu/chimera/) except the electron densities which were drawn with ccp4mg.

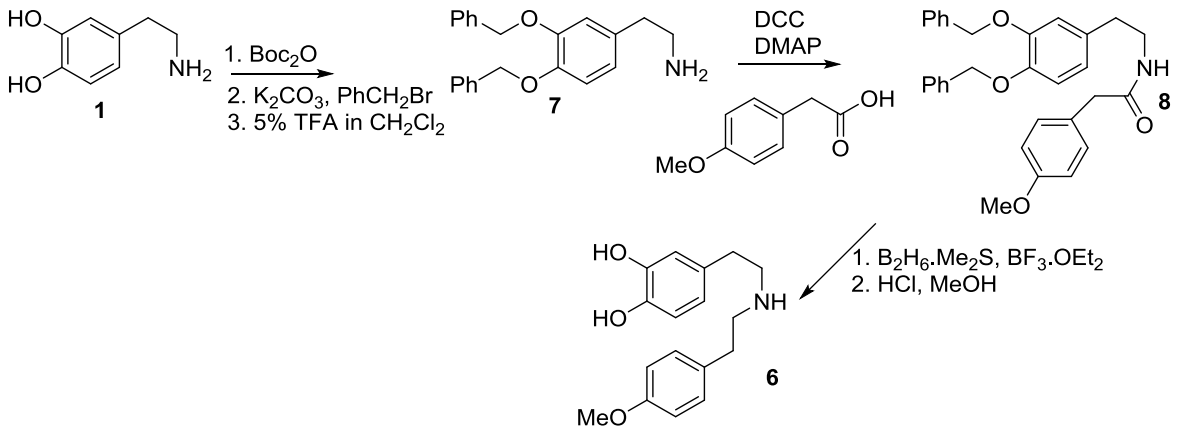
Computational docking

Subunit A of mimic-bound structure 5NON was used for docking experiments. Ligands and water molecules were removed. Ligands were MM2 energy minimised in ChemBio3D before docking with Chimera UCSF, using the AutoDock Vina plug-in[10]. The protein molecule was centred, and the docking box was position (-17.95, -7.43, 16.19) and size (18.14, 19.02, 28.24). The software was run with the settings: energy-range 3, exhaustiveness 8 and number of modes 10. Binding modes relevant to the dopamine-first mechanism were selected (see Table S2).

Enzyme assays

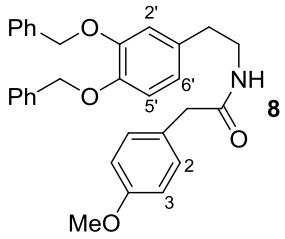
The time-courses of ΔN33C196*Tf*NCS and Δ29*Tf*NCS (Figure S2) was conducted in triplicate. Each assay contained 10 mM dopamine, 10 mM hexanal, 10% v/v MeCN, 0.1 mg/mL purified enzyme, 5 mM sodium ascorbate and 50 mM Hepes pH 7.5. Samples were quenched with 100 mM HCl, diluted and analysed by HPLC. Enzyme activities (initial rates) for Δ29*Tf*NCS and Δ29*Tf*NCS-A79I (Figure S3) were conducted in triplicate as previously reported[11]. Reactions contained dopamine (10 mM) and 4-HPAA or hexanal (2.5 mM) and were quenched after 30 seconds and analysed by HPLC. HPLC analyses were performed on a HPLC system consisting of an LC Packing FAMOS Autosampler, a P680 HPLC Pump, a TCC-100 Column oven and a UVD170U Ultraviolet detector (Dionex, Sunnyvale, CA, USA), and a C18 (150 x 4.6 mm) column (ACE, Aberdeen, UK). Samples were run with a gradient of H_2O (0.1% trifluoroacetic acid) /MeCN from 9:1 to 3:7 over 6 min, at a flow rate of 1 $mL.min^{-1}$. The column temperature was 30 °C, and compounds were detection by monitoring A280. Retention times and concentrations were calculated based on chemically verified standards.

Synthesis of 4-{2-[(4-Methoxyphenethyl)amino]ethyl}benzene-1,2-diol



General. All chemicals were obtained from commercial suppliers and used as received unless otherwise stated. Thin layer chromatography (TLC) analysis was performed on Merck Kieselgel precoated aluminium-backed silica gel plates and compounds visualised by exposure to UV light, potassium permanganate or ninhydrin stains. Flash column chromatography was carried out using silica gel (particle size 40-60 µm). NMR: ^{1}H and ^{13}C NMR spectra were recorded at 298 K at the field indicated using Bruker Avance 300 and Brucker Avance 400 III spectrometers. Coupling constants (*J*) are measured in Hertz (Hz) and multiplicities for ^{1}H NMR couplings are shown as s (singlet), d (doublet), t (triplet) q (quartet) and m (multiplet). Chemical shifts (in ppm) are given relative to tetramethylsilane and referenced to residual protonated solvent. Mass spectrometry analyses were performed at the UCL Chemistry Mass Spectrometry Facility using a Finnigan MAT 900 XP and Waters LCT Premier XE ESI Q-TOF mass spectrometers. 3,4-Bis(benzoyloxy)dopamine **7** was synthesized as previously reported[12].

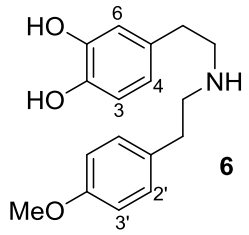
***N*-[3,4-Bis(benzyloxy)phenethyl]-2-(4-methoxyphenyl)acetamide 8**



3,4-Bis(benzoyloxy)dopamine **7** (200 mg, 0.60 mmol), 4-methoxyphenyl acetic acid (125 mg, 0.75 mmol), dicyclcohexyl carbodiimide (DCC) (193 mg, 0.94 mmol), and 4-dimethylaminopyridine (6 mg, 0.05 mmol) were stirred in dichloromethane (5 ml) for 18 h. The solution was filtered to remove the urea formed and the product was purified by silica chromatography (ethyl acetate/hexane, 1:2) to give **8** (200 mg, 69%) as a colourless oil. See Figure S7 for NMR spectra.

^{1}H NMR (300 MHz; $CDCl_3$) δ 2.61 (2H, t, *J* 6.7 Hz, CH_2CH_2Ar), 3.37 (2H, q, *J* 6.7 Hz, CH_2NH), 3.44 (2H, s, $ArCH_2CO$), 3.76 (3H, s, OMe), 5.09 (2H, s, $PhCH_2O$), 5.13 (2H, s, $PhCH_2O$), 5.29 (1H, br s NH), 6.56 (1H, dd, *J* 8.2 and 2.0 Hz, 5'-H), 6.68 (1H, d, *J* 2.0 Hz, 2'-H), 6.79 (1H, d, *J* 8.2 Hz, 6-H), 6.81 (2H, d, *J* 8.6 Hz, 2 x 3-H), 7.04 (2H, d, *J* 8.6 Hz, 2 x 2-H), 7.32-7.45 (10H, m, 2 x Ph); ^{13}C NMR (75 MHz; $CDCl_3$) δ 34.6, 40.2, 42.6, 54.9, 71.0, 71.1, 114.1, 114.7, 115.0, 115.3, 117.6, 121.3, 126.96 and 127.04, 127.46 and 127.49, 128.2, 130.2, 131.4, 131.7, 176.8; *m/z* (EI) 481($[M]^+$, 25%), 316 (49), 121 (27), 91 (100); HRMS $[M]^+$ calcd. for $C_{31}H_{31}NO_4$, 481.2248; found 481.2253.

4-{2-[(4-Methoxyphenethyl)amino]ethyl}benzene-1,2-diol 6



6

The reaction was carried out under anhydrous conditions. To amide **8** (250 mg, 0.52 mmol) in THF (20 mL), boron trifluoride etherate (33 μL, 0.26 mmol) was added and the solution heated at reflux for 10 min. Borane dimethylsulfide complex (2 M in THF; 780 μL, 1.56 mmol) was added dropwise and the reaction heated at reflux for 3 h. The reaction was cooled to 0 °C, 10% HCl solution (7 ml) was added and the reaction was stirred for 1 h at 0 °C and 2 h at room temperature. The solution was adjusted to pH 13 using NaOH and the product extracted with dichloromethane (3 x 10 mL), dried (Na_2SO_4) and concentrated to give *N*-[3,4-bis(benzyloxy)phenethyl]-2-(4-methoxyphenyl)ethan-1-amine, as a colourless oil (90 mg, 37%) which was taken directly through to the next step.

To *N*-[3,4-bis(benzyloxy)phenethyl]-2-(4-methoxyphenyl)ethan-1-amine (90 mg, 0.19 mmol) in methanol (10 ml) was added concentrated hydrochloric acid (2 ml). The reaction was heated at reflux for 24 h, water was added (2 ml) and the pH adjusted to 6 (NaOH solution). Solvents were removed *in vacuo* and the product (retention time 12.5 minutes) was purified by preparative HPLC to give **6** (55 mg, 99%).

Preparative HPLC conditions: Varian Prostar instrument with a UV-visible detector (monitoring at 280 nm) and a DiscoveryBIO wide Pore C18-10 Supelco column (25 Å~ 2.12 cm). A gradient of 5% to 90% of acetonitrile/water (0.1% TFA)) was used. See Figure S8 for NMR spectra.

^{1}H NMR (400 MHz; CD_3OD) δ 2.83 (2H, t, *J* 5.2 Hz, CH_2CH_2Ar), 2.91 (2H, t, *J* 5.2 Hz, CH_2CH_2Ar), 3.18-3.20 (4H, m, 2 x CH_2NH), 3.76 (3H, s, OMe), 6.57 (1H, dd, *J* 5.2 and 1.2 Hz, 4-H), 6.69 (1H, d, *J* 1.2 Hz, 6-H), 6.73 (1H, d, *J* 5.2 Hz, 3-H), 6.88 (2H, d, *J* 6.0 Hz, 2 x 3-H), 7.16 (2H, d, *J* 6.0 Hz, 2 x 2-H); ^{13}C NMR (100 MHz; CD_3OD) δ 32.5. 32.8, 50.2, 50.3, 55.7, 115.4, 116.75 and 116.79, 121.0, 128.9, 129.5, 130.8, 145.6, 146.8, 160.4; *m/z* (ES+) 288 ($[MH]^+$, 100%); HRMS (*m/z*) $[MH]^+$ calcd. for $C_{17}H_{22}NO_3$, 288.1600; found 288.1599.

Supplemental Figures and Tables

Table S1. X-ray data collection and refinement statistics.

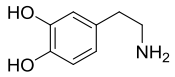
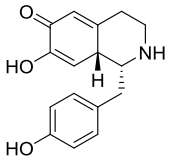
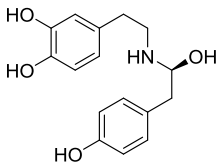
Data collection	Apo (5N8Q)	Mimic bound (5NON)
Space group	$P22_12_1$	$P22_12_1$
Unit-cell parameters		
a, b, c (Å)	38.05, 109.63, 136.74	38.31, 110.22, 136.90
α, β, γ (º)	90.0, 90.0, 90.0	90.0, 90.0, 90.0
Resolution range (Å)	109.6-2.0 (2.05-2.00)*	85.85-1.85 (1.89-1.85)
Total number of observation	211434 (15606)	429583 (27284)
Total number unique	39359 (2839)	50659 (3072)
Completeness	99.3 (99.7)	100.0 (100.0)
Multiplicity	5.4 (5.5)	8.5 (8.9)
<I/σ(I)>	6.1 (1.7)	11.2 (1.8)
$CC_{1/2}$	0.986 (0.725)	0.998 (0.839)
R_{merge}	0.136 (0.466)	0.087 (0.972)
R_{pim}	0.070 (0.535)	0.032 (0.350)
Molecule per ASU	3	3
Refinement		
Resolution Range (Å)	85.68-2.0 (2.05-2.00)	86.0-1.85 (1.9-1.85)
R_{work}	0.199 (0.259)	0.189 (0.267)
R_{free}	0.236 (0.299)	0.224 (0.295)
Reflection, working	37264	48057
Reflection, free	1960	2535
Average B factor	35.5	36.9
Rmsd bond angle	2.078	2.793
Rmsd bond length (Å)	0.022	0.032
Ramachandran plot		
Preferred region (%)	97.7	98.0
Allowed region (%)	2.3	2.0
Outliers (%)	0	0

$R_{merge}=\Sigma(I-\langle I\rangle)/\Sigma\langle I\rangle$

$R_{work}=\Sigma(|F_{obs}|-|F_{calc}|)/\Sigma|F_{obs}|$ for 95% of the data. R_{free} is the same definition but for the 5% of the data excluded from refinement.

*Values in parentheses are for highest-resolution shell.

Table S2. Computational docking output (Autodock Vina).

Ligand	Structure	Rank	Energy (kcal/mol)	Figure
Dopamine	5NON	1	-5.3	3C
		2	-5.2	
		3	-5.1	3D
		4	-5.1	
		5	-5.1	
		6	-4.8	
		7	-4.7	
		8	-4.5	
		9	-4.4	
		10	-4.4	
Quinone	5NON	1	-8.2	
		2	-7.8	
		3	-7.5	4C
		4	-7.2	
		5	-7.0	
		6	-7.0	
		7	-6.9	
		8	-6.8	
		9	-6.8	
		10	-6.7	
(*S*)-hemiaminal	5NON	1	-7.1	4D
		2	-7.1	
		3	-7.0	
		4	-6.8	
		5	-6.6	
		6	-6.6	
		7	-6.5	
		8	-6.4	
		9	-6.4	
		10	-6.3	

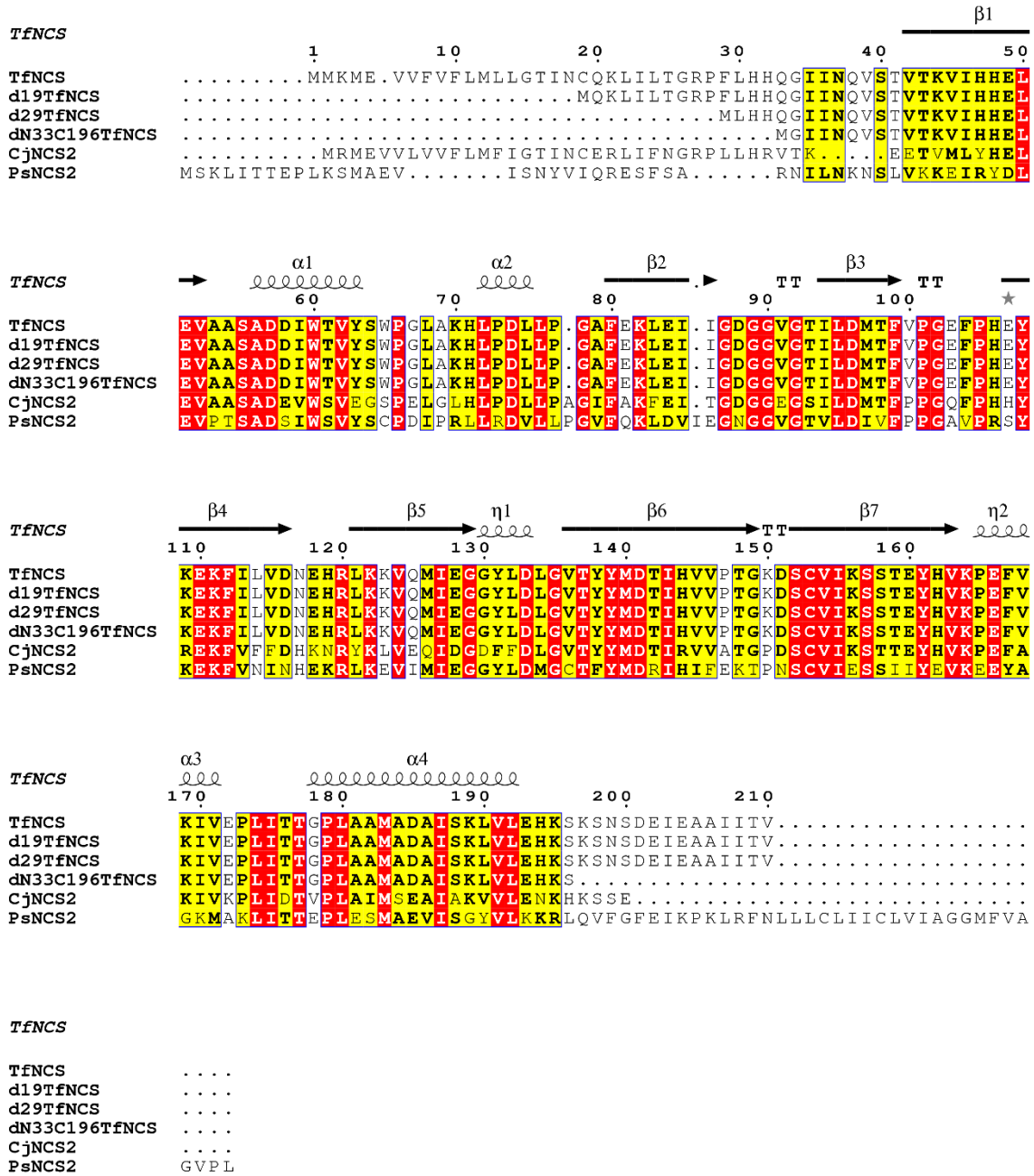


Figure S1. Multiple sequence alignment of NCS sequences. Alignment performed with ClustalOmega[13] and visualised with ESPript[14]. Amino acid numbering in the paper is relative to full length *Tf*NCS. NCS sequences used: *Tf*NCS[1], *Cj*NCS2[15], *Ps*NCS2[16]. Secondary structure elements derived from 5N8Q. For NCS sequence alignment with more proteins, see Li *et al*[17].

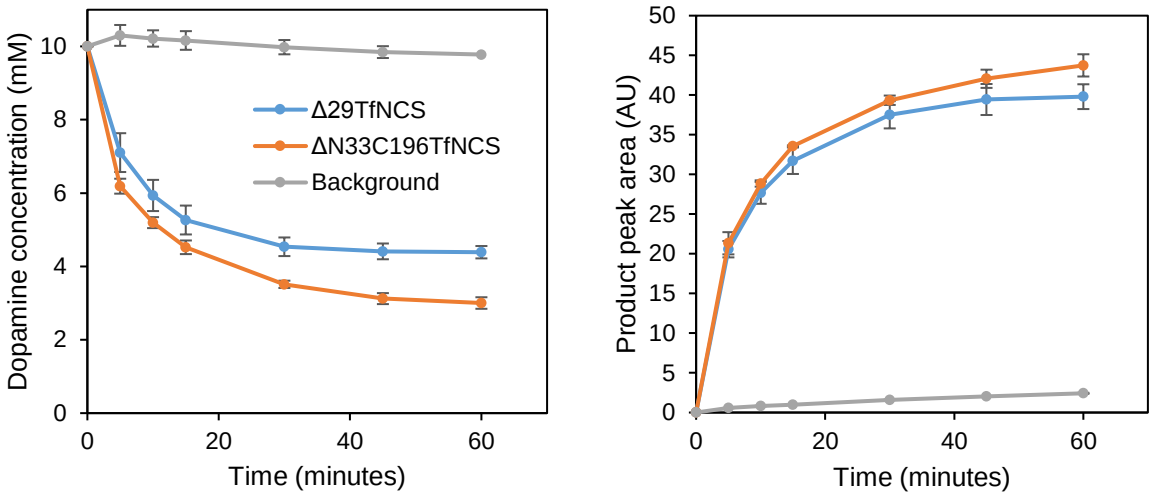


Figure S2. Time course comparison of ΔN33C196*Tf*NCS and Δ29*Tf*NCS. with dopamine (10 mM) and hexanal substrates (10 mM). Values are the mean of three separate measurements, error bars indicate standard deviations.

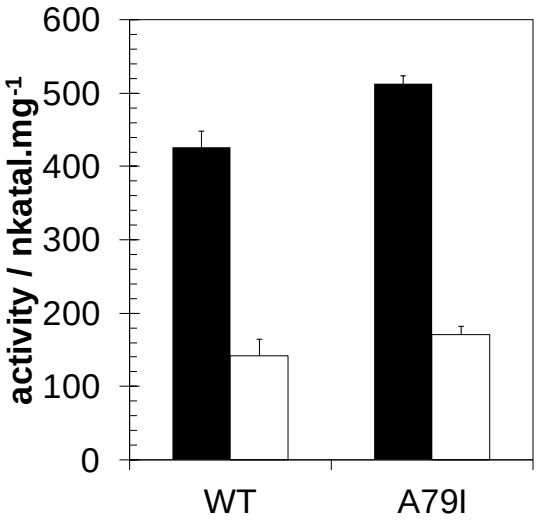


Figure S3. Enzyme activities of WT and A79I Δ29*Tf*NCS. Initial rates between dopamine (10 mM) and aldehydes (2.5 mM) catalysed by Δ29*Tf*NCS variants. 4-HPAA (black bars) and hexanal (white bars). Values are the mean of three separate measurements, error bars indicate standard deviations. Background activity has been subtracted from all measurements.

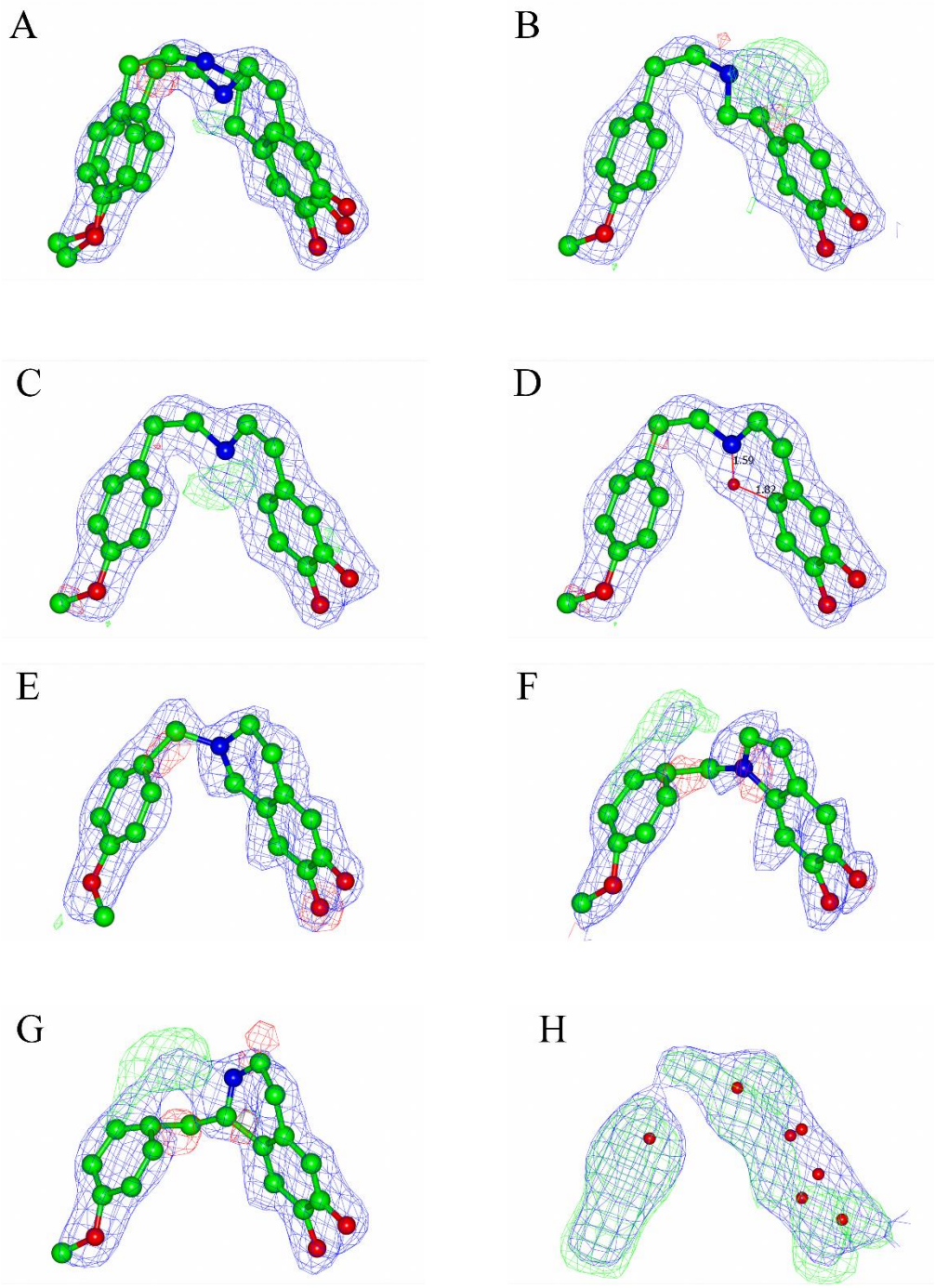


Figure S4. Different interpretations of the ligand density. A. Mixture of Productive and Unproductive mimic conformations. B. Unproductive mimic conformation. C. Productive mimic conformation. D. Productive mimic conformation plus water (distances to N1 and C10 shown). E. Tertiary amine. F. Ring closure of mimic. G. (*S*)-argemexirine **5**. H. Original density after one round of refinement of apo structure (including waters) direct with Refmac. The two data sets were isomorphous enough to obviate a molecular replacement step. 2Fo-Fc maps in blue at 1 sigma. Fo-Fc at +3 sigma (green) and -3 sigma (red). All maps clipped to the double mimic coordinates at 1.5 A (Fo-Fc) and 2 A difference maps. Drawn with CCP4mg.

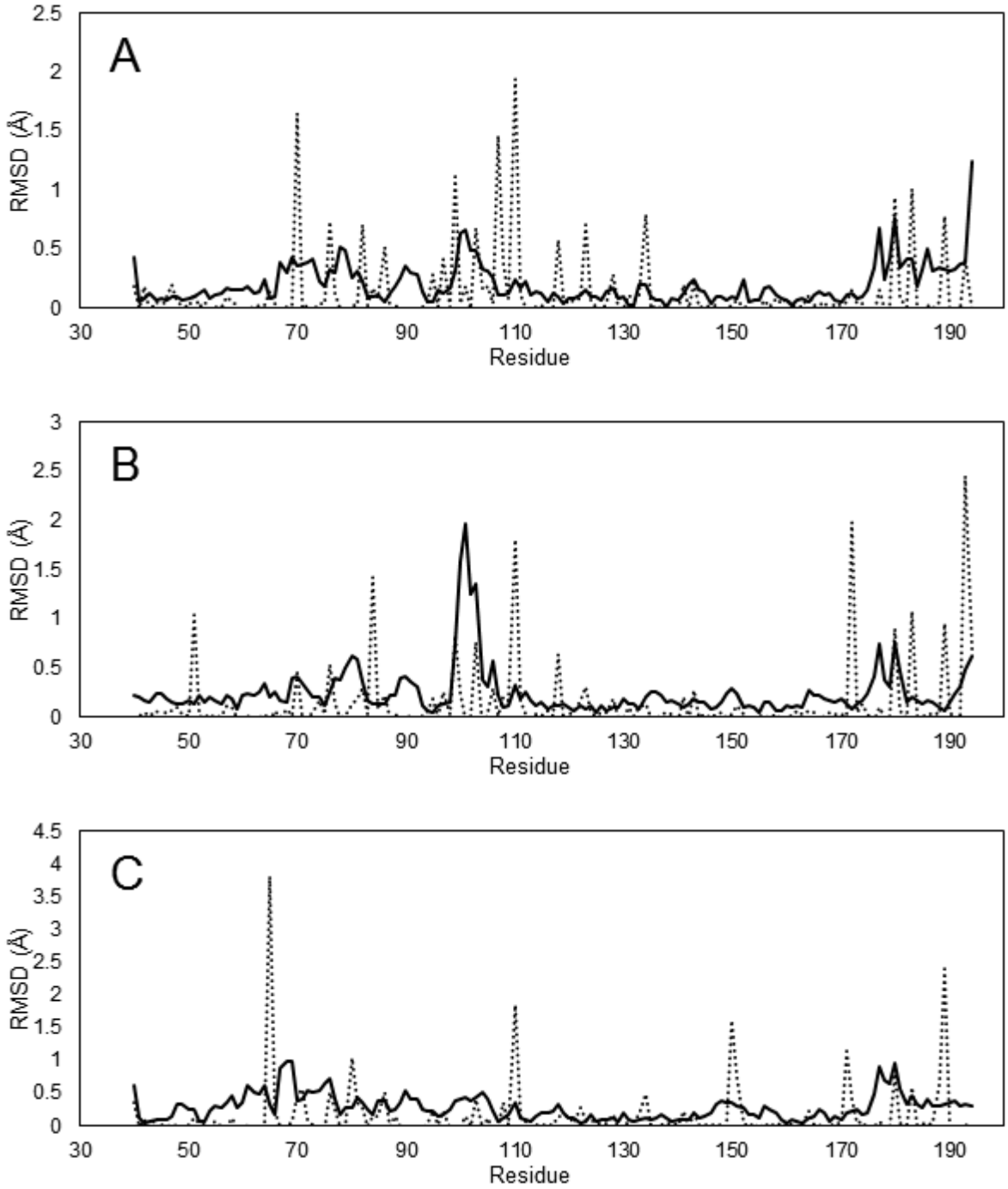


Figure S5. Changes to structure upon ligand binding. RMSD (Å) calculated with UCSFChimera for each subunit pair of 5N8Q and 5NON. Solid line is the Cα RMSD, dotted line the sidechain RMSD (calculated by subtracting the backbone RMSD from the full residue RMSD). A, B and C are the three chains of NCS in the asymmetric unit.

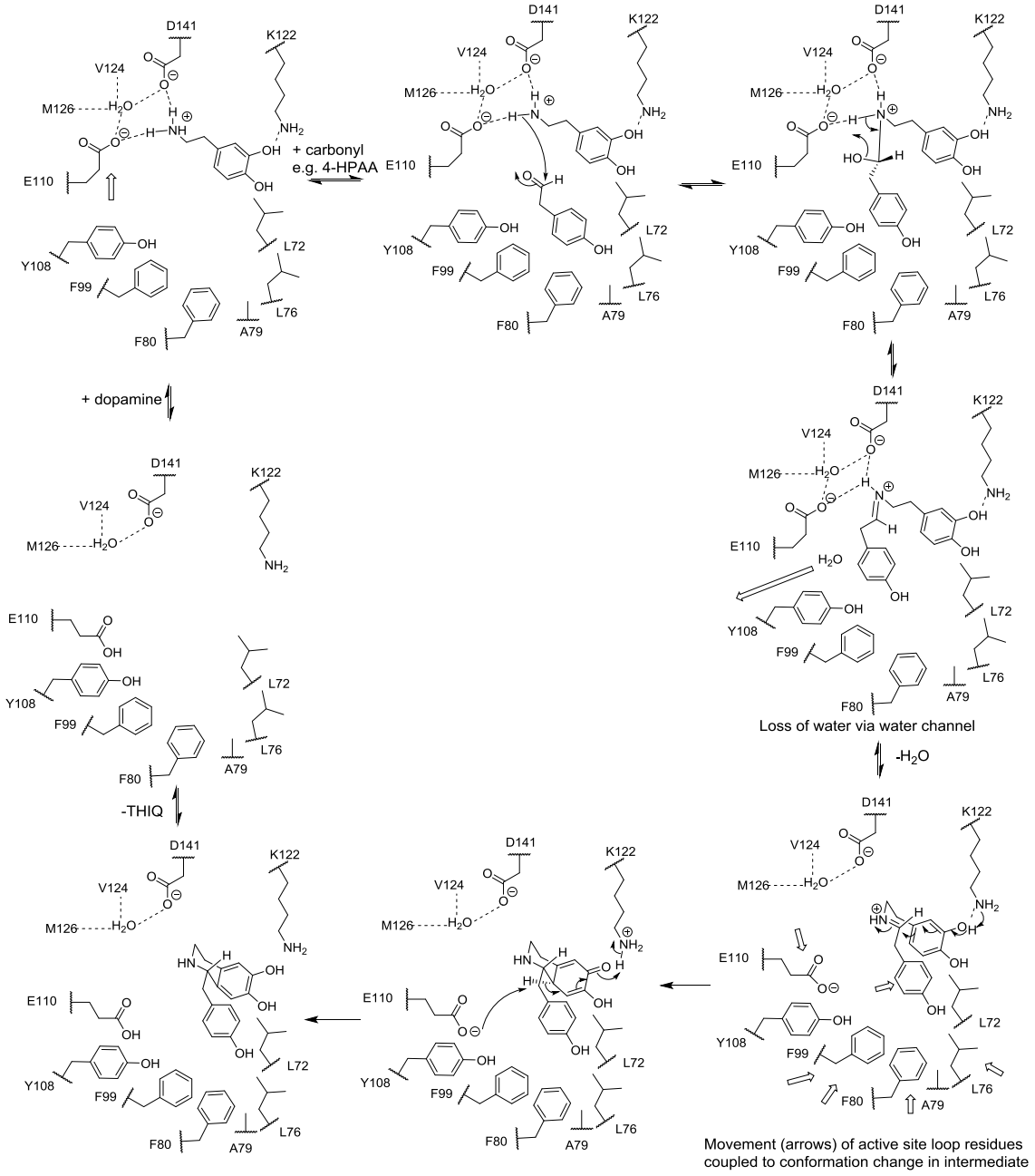
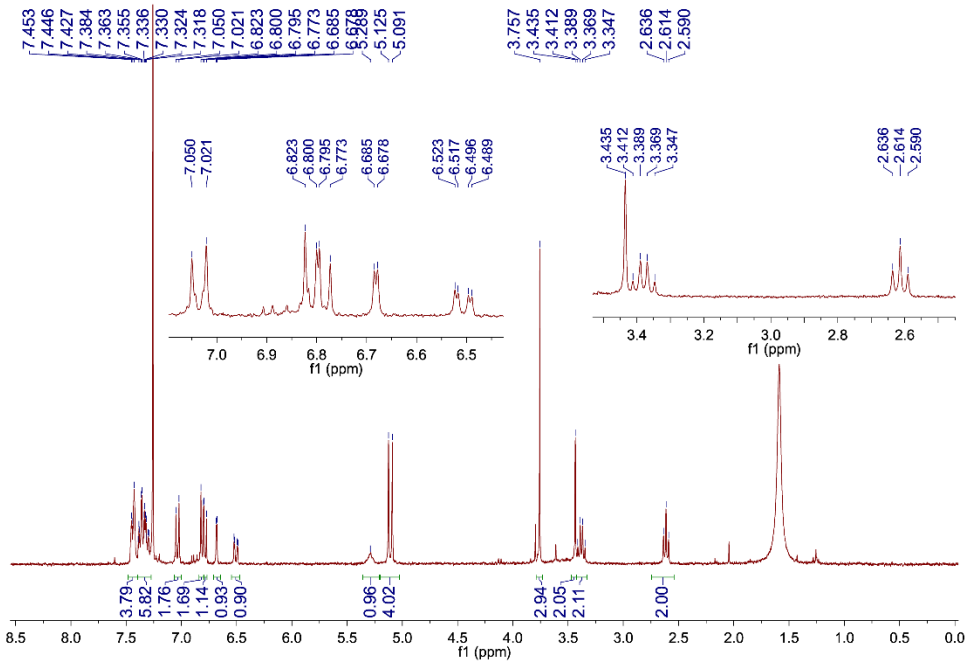


Figure S6. Full updated proposed dopamine-first mechanism.
Curly arrows represent electron movement, block arrows represent physical movement of residues/water.

^{1}H



^{13}C

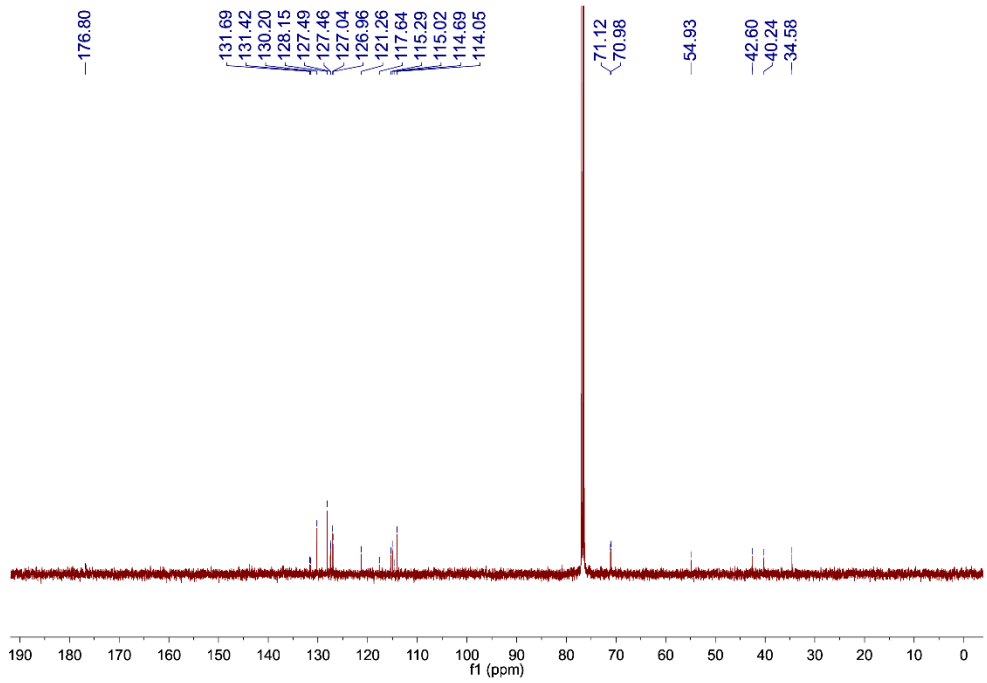
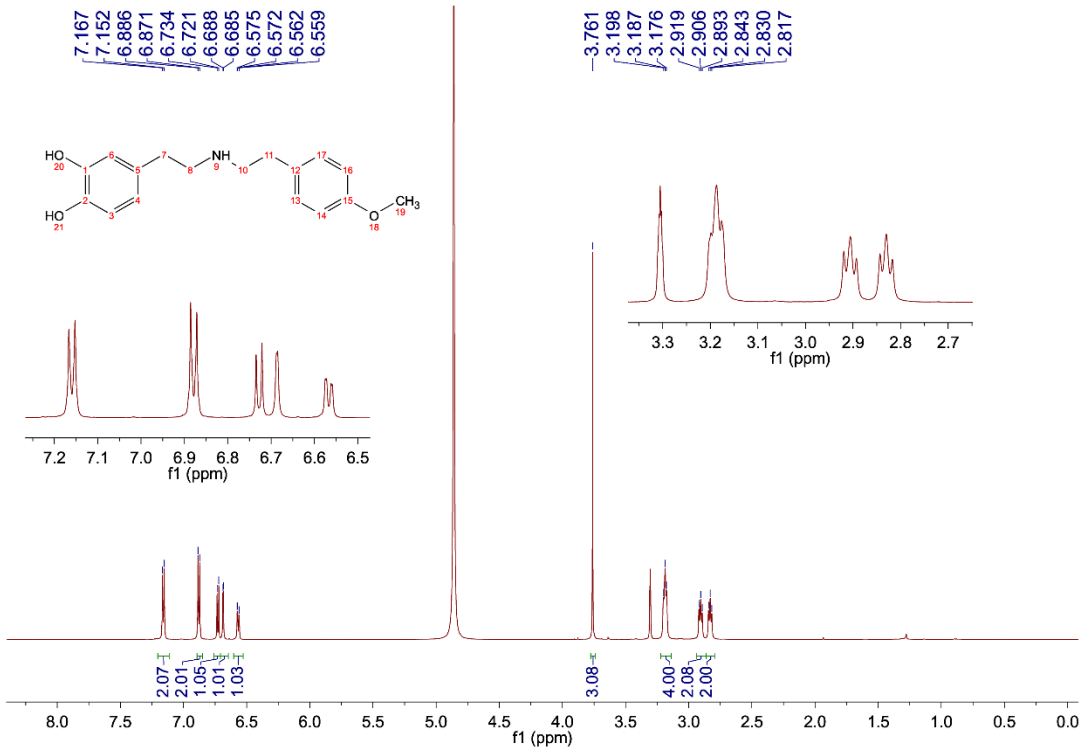


Figure S7. NMR spectra of synthetic intermediate 8.

^{1}H



^{13}C

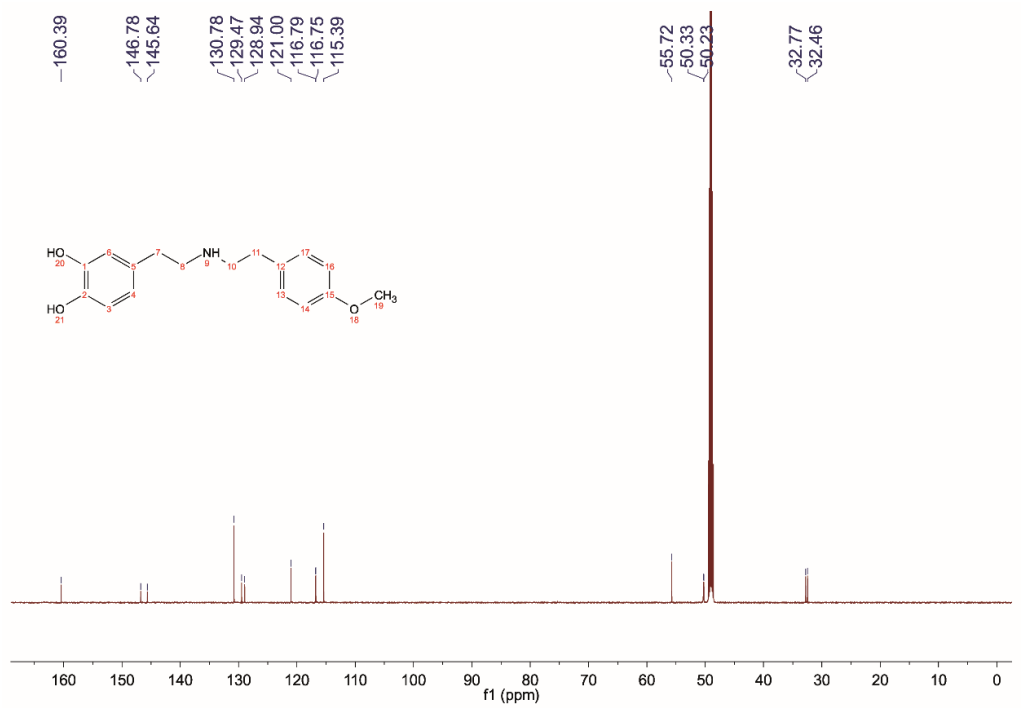


Figure S8. NMR spectra of mimic 6.

References

(1) Samanani, N., Liscombe, D. K., and Facchini, P. J. (2004). *Plant J. 40*, 302–313.

(2) Gustafsson, C., Minshull, J., Govindarajan, S., Ness, J., Villalobos, A., and Welch, M. (2012). *Protein Expr. Purif. 83*, 37–46.

(3) Kabsch, W. (2010). *Acta Crystallogr. Sect. D Biol. Crystallogr. 66*, 125–132.

(4) Evans, P. R. (2011). *Acta Crystallogr. Sect. D Biol. Crystallogr. 67*, 282–292.

(5) McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007). *J. Appl. Crystallogr. 40*, 658–674.

(6) Ilari, A., Franceschini, S., Bonamore, A., Arenghi, F., Botta, B., Macone, A., Pasquo, A., Bellucci, L., and Boffi, A. (2009). *J. Biol. Chem. 284*, 897–904.

(7) Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010). *Acta Crystallogr. Sect. D Biol. Crystallogr. 66*, 486–501.

(8) Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997). *Acta Crystallogr. Sect. D Biol. Crystallogr. 53*, 240–255.

(9) Pearce, N. M., Krojer, T., and Delft, F. Von. (2017). *Acta Crystallogr. Sect. D Biol. Crystallogr. 73*, 256–266.

(10) Trott, O., and Olson, A. J. (2010). *J. Comput. Chem. 31*, 455–461.

(11) Lichman, B. R., Gershater, M. C., Lamming, E. D., Pesnot, T., Sula, A., Keep, N. H., Hailes, H. C., and Ward, J. M. (2015). *FEBS J. 282*, 1137–1151.

(12) Li, Y., Zhou, Y., Qi, B., Gong, T., Sun, X., Fu, Y., and Zhang, Z. (2014). *Mol. Pharm. 11*, 3174–3185.

(13) Sievers, F., and Higgins, D. G. (2014). *Methods Mol. Biol. 1079*, 105–116.

(14) Robert, X., and Gouet, P. (2014). *Nucleic Acids Res. 42*, W320–W324.

(15) Minami, H., Dubouzet, E., Iwasa, K., and Sato, F. (2007). *J. Biol. Chem. 282*, 6274–6282.

(16) Lee, E.-J., and Facchini, P. (2010). *Plant Cell 22*, 3489–3503.

(17) Li, J., Lee, E.-J., Chang, L., and Facchini, P. J. (2016). *Sci. Rep. 6*, 39256.