Supplementary Information

Synthesis and Evaluation of 2,5-Dihydrochorismate Analogues as Inhibitors of the Chorismate-Utilising Enzymes

Richard J. Payne,[ab]* Esther M. M. Bulloch,[a] Miguel M. Toscano,[a] Michelle A. Jones,[a] Olivier Kerbarh[a] and Chris Abell[a]*

[a] Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, UK.. Fax: +44 1223 336362; Tel: +44 1223 336405; **ca26@cam.ac.uk**

[b]School of Chemistry, The University of Sydney, Sydney, NSW 2006, Australia. **payne@chem.usyd.edu.au**

Table of contents

GOLD dockings of ADC and isochorismate

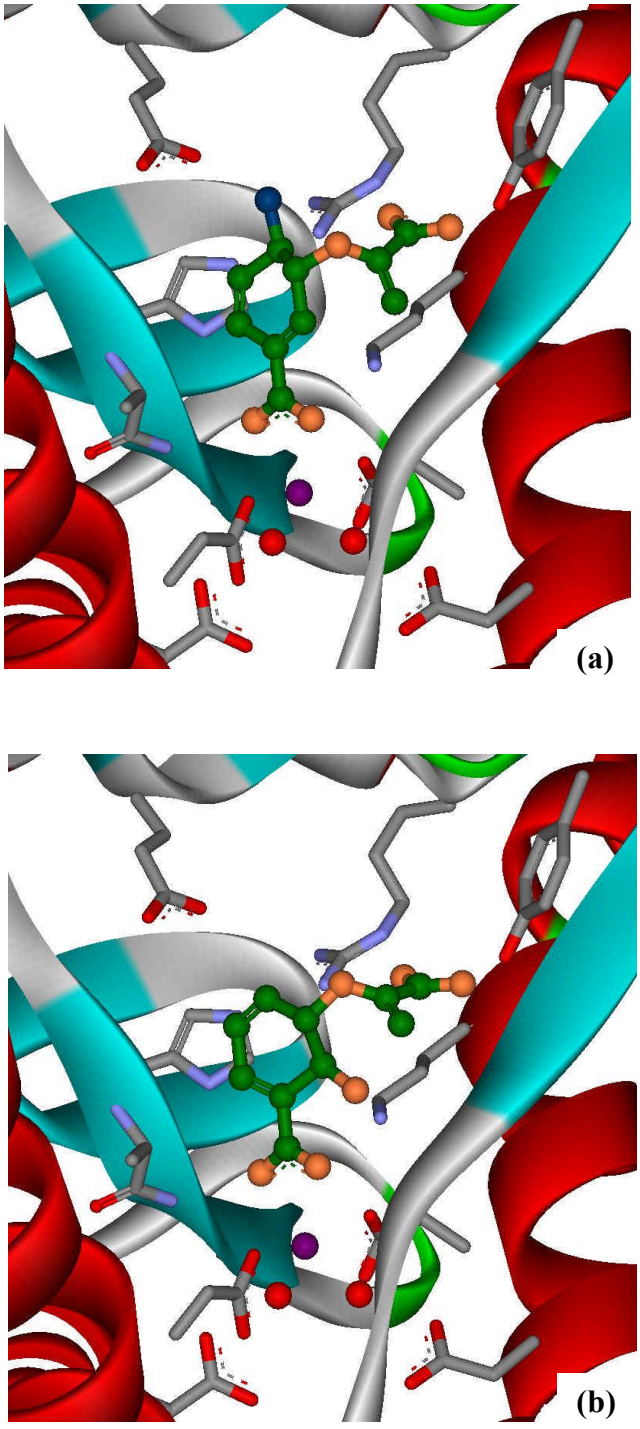


Figure S1. GOLD docking of (a) 4-deoxy-4-aminochorismate (ADC) and (b) isochorismate into the active-site of *S. marcescens* anthranilare synthase.

Sequence alignments

(a)

Sequences are labelled as follows:

`PabB`: ADC synthase (PabB) sequence from *E.coli*

`TrpE`: Anthranilate synthase (TrpE) sequence from *S. marcescens*

`Irp9`: Salicylate synthase (Irp9) sequence from *Y. enterocolitica*

`EntC`: Isochorismate synthase (EntC) sequence from *E.coli*

Sequences were aligned using the CLUSTAL W (1.82) multiple sequence alignment software

"*" indicates that the amino acid residues in that column are identical in all sequences of the alignment

":" indicates that conserved residues have been identified amongst the sequences in the alignment

"." indicates that semi-conserved substitutions have been observed amongst the sequences in the alignment

```
PabB            ---MKTLSPAVITLLWRQDAAEFYFSRLSHLPWAMLLHSGYADH-PYSRFDIVVAEPICT 56
TrpE            MNTKPQLTLLKVQASYRGDPTTLFHQLCGARPATLLLESAEIND-KQNLQSLLVIDSALP 59
Irp9            --------MKISEFLHLALPEEQWLPTISGVLRQFAEEECYVYERPPCWYLGKGCQARLH 52
EntC            --------------MDTSLAEEVQQTMATLAPNRFFFMS-----------------PYRS 29
                                     .             :   .                .

PabB            LTTFGKETVVSESEKR------------TTTTDDPLQVLQQVLDRADIRPTHNEDLPFQG 104
TrpE            ITALGHTVSVQALTANGPALLPVLDEALPPEVRNQARPNGRELTFPAIDAVQDEDARLRS 119
Irp9            INADGTQATFIDDAGEQ-----------KWAVDSIADCARRFMAHPQVKGRR-------- 93
EntC            FTTSGCFARFDEPAVNG---------------DSPDSPFQQKLAALFADAKAQ------- 67
                :.: *  . .     .                .      : :

PabB            G------------------------ALGLFGYDLGRRFESLPEIAEQDIVLPDMAVGIYD 140
TrpE            LSVFDALRTLLTLVDSPADEREAVMLGGLFAYDLVAGFENLPAVRQ-DQRCPDFCFYLAE 178
Irp9            ------------------------VYGQVGFNFAAHARGIAFNAG---EWPLLTLTVPR 125
EntC            -----------------------------GIKNPVMVGAIPFDPR-----QPSSLYIPE 92
                                          . .      :.            . :

PabB            WALIVDHQRHTVSLLS-------------HNDVNARRAWLESQQFSPQEDFTLTSDWQSN 187
TrpE            TLLVLDHQRGSARLQASVFSEQASEAQRLQHRLEQLQAELQQPPQPIPHQKLENMQLSCN 238
Irp9            EELIFEKGNVTVYADS-------------ADGCRRLCEWVKEASTTTQN---APLAVDTA 169
EntC            SWQSFSRQEKQASARR----------------------------FTRSQSLNVVERQAI 123
                    ..: .  .                                              .

PabB            MTREQYGEKFRQVQEYLHSGDCYQVNLAQRFHATYSGDEWQAFLQLNQANRAPFSAFLRL 247
TrpE            QSDEEYGAVVSELQEAIRQGEIFQVVPSRRFSLPCPAP-LGPYQTLKDNNPSPYMFFMQD 297
Irp9            LNGEAYKQQVARAVAEIRRGEYVKVIVSRAIPLPSRIDMPATLLYGRQANTPVRSFMFRQ 229
EntC            PEQTTFEQMVARAAALTATPQVDKVVLSRLIDITTDAAIDSGVLLERLIAQNPVSYNFHV 183
                     :   . .        :  :*  :: :  .            .          ::
```

```
PabB            ---EQGAILSLSPERFILCDNS--EIQTRPIKGTLPR------LPDPQEDSKQAVKLANS 296
TrpE            ---DDFTLFGASPESALKYDAGNRQIEIYPIAGTRPRGRRADGSLDLDLDSRIELEMRTD 354
Irp9            ---EGREALGFSPELVMSVTGN--KVVTEPLAGTRDR------MGNPEHNKAKEAELLHD 278
EntC            PLADGGVLLGASPELLLRKDGE--RFSSIPLAGSARR------QPDEVLDREAGNRLLAS 235
                   :    :. ***  :       ..   *: *:  *        :   :     .:  .

PabB            AKDRAENLMIVDLMRNDIGRVAVAGSVKVPELFVVEPFPAVHHLVSTITAQLPEQLHASD 356
TrpE            HKELAEHLMLVDLARNDLARICQAGSRYVADLTKVDRYSFVMHLVSRVVGTLRADLDVLH 414
Irp9            SKEVLEHILSVKEAIAELEAVCLPGSVVVEDLMSVRQRGSVQHLGSGVSGQLAENKDAWD 338
EntC            EKDRHEHELVTQAMKEVLRERSSE--LHVPSSPQLITTPTLWHLATPFEGKANSQENALT 293
                 *:  *: : ..     :   .      * .   :     : ** : . .    : ..

PabB            LLRAAFPGGSITGAPKVRAMEIIDELEPQRRNAWCGSIGYLSFCGNMDTSITIRTLTAIN 416
TrpE            AYQACMNMGTLSGAPKVRAMQLIAALRSTRRGSYGGRVGYFTAHRHLDTCIVIRSAYVED 474
Irp9            AFTVLFPSITASGIPKNAALNAIMQIEKTPRELYSGAILLLDDTR-FDAALVLRSVFQDS 397
EntC            LACLLHPTPALSGFPHQAATQVIAELEPFDRELFGGIVGWCDSEGNGEWVVTIRCAKLRE 353
                         : :* *:  * : *  :.   *  : * :         :  :.:*     .

PabB            GQIFCSAGGGIVADSQEEAEYQETFDKVNRILKQLEK-------- 453
TrpE            GHRTVQAGAGVVQDSIRRREADETRNKARAVLRAIATAHHAKEVF 519
Irp9            QRCWIQAGAGIIAQSTPERELTETREKLASIAPYLMV-------- 434
EntC            NQVRLFAGAGIVPASSPLGEWRETGVKLSTMLNVFGLH------- 391
                 :    **.*::  *    *  **  *   :   :
```

(b)

TrpE - PabB 27%

TrpE - EntC 21%

TrpE - Irp9 24%

PabB - EntC 21%

PabB - Irp9 22%

Irp9 - EntC 24%

Figure S2. (a) Clustal W sequence alignment for four chorismate-utilising enzymes PabB (ADC synthase from *E.coli*), TrpE: (Anthranilate synthase from *S. marcescens*), Irp9 (salicylate synthase from *Y. enterocolitica*) and EntC (isochorismate synthase from *E.coli* (b) pairwise sequence similarities for the four chorismate-utilising enzymes.

Structural similarities of chorismate-utilising enzyme active sites

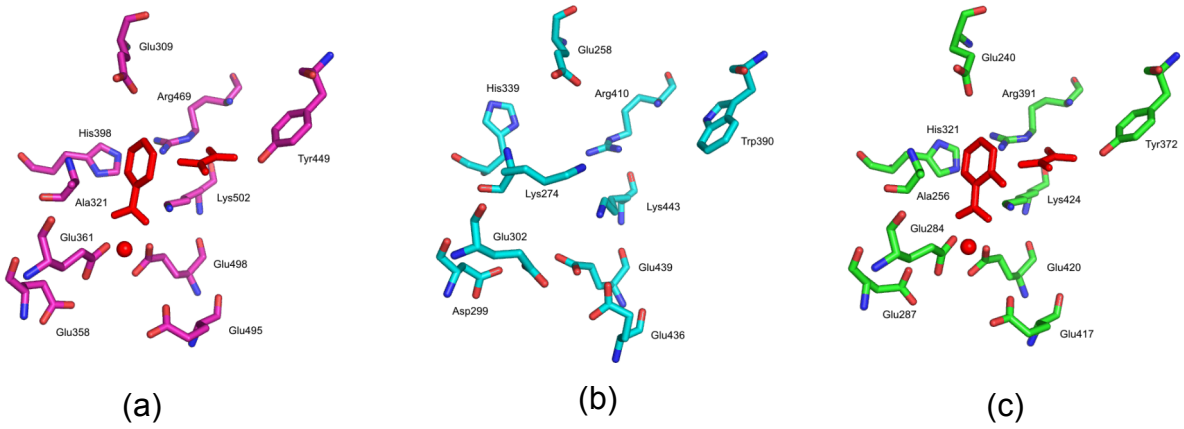
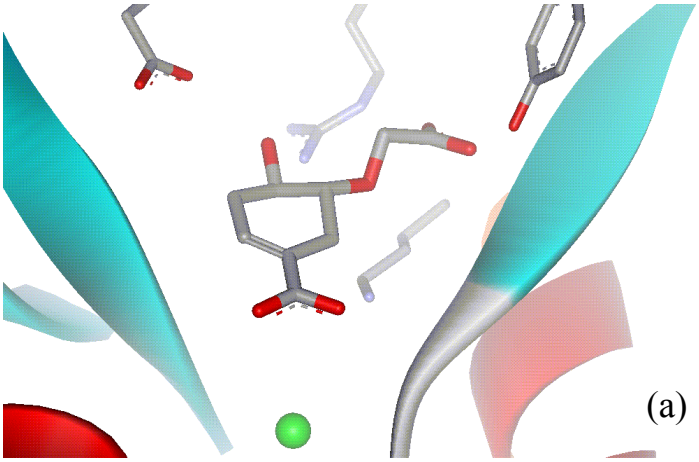
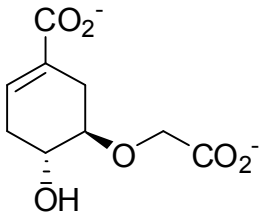


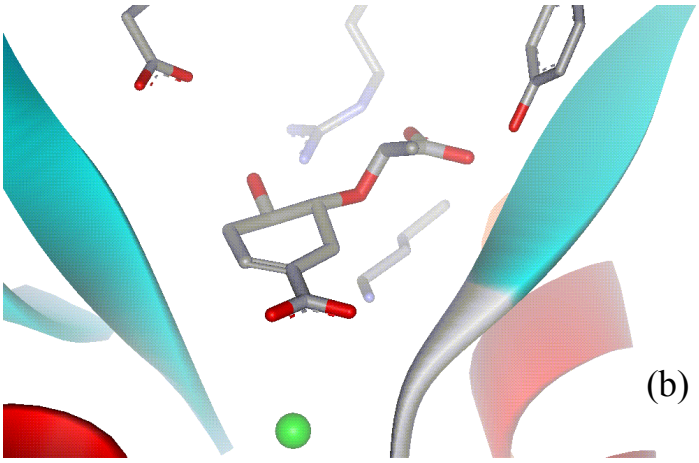
Figure S3. Active site comparisons of (a) TrpE subunit of *S. marcescens* anthranilate synthase with benzoate and pyruvate bound; (b) *E. coli* PabB (ADC synthase) and (c) *Y. enterocolitica* salicylate synthase with salicylate and pyruvate bound.

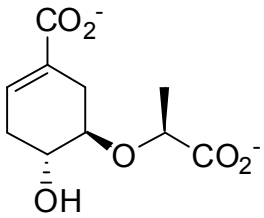
GOLD dockings of 2,5-dihydrochorismate inhibitors 11, 12 and 14



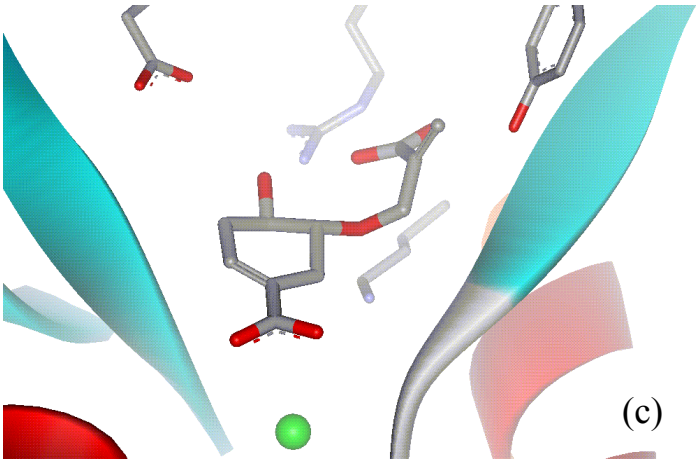


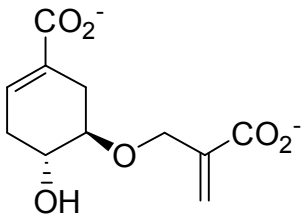
11





12





14

Figure S4. GOLD docking of 2,5-dihydrochorismate-based inhibitors (a) **11** (b) **12** and (c) **14** into the active-site of *S. marcescens* anthranilare synthase.

Experimental:

General methods: All non-aqueous reactions were carried out in pre-dried glassware under an inert atmosphere (N_2 or Ar). Organic solvents were freshly distilled prior to use and milli-Q deionised water was used for all biochemical work. Analytical thin layer chromatography was carried out on commercial silica gel 60 0.25 mm plates using either UV absorption or potassium permanganate stain (3 g potassium permanganate, 20 g potassium carbonate, 5 ml of 5% sodium hydroxide, 300 ml water) for visualisation. R_F values are quoted with respect to the solvent system used to develop the plate. Column chromatography was carried out using 230-400 mesh silica gel 60. Unless otherwise stated petroleum ether refers to the fraction collected between 40 – 60 ^oC. ^1H NMR spectra were recorded on a Bruker AM-400 spectrometer or a Bruker Avance 500 spectrometer in deuterated solvents, as indicated. ^{13}C NMR spectra were recorded on a Bruker AM-400 spectrometer or a Bruker Avance 500 spectrometer linked to a Bruker 5mm dual Cryoprobe (operating at 100 MHz and 125 MHz, respectively). All chemical shifts are quoted in parts per million (ppm) δ. Coupling constants for ^1H NMR spectroscopy are assigned where possible and are given in Hz. Infrared spectra were recorded on a Perkin Elmer Spectrum One FTIR spectrometer using attenuated transmittance reflectance (ATR). Liquid-chromatography mass-spectrometry (LC-MS) was carried out using an Alliance HT Waters 2795 Separations Module coupled to a Waters Micromass ZQ Quadrapole Mass Analyzer. Samples were detected using a photomultiplier detection system. Samples were run on a gradient from 10 mM ammonium acetate containing 0.1% formic acid to 95% acetonitrile over a period of 8 min.

Docking Studies. *Protein preparation:* The crystal structure of *S. marcescens* anthranilate synthase was downloaded from the Brookhaven Protein Databank (PDB code: 1K0E).[1] Hydrogens were added to the protein using SYBYL6.5 and the products, benzoate and pyruvate, were abstracted from the active site of the complexed protein pdb file. Crystallographic waters were removed from the structure, except for H_2O-518 and H_2O-519 as these mediated protein interactions to the magnesium ion in the active site. The hydrogens of these two water molecules were manually adjusted to optimise hydrogen-bonding to the protein. *Inhibitor preparation:* Inhibitors were built in SYBYL6.5 and used as MOL2 files.[2] Each molecule had hydrogens added. Structures were minimised to relax bond lengths and fix angles using a Tripos force field and Gasteiger-Hückel charges calculated.[3] *Docking:* Ligands were docked into the active sites of the

enzymes prepared above using GOLD2.1.[4] For each independent genetic algorithm (GA) run, a maximum of 100 000 operations were performed on a population of 5 islands, each of 100 individuals. Operator weights for crossover, mutation and migration were set to 95, 95 and 10 respectively. To allow for development of close contacts and poor hydrogen bonds occurring at the beginning of each GA run, the initial external Van der Waals (vdw) energies were cut off at $2.5k_{ij}$, where k_{ij} is the depth of the vdw well between atoms i and j, and the maximum distance between a donor atom and a fitting point was set to 4 Å. Ring flipping and internal hydrogen bonds were allowed. The dockings were terminated after 25 runs for each of the inhibitors.

Isolation of shikimic acid 15.[5] Star anise seeds and carpels (900 g) were suspended in ethanol (1 l) and ground to a fine consistency using a blender. The resulting suspension was subjected to a Soxhlet extraction with 95% ethanol (4 l) for 24 h. The dark brown extract was evaporated to dryness *in vacuo* to give a thick green oil. This oil was dissolved in water (5 l) and warmed to 80 ^{o}C, whereupon a dark green oil formed on the surface which was removed by pipette and discarded. Aqueous formaldehyde (5 ml of a 37-40% solution) was added to the hot solution, which was subsequently boiled for 5 min and then allowed to cool. The precipitate that formed was removed by filtration to leave a clear orange solution, which was subsequently passed down a column of Amberlite IRA-400 (Cl) anion exchange resin (standard grade, 500 g, as the acetate). After washing with water (3 l), the product was eluted with aqueous acetic acid (25% v/v, 4 l). The resulting yellow solution was evaporated under reduced pressure to yield the crude product as a pale orange solid that was dissolved in a minimum volume of water and was applied to a 15 cm column of Celite® in water. Elution with water afforded a pale yellow solution that was evaporated under reduced pressure to afford the crude product as an orange solid. The solid was dissolved in hot methanol and the coloured impurities removed by the addition of activated charcoal powder (5 g). The off white solid that resulted was recrystallised from toluene and methanol to give shikimic acid **15** as a white solid (45 g, 5% yield from Star anise). mp. 184 – 186 ^{o}C; $[\alpha]_D^{25}$ –166^{o} (c 4.0 H_2O); [Lit mp. 184 ^{o}C; $[\alpha]_D^{25}$ –176^{o} (*c* 2.0, EtOH)][5]; ^{1}H NMR (500 MHz, D_2O) 2.12 (1H, ddt, *J* 18.0, 6.5, 1.5 Hz, H-6β), 2.63 (1H, ddt, *J* 18.0, 5.0, 1.5 Hz, H-6α), 3.66 (1H, dd, *J* 8.5, 4.0 Hz, H-4), 3.94 (1H, ddd, *J* 8.5, 6.5, 5.0 Hz, H-5), 4.34 (1H, tt, *J* 1.5, 4.0 Hz, H-3), 6.72 (1H, dt, *J* 1.5, 4.0 Hz, H-2); ^{13}C NMR (125 MHz, D_2O) 36.3, 71.7, 72.5, 77.0, 135.6, 143.3, 176.1; HRMS calcd for $C_7H_{10}O_5Na$: *M*Na^+, 197.0427. Found: MNa^+, 197.0426. These data are in agreement with those previously reported by Adams *et al*.[5]

Enzyme overexpression and purification protocols:

***Serratia marcescens* anthranilate synthase.**

S. marcescens anthranilate synthase was expressed and purified as described previously.[6]

***Escherichia coli* isochorismate synthase (EntC), *Pseudomonas aeruginosa* isochorismate pyruvate lyase (PchB) and *Yersinia enterocolitica* salicylate synthase.**

The *pchB* gene from *P. aeruginosa* was amplified by PCR from the vector pME3324 which was kindly supplied by Professor Dieter Haas, Laboratoire de Microbiologie Microbienne, Université de Lausanne, Switzerland. The *entC* gene was amplified from *E. coli* genomic DNA. Genes were cloned into the vector pET-28a (Novagen). Enzymes were over-expressed with an *N*-terminal hexahistidine affinity-tag and purified using a Ni-NTA column (Qiagen). Concentrated stock solutions of *E. coli* isochorismate synthase (EntC, 15.9 mg/ml, 352 μM), *P. aeruginosa* isochorismate pyruvate lyase (PchB, 15.6 mg/ml, 573 μM) were stored as aliquots in 100 mM potassium phosphate (pH 7.0) at -80 ^{o}C.

***Yersinia enterocolitica* salicylate synthase (Irp9)**

The *irp9* gene was amplified by PCR from the *Y. enterocolitica* (strain 8081) genomic DNA, which was kindly supplied by Professor Brendan Wren, Department of Infectious & Tropical Diseases, London School of Hygiene & Tropical Medicine, UK, and was cloned into the expression vector pET-28a (Novagen). Irp9 was over-expressed with an *N*-terminal hexahistidine affinity-tag and purified using a Ni-NTA column (Qiagen). Concentrated stock solutions of *Y. enterocolitica* salicylate synthase (Irp9) were stored as aliquots in 100 mM potassium phosphate (pH 7.0) at –80 ^{o}C.

***Escherichia coli* ADC synthase (PabA, PabB) and ADC lyase (PabC)**

E. coli pabA and *pabB* genes coding for ADC synthase were amplified from the vectors pNPA and pNPB, which were kindly supplied by Prof. C. T. Walsh, Harvard Medical School, Massachusetts, USA. *E. coli pabC* was amplified from *E. coli* genomic DNA. Genes were cloned into the vector

mini-pRSETA (Invitrogen). Enzymes were over-expressed with an *N*-terminal hexahistidine tag and were purified using a Ni-NTA column (Qiagen). Concentrated solutions of each of these enzymes were stored as aliquots in 50 mM Tris-HCl (pH 7.8) at –80 ^{o}C.

Enzyme Assays:

Anthranilate synthase assay: Kinetic parameters for the glutamine-dependent activity of *S. marcescens* anthranilate synthase were determined using the kinetic fluorescence assay developed by Bauerle *et al*.[6,7] Fluorescence was detected at 390 nm after initial excitation at 313 nm at 25 ^{o}C. Assays were performed in duplicate in Nunc 96-well fluorescence plates with a total volume of 200 μl in each well. Reactions were initiated by the addition of anthranilate synthase (20 μl of a 0.01 mg/ml solution, final concentration = 16.9 nM) to the following assay mixture that had been pre-incubated at 25 ^{o}C for 5 min: 100 mM potassium phosphate buffer (pH = 7.0), 10 mM magnesium chloride, 20 mM glutamine, 1-50 μM chorismate. Initial rates were determined by measuring the increase in fluorescence over the first min and the data fitted to a Michaelis-Menten kinetic model using the software GraFit.[8] For the inhibitor studies, various concentrations of inhibitor (1 μM – 1 mM) were added to the assay mixture. Inhibition contants (K_I's) were calculated using the software GraFit and an F-test was used to confirm that the data satisfied a competitive inhibition model.[8]

Isochorismate synthase (EntC) assay: Kinetic parameters for the activity of *E. coli* isochorismate synthase (EntC) were determined using a coupled assay. *P. aeruginosa* isochorismate pyruvate lyase (PchB) was used to convert isochorismate to salicylate which was detected by fluorescence at 440 nm after initial excitation at 305 nm at 25 ^{o}C.[9,10] Assays were performed in duplicate in Nunc 96-well fluorescence plates with a total volume of 200 μl in each well. Reactions were initiated by the addition of isochorismate synthase (EntC) (10 μl of a 2.07 μM solution, final concentration 104 nM) to the following assay mixture that had been pre-incubated at 25 ^{o}C for 5 min: 100 mM potassium phosphate, pH 7.0, 10 mM magnesium chloride**,** 18-90 μM chorismate**,** 10 μl of a 19.1 μM solution of PchB, (final concentration 478 nM). Initial rates were determined by measuring the increase in fluorescence over the first min and the data fitted to a Michaelis-Menten kinetic model using the software GraFit.[8] For the inhibitor studies, various concentrations of inhibitor (10 μM – 1 mM) were added to the assay mixture. Inhibition contants (K_I's) were

calculated using the software GraFit and an F-test was used to confirm that the data satisfied a competitive inhibition model.[8]

Salicylate synthase (Irp9) activity assay: Kinetic parameters for the activity of *Y. enterocolitica* salicylate synthase were determined using a kinetic fluorescence assay where fluorescence was detected at 440 nm after initial excitation at 305 nm at 25 ^{o}C.[11] Assays were performed in duplicate in Nunc 96-well fluorescence plates with a total volume of 200 μl in each well. Reactions were initiated by the addition of salicylate synthase (20 μl of a 0.5 mg/ml solution, final concentration = 1 μM) to the following assay mixture that had been pre-incubated at 25 ^{o}C for 5 min: 100 mM potassium phosphate buffer, pH 7.0, 10 mM magnesium chloride**,** 1.8 – 22.5 μM chorismate**.** Initial rates were determined by measuring the increase in fluorescence over the first min and the data fitted to a Michaelis-Menten kinetic model using the software GraFit.[8] For the inhibitor studies, various concentrations of inhibitor (10 μM – 1 mM) were added to the assay mixture. Inhibition contants (K_I's) were calculated using the software GraFit and an F-test was used to confirm that the data satisfied a competitive inhibition model.[8]

ADC synthase activity assay: Glutamine-dependent ADC synthase activity was measured by a coupled assay with ADC lyase and lactate dehydrogenase.[12] The initial rates were measured by the consumption of NADH which was monitored at 340 nm (ε = 6220 $Lmol^{-1}cm^{-1}$) at 37 ^{o}C. Assays were carried out in 96 well plates with a total volume of 200 μl in each well and were performed in duplicate. Assays were initiated by the addition of His_6-tagged PabA (10 μl of a 0.55 mg/ml solution, final concentration 1.21 μM) to to the following assay mixture that had been pre-incubated at 37 ^{o}C for 5 min: 50 mM Tris.HCl buffer, pH 7.8, 5 mM magnesium chloride**,** 50 mM glutamine**,** 2.7 – 22.5 μM chorismate**,** 352 μM NADH**,** 357 nM (1.4 units) lactate dehydrogenase (rabbit muscle type XI, purchased from Sigma)**,** 190 nM His_6-tagged PabB**,** ~1.75 μM His_6-tagged PabC (ADC lyase). Initial rates were determined by measuring the decrease in absorption over the first min and the data fitted to a Michaelis-Menten kinetic model using the software GraFit.[8] For the measurement of inhibition constants of the ADC synthase system, various concentrations of inhibitor (50 μM – 2 mM) were added to the assay mixture. Inhibition contants (K_I values) were calculated using the software GraFit and an F-test was used to confirm that the data satisfied a competitive inhibition model.[8]

References

1. H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weisseg, N. Shindyalov, P. E. Bourne, *Nucleic Acids Res.* **2000**, *28*, 235-242.
2. *Tripos SYBYL version 6.5*; Tripos, Inc., 1699, South Hanley Road, St Louis, Missouri, 63144, USA.
3. M. Clark, R. D. I. Cramer, N. van Opdenbosch, *J. Comp. Chem.* **1989**, *10*, 982-1012.
4. G. Jones, P. Willett, R. C. Glen, A. R. Leach and R. Taylor, *J. Mol. Biol*., 1997, **267**, 727-748.
5. H. Adams, N. A. Bailey, R. Brettle, R. Cross, M. Frederickson, E. Haslam, F. S. MacBeath and G. M. Davies, *Tetrahedron*, 1996, **52**, 8565-8580.
6. R. J. Payne, M. D. Toscano, E. M. M. Bulloch, A. D. Abell and C. Abell, *Org. Biomol. Chem*., 2005, **3**, 2271-2281
7. R. Bauerle, J. Hess and S. French, *Methods Enzymol*. 1987, **142**, 366-386.
8. *GraFit (version 5.0.6)*, 1989-2003, Erithacus Software Limited.
9. C. Gaille, C. Reimmann and D. Haas, *J. Biol. Chem*., 2003, **278**, 16893-16898.
10. C. Gaille, P. Kast, D. Haas, *J. Biol. Chem*. 2002, **277**, 21768-21775.
11. R. J. Payne, O. Kerbarh, R. N. Miguel, A. D. Abell and C. Abell, *Org. Biomol. Chem*., 2005, **3**, 1825-1827.
12. Q. Z. Ye, J. Liu, C. T. Walsh, *Proc. Natl. Acad. Sci. USA* 1990, **87**, 9391-9395.