Supporting Information for

Tracing the Evolution of Angucyclinone Monooxygenases: Structural Determinants for C-12b Hydroxylation and Substrate Inhibition in PgaE

Pauli Kallio,¹ Pekka Patrikainen,¹ Georgiy A. Belogurov,¹ Pekka Mäntsälä,¹ Keqian Yang,² Jarmo Niemi¹ and Mikko Metsä-Ketelä^{1,*}

¹ Department of Biochemistry and Food Chemistry, University of Turku, FIN-20014 Turku, Finland

² State Key Laboratory of Microbial Resources, Institute of Microbiology, Chinese Academy of Sciences, Beijing 100101, PR China.

Supplementary Text: Analysis of prejadomycin conversion kinetics.

The concentration dependences of initial rate (~ 1 min reaction time) of prejadomycin conversion by wild-type PgaE revealed the apparent inhibition of catalysis by excess of substrate. To circumvent the problem of uncertainty in the initial rate determination a longer (10-15 min, Figure S4), notably nonlinear parts of prejadomycin conversion curves were fit to various rate models using numerical integration capabilities of Scientist 2.01 software (Micromath, St Louis, MO, USA). Kinetic models incorporating minimum of two prejadomycin binding sites (Figure S5) were consistent with the data but allowed for considerable flexibility in parameter values largely because the sensitivity of the assay was insufficient to accurately determine reaction rate below $10 \mu\text{M}$ substrate, thereby precluding accurate determination of the dissociation constant of the mono-substrate complex (K_{ES}). Given that time courses of prejadomycin conversion failed to provide sufficient information to resolve K_{ES} , the dissociation constant of the bi-substrate enzyme complex (K_{ES2}), and the activities of mono-substrate (V_{ES}) and bi-substrate (V_{ES2}) enzyme complexes, we used structural considerations to constrain to the fit parameters. Thus, PgaE is a relatively small homodimeric protein unlikely capable of accommodating the second prejadomycin molecule outside the active site. Moreover, we identified mutations affecting the pattern of substrate inhibition in the active site and at the dimerization interface (see below). The above considerations suggest that the inhibition by excess of prejadomycin is due to allosteric interactions of active sites within the dimer. The simplest scenario capable of describing the inhibition assumes the independent binding of prejadomycin to the active sites of the monomers within the dimer and V_{ES2} being lower than V_{ES} . In quantitative term such scenario implies at least a 4-fold difference between the first and the second substrate binding constants in the Bjerrum equation ($4 \times K_{\text{ES}} = K_{\text{ES2}}$; Figure S5, Equation 1a). In other words, the substrate binding is described by a single parameter, K_{m} that links $K_{\text{ES}} = K_{\text{M}}/2$ and $K_{\text{ES2}} = 2 \times K_{\text{M}}$. Analysis of the wild-type data suggested 7-fold inhibition of prejadomycin

conversion rate upon binding of the second prejadomycin molecule ($V_{ES2}/V_{ES}=0.15$) and K_M of 3-4 μM .

P78Q, $\beta 4$, P78Q/I79F variant PgaE proteins catalyzed prejadomycin conversion qualitatively similar kinetic behavior to that of wild-type enzyme whereas I79F, $\beta 5$ and $\beta 4$ - $\beta 5$ PgaE proteins displayed broadened activity maximum with reduced amplitude. To evaluate whether changes in K_M could be responsible for the decrease in apparent substrate inhibition in variant PgaE proteins we fit the variant data to equation 1a allowing K_M to vary while keeping V_{ES2}/V_{ES} fixed at 0.15. The analysis demonstrated that K_M of P78Q/I79F PgaE variant was indistinguishable from the wild-type enzyme within the margin of error, ~ 1.5 -fold elevated K_M was required to fit P78Q and $\beta 4$ chimera datasets, whereas I79F, $\beta 5$ and $\beta 4$ - $\beta 5$ chimeras displayed ~ 3 -fold elevated K_M (Figure S5A). The above analysis suggested that the apparent decrease in substrate inhibition in variant PgaE proteins may be attributed solely to the increase in K_M for prejadomycin.

We then investigated whether changes in V_{ES2}/V_{ES} could be responsible for the decrease in apparent substrate inhibition in variant PgaE proteins. We fit the data to equation 1a, this time allowing V_{ES2}/V_{ES} to vary while keeping K_M fixed at 4 μM for wild-type and P78Q/I79F chimera and at 7 μM for all the other variant enzymes (P78Q, I79F and $\beta 4$, $\beta 5$ and $\beta 4$ - $\beta 5$ chimeras data necessitated the use of slightly elevated K_M to obtain the adequate fit). The analysis demonstrated that approximately 2-fold elevated V_{ES2}/V_{ES} was required to describe I79F, $\beta 5$ and $\beta 4$ - $\beta 5$ chimeras kinetics under such conditions ($K_M=7 \mu M$; Figure S5B). In contrast, the fit did not suggest changes in V_{ES2}/V_{ES} in P78Q, $\beta 4$ and P78Q/I79F enzymes.

Finally, we investigated whether the increase in K_{ES2} could be responsible for the decrease in apparent substrate inhibition in variant PgaE proteins. We fit the data to equation 1b which is a

more general form of equation 1a that does not assume 4-fold difference between K_{ES} and K_{ES2} (Figure S5C). In other words, in equation 1b the binding of prejadomycin to two sites on the enzyme is not required to be independent and a more general successive binding model is used (Bjerrum equation). V_{ES2}/V_{ES} was fixed at 0.15 for all enzymes, whereas K_{ES} was fixed at 2 μ M for wild-type and P78Q/I79F chimera and at 3.5 μ M for all the other variant enzymes (K_{ES} value was constrained at $K_M/2$ from fits in Figures S5A and S5B, whereas K_{ES2} was set to vary freely; P78Q, I79F, β 4, β 5 and β 4- β 5 variant data necessitated the use of slightly elevated K_{ES} to obtain an adequate fit). The analysis demonstrated that approximately 2-3-fold higher K_{ES2} than that of the wild-type PgaE was required to describe I79F, β 5 and β 4- β 5 variants kinetics under such conditions (V_{ES2}/V_{ES} = 0.15, K_{ES} = 2 or 3.5 μ M, see above). In contrast, the fit did not suggest changes in K_{ES2} in P78Q, β 4 and P78Q/I79F enzymes.

To summarize the result above, analysis of the prejadomycin conversion time courses led to the following conclusions. The P78Q/I79F variant demonstrated kinetic behavior indistinguishable from the wild-type enzyme. P78Q, β 4 variants were characterized by $\sim$ 1.5-fold elevated K_M (or K_{ES}). The apparent decrease in inhibition by excess prejadomycin in I79F, β 5 and β 4- β 5 variants can originate from (i) $\sim$ 1.5-fold elevated K_M (or K_{ES}) and 2-fold elevated V_{ES2}/V_{ES} ; (ii) $\sim$ 1.5-fold elevated K_{ES} and 2-3-fold elevated K_{ES2}/K_{ES} ; (iii) 3-fold elevated K_M . The data did not allow discrimination between the three scenario above because the qualities of the fits (judged by the sum of squared deviations) differed by less than 5 %.

Finally, the H73A PgaE variant was distinct from other PgaE proteins examined in this study in displaying simple kinetic behavior, allowing for unambiguous quantitative interpretation. In the case of H73A, inhibition by excess substrate did not occur and the data fit a simple model assuming

competitive binding of prejadomycin and the C-12 hydroxylated reaction product to a single site on the enzyme with approximately equal affinities (Figure S5D, Equation 2).

Figure S1. SDS-PAGE showing the purity of the PgaE enzyme variants utilized for kinetic measurements.

Figure S2. Initial rates of the 12-hydroxylated prejadomycin conversion by wild-type and variant PgaE enzymes versus the 12-hydroxylated prejadomycin concentrations. The kinetic parameters for A) wild-type PgaE, B) β 4- β 5 chimera, C) β 4 chimera, D) β 5 chimera, E) P78Q/I79F double mutant, F) P78Q mutant, G) I79F mutant were obtained by fitting the data with the equation H.

Diagram showing the first domain of the protein structure, including helices $\beta 1$, $\alpha 1$, $\beta 2$, $\beta 3$, and $\alpha 2$. The sequence alignment for five proteins (PgaE, LanE, CabE, UrdE, AlpG, and GilOI) is shown below the diagram. The residues are color-coded: red for conserved, blue for variable, and green for specific mutations. The alignment is as follows:

Protein	1	10	20	30	40	50	60
PgaE	MDAA	VIVVGAGP	AGMMLAGELRL	AGVEV	VVLERL	VER	TGESRGLGFTARTMEVFDQRGIL
LanE	MDAA	VIIAGAGP	AGLMLAGELRL	AGVDV	IVLERL	LAERT	TGESRGLGFTARTMEVFDQRGIL
CabE	MDAS	VIVVGAGP	AGLMLAGELRL	AGVDV	IVLERL	LAERT	TGESRGLGFTARTMEVFDQRGIL
UrdE	MDAS	VIVVGAGP	AGLMLAGELRL	AGVDV	IVLERL	LAERT	TGESRGLGFTARTMEVFDQRGIL
JadH	MTTTP	EAAGFD	ADVIVVGAGP	TGLMLAGELRL	LGARV	LVLAERL	ERTGESRGLGFTARTMEVFDQRGIL
AlpG	MEGTT	ADTD	VIVVGAGP	TGLMLAGELRL	LGARV	VVIEKL	AAPTGESRGLGFTARTMEVFDQRGIL
GilOI	MTLHAA	EAI	PSHVP	VLVVGAGP	TGLMLAGELRL	HGSRP	LVLDALPSPTGESRGLGFTARTMEVFDQRGIL

Diagram showing the second domain of the protein structure, including helices $\beta 4$, $\beta 5$, $\eta 2$, $\eta 3$, $\beta 6$, $\alpha 3$, $\beta 7$, and $\beta 8$. The sequence alignment for five proteins (PgaE, LanE, CabE, UrdE, AlpG, and GilOI) is shown below the diagram. The residues are color-coded: red for conserved, blue for variable, and green for specific mutations. The alignment is as follows:

Protein	70	80	90	100	110	120	
PgaE	PRFGE	VETSTQGHFGGLP	IDFGVL	LEGA	WQA	AKTV	PQSV
LanE	ARFGD	VETSTQGHFGGLP	IDFGVL	LEGA	WQA	AKTV	PQSV
CabE	PRFGP	VETSTQGHFGGLP	IDFGVL	LEGA	WQA	AKTV	PQSV
UrdE	HRLGD	MGTSNAGHFGGLP	VDVFGVL	LEGA	WQA	AKTV	PQSV
JadH	PRFGQ	GETLET	SPMGHFGGLP	VDVFGVL	LEGA	WQA	AKTV
AlpG	PRFGQ	QDGLAT	SPMGHFGGLP	VDVFGVL	LEGA	WQA	AKTV
GilOI	GRFQG	...LAP	VPGVH	EAGLS	IKGDH	LS	SSMRP

Diagram showing the third domain of the protein structure, including helices $\beta 9$, $\beta 10$, $\beta 11$, $\alpha 4$, $\beta 12$, $\beta 13$, and $\beta 14$. The sequence alignment for five proteins (PgaE, LanE, CabE, UrdE, AlpG, and GilOI) is shown below the diagram. The residues are color-coded: red for conserved, blue for variable, and green for specific mutations. The alignment is as follows:

Protein	130	140	150	160	170	180	190
PgaE	DGAC	VTVEVR	GP	EGKH	TLRA	AYLV	GCDGGR
LanE	HGDA	VEVEVR	GP	EGLR	TLRA	AYLV	GCDGGR
CabE	EGDH	VVEVR	GP	DGPR	SLT	TRYV	VGCDGGR
UrdE	HGDH	VEA	VEVR	GP	DEKI	RLTA	PYLV
JadH	DDD	AVE	ITV	RT	PE	GER	RLRA
AlpG	QDGD	SV	ITV	RT	PE	GER	RLRA
GilOI	TGTG	YR	C	V	L	S	GP

Diagram showing the fourth domain of the protein structure, including helices $\beta 15$, $\beta 16$, $\alpha 5$, $\beta 17$, and $\beta 18$. The sequence alignment for five proteins (PgaE, LanE, CabE, UrdE, AlpG, and GilOI) is shown below the diagram. The residues are color-coded: red for conserved, blue for variable, and green for specific mutations. The alignment is as follows:

Protein	200	210	220	230	240	250	260
PgaE	LPG	GMVM	VGP	PL	PGG	ITRI	IV
LanE	LPG	GMVM	VGP	PL	PGG	ITRI	IV
CabE	VPL	GMVM	SAP	PL	GDG	VDR	IV
UrdE	HPG	GMVM	SAP	PL	GDG	VDR	IV
JadH	LPG	GMVM	SAP	PL	GDG	VDR	IV
AlpG	LNG	GMVM	SAP	PL	GDG	VDR	IV
GilOI	HEK	GMVM	SAP	PL	GDG	VDR	IV

Diagram showing the fifth domain of the protein structure, including helices $\beta 19$, $\beta 20$, $\eta 4$, $\beta 21$, $\alpha 6$, and $\alpha 7$. The sequence alignment for five proteins (PgaE, LanE, CabE, UrdE, AlpG, and GilOI) is shown below the diagram. The residues are color-coded: red for conserved, blue for variable, and green for specific mutations. The alignment is as follows:

Protein	270	280	290	300	310	320	330
PgaE	QVTE	YRR	GRV	IL	LAGD	SAH	IHL
LanE	QASE	YRR	GRV	IL	LAGD	SAH	IHL
CabE	QVS	YRR	GRV	IL	LAGD	SAH	IHL
UrdE	LVTE	YRR	GRV	IL	LAGD	SAH	IHL
JadH	QATE	YRR	GRV	IL	LAGD	SAH	IHL
AlpG	QAAH	YRR	GRV	IL	LAGD	SAH	IHL
GilOI	LVE	S	YR	S	G	R	A

Diagram showing the sixth domain of the protein structure, including helices $\eta 5$, $\alpha 8$, and $\alpha 9$. The sequence alignment for five proteins (PgaE, LanE, CabE, UrdE, AlpG, and GilOI) is shown below the diagram. The residues are color-coded: red for conserved, blue for variable, and green for specific mutations. The alignment is as follows:

Protein	340	350	360	370	380	390
PgaE	MNT	QAQ	LL	FL	SG	PEV
LanE	MNT	QAQ	LL	LL	GG	PEV
CabE	MNT	QAQ	LL	FL	SG	PEV
UrdE	MNT	QAQ	LL	FL	SG	PEV
JadH	MNT	QAQ	LL	FL	SG	PEV
AlpG	MNT	QAQ	LL	FL	SG	PEV
GilOI	MNT	QAQ	LL	FL	SG	PEV

Figure S3. Structure-based sequence alignment of homologous proteins PgaE, LanE, CabE, UrdE, JadH, AlpG and GilOI. Regions investigated in this study are shown in light blue.

D

E

F

Figure S4. Time courses (left panels) and initial rates (right panels) of prejadomycin conversion by wild-type and variant PgaE enzymes. The best-fit curves to the model described in Figure S5A are plotted in red. A) wild-type PgaE, B) P78Q mutant, C) I79F mutant, D) $\beta 4$ chimera, E) $\beta 5$ chimera, F) $\beta 4$ - $\beta 5$ chimera, G) P78Q/I79F double mutant and H) H73A mutant.

Figure S5. Kinetic analysis of prejandomycin conversion by wild-type and variant PgaE enzymes. Prejandomycin conversion time courses (Figure S4) were fit to two-prejandomycin-binding-site model (A-C) and one-prejandomycin-binding-site model incorporating competitive inhibition by the C-12 hydroxylated reaction product (D). Each panel depicts the kinetic scheme and the corresponding equation (top) followed by bar graph and table (bottom) describing the best fit parameters. The

difference between the wild-type and variant PgaE enzymes is modeled as: (A) changes in affinity for substrate assuming independent binding of prejadomycin to two sites; (B) changes in the ratio of activities of bi-substrate and mono-substrate complexes assuming independent binding of prejadomycin to two sites; (C) changes in the ratio of dissociation constants of bi-substrate and mono-substrate complexes. For each panel the variable parameter intended to account for the variations between the PgaE enzymes is depicted in red.

Figure S6. Stereoview image of the electron density map (2Fo-Fc) around the $\beta 4$ - $\beta 5$ region of the PgaE P78Q/I79F double mutant at 2.4 Å resolution ($\sigma = 1.2$).

Table S1. Oligonucleotide primers used in this study. Regions complementary to the template are written with upper case and non-complementary with lower case; sequences complementary to *jadH* are in grey and restriction sites in outer primers in red.

Region	Fragment	Primer	Sequence
β4-β5	N-terminal	FOR	5'-GTAACAAAGCGGGACCAAAGC-3'
		REV	5'-gggtgtagtcgaactgcacaccgccgaagtggcccatcggGCTGGTCTCGACCTCGCC-3'
	C-terminal	FOR	5'-ccgatgggccacttcggcggtgtgcagttcgactacaccGTCCTCGAAGGCGCCTGG-3'
		REV	5'-tat aagctt TCAGCCGGTGAGCGGTGC-3'
β4	N-terminal	FOR	5'-GTAACAAAGCGGGACCAAAGC-3'
		REV	5'-GAAGTGTCccatcggGCTGGTCTCGACCTCGCC-3'
	C-terminal	FOR	5'-GAGACCAGCccgatgGGACACTTCGGCGGGCTC-3'
		REV	5'-tat aagctt TCAGCCGGTGAGCGGTGC-3'
β5	N-terminal	FOR	5'-GTAACAAAGCGGGACCAAAGC-3'
		REV	5'-GACgggtgtagtcgaactgcacCCCGCCGAAGTGTCCTG-3'
	C-terminal	FOR	5'-GGGgtgcagttcgactacaccGTCCTCGAAGGCGCCTGG-3'
		REV	5'-tat aagctt TCAGCCGGTGAGCGGTGC-3'
β6	N-terminal	FOR	5'-GTAACAAAGCGGGACCAAAGC-3'
		REV	5'-GGGGgatgcgcgGGCGGCCTGCCAGGC-3'
	C-terminal	FOR	5'-CGCCcgcgcatcCCCCAGTCGGTGACCG-3'
		REV	5'-tat aagctt TCAGCCGGTGAGCGGTGC-3'
β14	N-terminal	FOR	5'-GTAACAAAGCGGGACCAAAGC-3'
		REV	5'-TCTCACCgaggaaGCGGGGCTGGAGTTCTAC-3'
	C-terminal	FOR	5'-AGCCCCGcttctcGGTGAGACGCTGCCCCGGC-3'
		REV	5'-tat aagctt TCAGCCGGTGAGCGGTGC-3'
β15-β16	N-terminal	FOR	5'-GTAACAAAGCGGGACCAAAGC-3'
		REV	5'-gtcgcgcccctcggcgagcggcgccgcCATGACCATGCCGCCGGGCA-3'
	C-terminal	FOR	5'-gcggcgccgctcgccgagggcgctcgacCGCATCATCGTCTGCGAG-3'
		REV	5'-tat aagctt TCAGCCGGTGAGCGGTGC-3'
P78Q/I79F	N-terminal	FOR	5'-GTAACAAAGCGGGACCAAAGC-3'
		REV	5'-CACCGAAGTCGAaCtGGAGCCC-3'
	C-terminal	FOR	5'-GGGCTCCaGtTCGACTTCGGTG-3'
		REV	5'-tat aagctt TCAGCCGGTGAGCGGTGC-3'
P78Q	N-terminal	FOR	5'-GTAACAAAGCGGGACCAAAGC-3'
		REV	5'-AAGTCGATctgGAGCCCGCCGAAGTG-3'
	C-terminal	FOR	5'-GCGGGCTCagATCGACTTCGGTGTCTCG-3'
		REV	5'-tat aagctt TCAGCCGGTGAGCGGTGC-3'
I79F	N-terminal	FOR	5'-GTAACAAAGCGGGACCAAAGC-3'
		REV	5'-CACCGAAGTCGAaCGGGAGCCC-3'
	C-terminal	FOR	5'-GGGCTCCCGtTCGACTTCGGTG-3'
		REV	5'-tat aagctt TCAGCCGGTGAGCGGTGC-3'
H73A	N-terminal	FOR	5'-tga agatct GACGCTGCCGTGATTGTC-3'
		REV	5'-GCCCGCCGAAGgcTCCCTGGGT-3'
	C-terminal	FOR	5'-ACCCAGGGAGcCTTCGGCGGGC-3'
		REV	5'-cag aagctt tcaGCCGGTGAGCGG-3'

Table S2. X-ray data collection and crystallographic refinement statistics. Values in parenthesis are for the highest resolution shell.

PgaE P78Q/ I79F	
Data collection	
PDB entry	4ICY
X-ray source	ESRF ID14-1
Space group	F222
Unit cell (Å)	a=66.0 b=171.0 c=212.9
Resolution (Å)	46.5 – 2.4 (2.53-2.40)
Wavelength (Å)	0.933
No. unique reflections	23686
Multiplicity	5.4 (3.7)
Completeness (%)	99.5 (97.3)
Mean (<I> / <σ _I >)	16.3 (2.4)
R-sym (%)	6.6 (52.2)
Wilson B-factor (Å ²)	49.6
Refinement	
R-factor (all reflections) (%)	19.9
R-free (%) ^a	25.9
Number of atoms	3486
Number of water molecules	113
Number of other molecules	1 FAD / 2 CL / 2 SO4 4 EDO / 2 GOL
RMSD bond lengths (Å)	0.009
RMSD bond angles (°)	1.340
Average B-factor (Å ²)	
All atoms	52.4
Protein	52.6
Water	44.0
FAD	44.0
Other molecules	69.1
Ramachandran plot	
Favoured regions (%)	435 (96.0)
Allowed regions (%)	17 (3.8)
Outliers (%)	1 (0.2)

^a 5 % of the reflections were used in the R-free calculations.

Legend: FAD, flavin adenine dinucleotide; CL, chloride; SO4, sulphate; EDO, ethylene glycol; GOL, glycerol.