

A Predictive Model for Epoxide Hydrolase-generated Stereochemistry in the Biosynthesis of 9-Membered Eneidyne Antitumor Antibiotics

Geoffrey P. Horsman,^{†,¶} Anna Lechner,[‡] Yasuo Ohnishi,[§] Bradley S. Moore,[‡] and Ben Shen^{*,†,||,Ⓜ,¥}

[†]Division of Pharmaceutical Sciences, University of Wisconsin-Madison, Madison, WI 53705, USA; [‡]Scripps Institution of Oceanography and the Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California-San Diego, La Jolla, CA 92093, USA; [§]Department of Biotechnology, University of Tokyo, Tokyo, Japan; and ^{||}Department of Chemistry, ⁺Department of Molecular Therapeutics, and [¥]Natural Products Library Initiative at The Scripps Research Institute, The Scripps Research Institute, Jupiter, FL 33458, USA.

[¶]Present address: Department of Chemistry, Wilfrid Laurier University, Waterloo, ON N2L 3C5, Canada.

*To whom correspondence should be addressed: The Scripps Research Institute, 130 Scripps Way, #3A1, Jupiter, FL 33458. Tel: (561) 228-2456. Fax: (561) 228-2472. Email: shenb@scripps.edu

Supporting Information

Table S1. Primers used for described cloning and mutagenesis procedures	S2
Table S2. Plasmids used	S3
Table S3. Sequence homology among enediyne biosynthesis-associated EHs	S4
Table S4. EH sequences used to construct phylogram in Fig. 6B.	S5
Figure S1. Sequence alignment of enediyne biosynthesis-associated EHs	S8
Figure S2. SDS-PAGE analysis on a 12% acrylamide gel of purified EHs	S9
Figure S3. Steady-state kinetic analysis of SgcF variants, SpoF, and SgrF	S10

Table S1. Primers used for described cloning and mutagenesis procedures

Primer Name	Sequence
SGR625-NdeI-F	5'-CATATGGCACATCACCACCACCATCACGTGGAT GACGACGACAAGGTGCGCCCCTTCCGCATCGAC -3'
SGR625-XhoI-R	5'- <u>CTCGAGT</u> CACCGCAGTGAACGGGCGAAC-3'
SpoF-NdeI-F	5'-CATATGGCACATCACCACCACCATCACGTGGAT GACGACGACAAGATGCGCCCGTTCCGGATCCATG -3'
SpoF-XhoI-R	5'- <u>CTCGAGT</u> CAGGGCAGCGGACGGATGAAG-3'
SgcF-W236Y-F	5'-GGCTCGGTCGCCGAGTTCC-3'
SgcF-W236Y-mutF	5'-GGATGCGCTG GT AGCCGATGC-3'
SgcF-W236Y-mutF	5'-GCATCGGCT ACC AGCGCATCC-3'
SgcF-W236Y-R	5'-CATCACGTGGGTGAGCAGCC-3'
SgcF-W176L-mutF	5'-GGCGGCGAC CTGGG AAAGGTTCG-3'
SgcF-W176L-mutR	5'-CGACCTT TCC AGGTTCGCCGCC-3'
SgcF-Q237M-mutF	5'-CATCGGCTGG AT GCGCATCCAG-3'
SgcF-Q237M-mutR	5'-CTGGATGCG CAT CCAGCCGATG-3'
SgcF-W236Y/Q237M-mutF	5'-CATCGGCTAC AT GCGCATCCAGC-3'
SgcF-W236Y/Q237M-mutR	5'-GCTGGATGCG CAT GTAGCCGATG-3'
NcsF2-Y235W-F	5'-TTCGGTCCCGAGTTCCTCG-3'
NcsF2-Y235W-R	5'-CTAGGACGTCCGCAGCGCG-3'
NcsF2-Y235W-mutF	5'-GGCTCCGGG TGG ATGAAGATCC-3'
NcsF2-Y235W-mutR	5'-GGATCTTCAT CC ACCCGGAGCC-3'
NcsF2-L176W-mutF	5'-GGCGGCGAT TGGGG TGCCTGG-3'
NcsF2-L176W-mutR	5'-CCAGGCAC CCCA ATCGCCGCC-3'
NcsF2-M236Q-mutF	5'-TCCGGGTAC CA GAAGATCCAGTCG-3'
NcsF2-M236Q-mutR	5'-CGACTGGATCTT CTGGT ACCCGGA-3'
NcsF2-Y235W/M236Q-mutF	5'-CCGGGTGG CA GAAGATCCAGTCG-3'
NcsF2-Y235W/M236Q-mutR	5'-CGACTGGATCTT CTG CCACCCGG-3'

Bold letters represent bases added or mismatched from template. Restriction sites are underlined and mutated codons are italicized.

Table S2. Plasmids used

Plasmid	Description	Reference
pBS1096	Stp ^R ; <i>sgcF</i> in pCDF-2 Ek/LIC	3
pBS5042	Stp ^R ; <i>ncsF2</i> in pCDF-2 Ek/LIC	4
pBS1118	Amp ^R ; <i>sgrF</i> in pGEM-T Easy	This study
pBS1119	Kan ^R ; <i>sgrF</i> in pET29	This study
pBS1120	Amp ^R ; <i>spoF</i> in pGEM-T Easy	This study
pBS1121	Kan ^R ; <i>spoF</i> in pET29	This study
pBS1122	Amp ^R ; 566 bp fragment of <i>sgcF</i> W236Y in pGEM-T Easy	This study
pBS1123	Stp ^R ; <i>sgcF</i> W236Y in pCDF-2 Ek/LIC	This study
pBS1124	Amp ^R ; 566 bp fragment of <i>sgcF</i> W176L in pGEM-T Easy	This study
pBS1125	Stp ^R ; <i>sgcF</i> W176L in pCDF-2 Ek/LIC	This study
pBS1126	Amp ^R ; 566 bp fragment of <i>sgcF</i> Q237M in pGEM-T Easy	This study
pBS1127	Stp ^R ; <i>sgcF</i> Q237M in pCDF-2 Ek/LIC	This study
pBS1128	Amp ^R ; 566 bp of <i>sgcF</i> W236Y/Q237M in pGEM-T Easy	This study
pBS1129	Stp ^R ; <i>sgcF</i> W236Y/Q237M in pCDF-2 Ek/LIC	This study
pBS1130	Amp ^R ; <i>sgcF</i> W176L/Q237M in pGEM-T Easy	This study
pBS1131	Amp ^R ; <i>sgcF</i> W176L/W236Y/Q237M in pGEM-T Easy	This study
pBS1132	Stp ^R ; <i>sgcF</i> W176L/Q237M in pCDF-2 Ek/LIC	This study
pBS1133	Stp ^R ; <i>sgcF</i> W176L/W236Y/Q237M in pCDF-2 Ek/LIC	This study
pBS5043	Amp ^R ; 860 bp fragment of <i>ncsF2</i> Y235W in pGEM-T Easy	This study
pBS5044	Stp ^R ; <i>ncsF2</i> Y235W in pCDF-2 Ek/LIC	This study
pBS5045	Amp ^R ; 860 bp of <i>ncsF2</i> L176W in pGEM-T Easy	This study
pBS5046	Stp ^R ; <i>ncsF2</i> L176W in pCDF-2 Ek/LIC	This study
pBS5047	Amp ^R ; 860 bp of <i>ncsF2</i> M236Q in pGEM-T Easy	This study
pBS5048	Stp ^R ; <i>ncsF2</i> M236Q in pCDF-2 Ek/LIC	This study
pBS5049	Amp ^R ; 860 bp of <i>ncsF2</i> Y235W/M236Q in pGEM-T Easy	This study
pBS5050	Stp ^R ; <i>ncsF2</i> Y235W/M236Q in pCDF-2 Ek/LIC	This study
pBS5051	Stp ^R ; <i>ncsF2</i> L176W/M236Q in pCDF-2 Ek/LIC	This study
pBS5052	Stp ^R ; <i>ncsF2</i> L176W/Y235W/M236Q in pCDF-2 Ek/LIC	This study

Resistance marked as: Amp, ampicillin; Stp, streptomycin; Kan, kanamycin

Table S3. Percent sequence identities/similarities between enediyne biosynthesis-associated EHs described in this study

	NcsF2	KedF	SgcF	SgrF	SpoF
NcsF2	-	63/74	62/75	63/73	64/75
KedF		-	64/75	65/75	67/77
SgcF			-	72/81	70/83
SgrF				-	72/82
SpoF					-

Table S4. List of EH sequences used to construct the phylogram in Fig. 6B.

Organism	Epoxide hydrolase sequences
Prokaryotes	
Acinetobacter sp. ADP1	YP_045093
Agrobacterium tumefaciens str. C58	NP_534160; NP_535865; NP_535669
Agrobacterium radiobacter AD1	CAA73331
Anabaena variabilis ATCC 29413	YP_320385
Azotobacter vinelandii AvOP	ZP_00417649; ZP_00417027; ZP_00419831
Bacillus subtilis subsp. subtilis str. 168	NP_388739
Bordetella bronchiseptica RB50	NP_887793
Bradrhizobium japonicum USDA 110	NP_767058; NP_774828; NP_767754; NP_771160; NP_769521; NP_773216; NP_773708; NP_774534; NP_767837; NP_774008; NP_772904; NP_770058; NP_774137
Brevibacterium linens BL2	ZP_00381314
Burkholderia fungorum LB400	ZP_00284192; ZP_00278609; ZP_00277201; ZP_00282077; ZP_00278366; ZP_00278122
Burkholderia pseudomallei K96243	YP_110633; YP_107687; YP_110932; YP_106956
Burkholderia thailandensis E264	YP_442876; YP_441357; YP_440002; YP_439655; YP_439670; YP_440442
Caulobacter crescentus CB15	NP_421885; NP_420043
Chloroflexus aurantiacus J-10-fl	ZP_00768880
Chromobacterium violaceum ATCC 12472	NP_901164
Clostridium acetobutylicum	NP_347356
Colwellia psychrerythraea 34H	YP_267577
Corynebacterium sp.	CAA11900
Deinococcus radiodurans R1	NP_296269
Gloeobacter violaceus PCC 7421	NP_924227; NP_924368; NP_927205
Kineococcus radiotolerans SRS30216	ZP_00615612; ZP_00615274; ZP_00617546
Leptospira interrogans serovar Lai str. 56601	NP_713968; NP_712707; NP_713280
Mesorhizobium loti MAFF303099	NP_107141; NP_104428; NP_107140; NP_105306; NP_103292
Mycobacterium avium subsp. paratuberculosis str. K10	NP_959380; NP_962216; NP_961601; NP_960889; NP_961819
Mycobacterium tuberculosis H37v	NP_216350; NP_218134; NP_216454; NP_215640; NP_216730; NP_214648
Nocardia farcinica IFM 10152	YP_118952; YP_117793; YP_119793; YP_116211
Nostoc punctiforme PCC 73102	ZP_00108314; ZP_00107532
<i>Nostoc</i> sp. PCC 7120	NP_488261; NP_484083
Novosphingobium aromaticivorans DSM 12444	ZP_00873504; ZP_00872884

Polaromonas sp. JS666	ZP_00503276
Pseudomonas aeruginosa PAO1	NP_252119; NP_252014; NP_251624; NP_252683; NP_250776
Pseudomonas aeruginosa UCBPP-PA14	
Pseudomonas fluorescens Pf-5	YP_258091; YP_260055; YP_261201
Pseudomonas fluorescens PfO-1	YP_347986; YP_347987; YP_347569; YP_346634; YP_348784; YP_348148; YP_351427; YP_349650
Pseudomonas putida KT2440	NP_746733
Ralstonia metallidurans CH34	ZP_00595941; ZP_00597123; ZP_00598949; ZP_00597041; ZP_00597898; ZP_00564175; ZP_00594247
Ralstonia solanacearum GMI1000	NP_519891
<i>Rhizobium</i> sp. NGR234	AAQ87477 CAB59813
Rhodopirellula baltica SH-1	NP_866425; NP_868953
Rhodopseudomonas palustris CGA900	NP_948077; NP_948116
Rubrivivax gelatinosus PM1	ZP_00243135
Salmonella typhimurium LT2	NP_459327
Silicibacter pomeroyi DSS-3	YP_165203; YP_166503; YP_166740
Silicibacter sp. TM1040	ZP_00621680
Sinorhizobium meliloti 1021	NP_436749
Streptomyces avermitilis MA-4680	NP_824987; NP_823281; NP_821315; NP_823898; NP_825429
Streptomyces carzinostaticus subsp. Neocarzinostaticus	AAM78002; AAM77995
Streptomyces coelicolor A3(2)	NP_627771; NP_628138; NP_630689; NP_627764; NP_625391; NP_630375; NP_631261; NP_625198
Streptomyces globisporus	AAL06662
<i>Synechocystis</i> sp. PCC 6803	NP_440962
Thermobifida fusca YX	YP_289300
Thermosynechococcus elongatus BP-1	NP_681708; NP_682856
Thrichodesium erythraeum IMS101	ZP_00672579
Eukaryotes	
Aspergillus nidulans FGSC A4	XP_660621; XP_662866
Candida albicans SC5314	XP_719809
Ceanorhabditis briggsae	CAE64799; CAE71892
Ceanorhabditis elegans	NP_504650; NP_497268; NP_505662
Cryptococcus neoformans var. neoformans B-3501A	EAL17134
<i>Danio rerio</i> (zebra fish)	NP_957362; AAH55594
Debaromyces hansenii CBS 767	XP_456371
Giberella zeae PH-1	XP_391372; XP_381941; XP_381892; XP_386920; XP_385919; XP_383988; XP_381913; XP_387982; XP_391218;

	XP_383909
Homo sapiens	AAP36260
Homo sapiens	P07099
Homo sapiens	P34913
Kluyveromyces lactis NRRL Y-1140	XP_451455
Magnaporthe grisea 70-15	XP_359602; XP_366994; XP_369638; XP_364758; XP_360844; XP_364277; XP_368501
Manduca sexta	Q25489
Mus musculus	NP_031966
Neurospora crassa	XP_322436; XP_330900; XP_330112
Oryza sativa (rice)	NP_912787; NP_922168; NP_922166; XP_470157
Saccharomyces cerevisiae	NP_014462
Saccharomyces paradoxus NRRL Y-17217	DAA02268;
Solanum tuberosum	AAA81892; AAA81893
Tetraodon nigroviridis	CAG05638; CAG00866
Yarrowia lipolytica CLIB122	XP_502171; XP_499652; XP_504164
Other	
Environmental DNA libraries ¹	BD8877; BD9300; BD9883; BD9884; BD10332; BD10721

- (1) Zhao, L.; Han, B.; Huang, Z.; Miller, M.; Huang, H.; Malashock, D. S.; Zhu, Z.; Milan, A.; Robertson, D. E.; Weiner, D. P.; Burk, M. J. (2004) Epoxide hydrolase-catalyzed enantioselective synthesis of chiral 1,2-diols via desymmetrization of meso-epoxides, *J. Am. Chem. Soc.* 126, 11156-11157.

Figure S2. SDS-PAGE analysis on a 12% acrylamide gel of purified enzymes. (A) SgcF variants as follows: lane 1, molecular weight markers; lane 2, WT; lane 3, W236Y; lane 4, Q237M; lane 5, W236Y/Q237M. (B) SpoF (lane 2) and SgrF (lane 3), with molecular weight markers in lane 1.

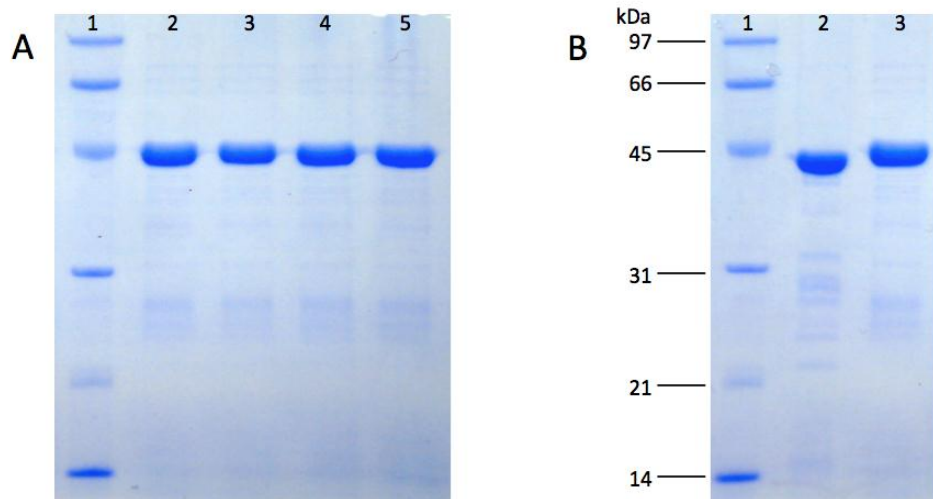


Figure S3. Steady-state kinetic analysis of (A) SgcF (W236Y), (B) SgcF (Q237M), (C) SgcF (W236Y/Q237M), (D) SpoF, and (E) SgrF towards (*S*- (left) and (*R*)-styrene oxide (right)). Solid lines represent best fits of appropriate equations to initial rate data (black squares) as follows: Michaelis-Menten equation for B [(*R*)-styrene oxide], D, and E; substrate inhibition for A and C; linear fit for estimation of k_{cat}/K_m in B [(*S*)-styrene oxide].

