Characterization of the Epoxide Hydrolase NcsF2 from the Neocarzinostatin Biosynthetic Gene Cluster

Shuangjun Lin,[†] Geoffrey P. Horsman,[†] and Ben Shen^{*,†,‡,§,}

[†]Division of Pharmaceutical Sciences, [‡]University of Wisconsin National Cooperative Drug Discovery Group, and [§]Department of Chemistry, University of Wisconsin-Madison, Madison, Wisconsin 53705, USA

bshen@pharmacy.wisc.edu

Supporting Information

General experimental procedures	S2
Cloning, overproduction, and purification of NcsF1 and NcsF2	S2
EH activity assays with styrene oxide as a substrate mimic	S2
Large-scale reaction with NcsF2	S 3
Enantioconvergent reaction using NcsF2/SgcF co-catalysis	S 3
Figure S1. SDS-PAGE analysis of purified NcsF1 and NcsF2	S4
Figure S2. Sequence alignment of NcsF1 and NcsF2 with selected EHs	S5
References	S 6

General experimental procedures

Substrates including racemic styrene oxide, (R)-styrene oxide, and (S)-styrene oxide, and products including racemic 1-phenyl-1, 2-ethanediol, (R)-1-phenyl-1, 2-ethanediol. (S)-1-phenyl-1, 2-ethanediol, and $[^{18}O]$ -H₂O were purchased from Sigma-Aldrich (St. Louis, MO). Dithiothreitol (DTT) was purchased from Research Products International Corp (Mt. Prospect, IL). Complete protease inhibitor tablet, EDTA-free, was from Roche Applied Science (Indianapolis, IN). Medium components and buffers were from Fisher Scientific (Pittsburgh, Synthetic DNA oligonucleotides were purchased from the University of PA). Wisconsin-Madison Biotechnology Center. PCR was performed with a PerkinElmer GeneAmp 2400. Electrospray ionization-mass spectroscopy (ESI-MS) was performed with an Agilent 1100 MSD SL ion trap mass spectrometer (Agilent Technologies, Inc. Santa Clara, CA). NMR spectra were recorded using a Varian UI-500 spectrometer (Varian, Inc., Palo Alto, CA). High performance liquid chromatography (HPLC) analyses were carried out on a Varian HPLC system equipped with Prostar 210 pumps, a photodiode array (PDA) detector, and an Alltech Appolo C18 reverse phase column (5 μ m, 4.6 \times 250 mm, Grace Davison Discovery Sciences, Deerfield, IL), using a 20 min linear gradient from 0 to 60% acetonitrile in water. The enantiomeric separation was performed on a Waters HPLC system equipped with 600 pumps, a 996 photodiode array (PDA) detector, and a Chiralcel OD-H chiral column (5 μ m, 4.6 \times 250 mm, Grace Davison Discovery Sciences) using a 60 min isocratic elution with 2.5% isopropanol in *n*-hexane.

Cloning, overproduction, and purification of NcsF1 and NcsF2

The gene encoding NcsF1 and NcsF2 were amplified by PCR from cosmid pBS50041 as template. PCRs were performed with Platinum Pfx polymerase (Invitrogen, Carlsbad, CA) using the following primers: forward 5'- GAC GAC GAC AAG ATG CGG CCG TTT CAG GTC C -3' and reverse 5'- GAG GAG AAG CCC GG TCA GAG CTT CTT GAG TGT CC -3' for NcsF1; forward 5'- GAC GAC GAC AAG ATG GAG CCG TTC CGT ATA G -3' and reverse 5'- GAG GAG AAG CCC GGT CTA GGA CGT CCG CAG CG -3' (starting and stop codons underlined) for NcsF2. The gel-purified PCR products were directly cloned into the pCDF-2 Ek/LIC vector using ligation-independent cloning as described by Novagen (Madison, WI) to give pBS5041 (for NcsF1) and pBS5042 (for NcsF2) for E. coli expression. The over-expression in E. coli BL21 (DE3) and purification of the recombinant protein by affinity chromatography using NTA-Ni agarose column (Qiagen, Valencia, CA) were performed following the previously described procedure for SgcF purification.² After the purified protein was desalted using a PD-10 column (GE Healthcare, Piscataway, NJ) twice to completely remove glycerol and then concentrated with a 10 K MWCO Vivaspin ultrafiltration device (Sartorius, Edgewood, NY), the NcsF2 protein was stored at -80 °C in 100 µL aliquots. The purity of the purified protein was assessed by 12% SDS-PAGE and the concentration was determined from the absorbance at 280 nm using a molar absorptivity ($\varepsilon = 68.41 \text{ mM}^{-1} \text{ cm}^{-1}$ for using the program NcsF1. 82.34 for NcsF2) calculated Protparam (http://www.expasy.ch/tools/protparam.html).

EH activity assays with styrene oxide as a substrate mimic

HPLC-based assays were carried out in 200 μ L reaction mixtures containing 2 mM racemic styrene oxide and 50 mM phosphate buffer (pH 7.5). The reaction was initiated by the addition of 50 μ M enzyme (NcsF1 or NcsF2) and incubated at 25 °C for 1 h. The reaction was quenched by extraction with 200 μ L ethyl acetate for three times and organic extracts were combined. After solvent was removed using a speed-vac, the resulting residue was dissolved in 50 μ L acetonitrile and 25 μ L was subjected to HPLC analysis. Control reactions were carried out under the identical conditions except with enzyme that had been boiled for 5 min.

To determine the kinetic parameters of NcsF2-catalyzed hydrolysis of styrene oxide, a spectrophotometric assay method was adopted to achieve continuous and accurate determination of epoxide hydrolase activity.^{3,4} Prior to kinetic analysis, the optimal pH for NcsF2 activity was tested in three different buffers: 50 mM sodium acetate (pH 5.0, 5.5, and 6.0); 50 mM sodium phosphate buffer (pH 5.5 to pH 8.5); and 50 mM glycine-NaOH (pH 8.5, 9.0, and 10.0). The reactions were carried out in 1 mL mixture containing 10 μ L of 300 mM sodium periodate in DMF, 20 μ L of 5 M styrene oxide in DMSO, and 50 mM buffer. The reactions were initiated by the addition of NcsF2 and carried out in triplicate. The absorbence at 290 nm was monitored in a 1-mL quartz cell and the velocity was calculated based on the rate of change of absorbance over several minutes. Steady-state kinetic parameters were obtained from reactions carried out in 50 mM phosphate buffer (pH 8.0) with varying styrene oxide concentrations [0.1-6.4 mM for (*R*)-styrene oxide with 2 μ M NcsF2 and 1-16 mM for (*S*)-styrene oxide with 8 μ M NcsF2]. The Michaelis-Menten equation was fitted to plots of velocity of 1-phenyl-1, 2-ethanediol formation versus substrate concentration to extract values for *K*_M and *k*_{cat}.

Large scale reaction of NcsF2

To prepare sufficient quantities of product for NMR characterization, a reaction was carried out in a 1-mL reaction mixture containing 10 mM styrene oxide and 200 μ M NcsF2 in 50 mM phosphate buffer (pH 8.0). During the course of the reaction, 10 mM styrene oxide and 50 μ M NcsF2 were added every hour. The reaction was allowed to proceed at 25 °C for 8 h, after which the products were extracted from the reaction mixture with 5 x 0.5 mL of ethyl acetate. The solvent was removed by evaporation under reduced pressure. ¹⁸O-labeled products were obtained following a similar procedure. The reactions were carried out in 0.5-mL reaction mixtures containing [¹⁸O]-H₂O. For (*R*)-styrene oxide, 5 mM substrate and 6 μ M NcsF2 were added every 30 min (~80% [¹⁸O]-H₂O final concentration). For (*S*)-styrene oxide, 20 mM substrate and 15 μ M NcsF2 were added every hour (~60% [¹⁸O]-H₂O final concentration).

Enantioconvergent reaction using NcsF2/SgcF co-catalysis

An enantioconvergent enzymatic preparation of (*R*)-1-phenyl-1,2-ethanediol was performed in a 10 mL reaction containing 14 mM racemic styrene oxide (16.8 mg), 35 μ M SgcF, and 10 μ M NcsF2 in 100 mM phosphate buffer (pH 8.0). After the reaction was incubated at room temperature for 2 h, another batch of fresh enzyme (22.5 μ M SgcF and 6.4 μ M NcsF2) was added to the reaction mixture and incubated at room temperature for an additional 3 h. The reaction mixture was extracted 5 times with 10 mL ethyl acetate after the styrene oxide was consumed as judged by TLC. The organic extracts were combined and dried with anhydrous Na₂SO₄. The removal of solvent gave 16.8 mg of (*R*)-1-phenyl-1,2-ethanediol (87% yield and 99% ee) as determined by chiral HPLC analysis.

Figure S1. SDS-PAGE analysis on a 12% acrylamide gel of purified NcsF1 (lane 1) and NcsF2 (lane 3) both with expected masses of ~44.5 kDa, and molecular weight markers (lane 2).



Fig. S2. Sequence alignment of NcsF1, NcsF2, and SgcF with selected epoxide hydrolase (EHs). The amino acid sequences were retrieved from NCBI with the following accession numbers: SgcF from *Streptomyces globisporus* (AAL06662); NcsF2 from *S. neocarzinostaticus* (AAM78002); NcsF1 from *S. neocarzinostaticus* (AAM77995); Bfueh1 from *Burkholderia fungorum* LB400 (ZP_00284192); AnEH from *Aspergillus niger* (CAB59813); human microsomal EH (P07099); juvenile hormone EH (Q25489). Shaded regions denote sequences conserved among EHs. Active site catalytic residues are marked as follows (NcsF2 numbering): ∇ , catalytic triad residues D¹⁷⁵, D³³⁶, H³⁶³; ●, oxirane-binding residues Y²³⁵ and Y³⁰⁴.

	** : : : **: : :	
SacE		27
NcsF2		27
NcsF1		26
Rfuchl		34
Drueni		40
Anen		92
Human	MWLEILLTSVLGFAIYWFISKDKEETLPLEDGWWGFGTKSAAKEDDSIRPFKVETSDEEIHDLHQKIDKFK-FT	/3
JHEH	-MYKILSSFVAGVAIGSGLVITYVLYNVPEPPELDLQR-WWGIGTRPTE-EDKSIRPFSIDFNDTVILDLKERLKNRRPFT	78
	* . : : * :::: :*:: :*: : *:: *:: *::	
SgcF	EIPGSGWDRGVPLSYLKELTDHWRHGYDWRAAEAELNAFPQFVTTIDGADVHFLHVRSPEPDAIPLILTHGWPG	101
NcsF2	EIPGTGWSRGVPLDYLKELVGYWRDGYDWRAAEDRLNTVPQFTTEIDGTNVHFMHIRSAEPDALPMIITHGWPG	101
NcsF1	ELVDVGWSRGAPLSYIKELAEYWRDGFDWRAAERRINQYPQFTTEIDGATIHFLHVRSPEPDATPMVITHGWPG	100
Bfueh1	RETVDDQSQGAQLDALRKLVEYWRTQYDWRRVEARINGYPQYKADIDGLGIHFLHIRSRHDNALPVVMTHGWPG	108
AnEH	PTYESLQADGRFGITSEWLTTMREKWLSEFDWRPFEARLNSFPQFTTEIEGLTIHFAALFSEREDAVPIALLHGWPG	119
Human	PPLEDSCFHYGFNSNYLKKVISYWRNEFDWKKQVEILNRYPHFKTKIEGLDIHFIHVKPPQLPAGHTPKPLLMVHGWPG	152
JHEH	KPLEGINSEYGMNTEYLETVLEYWLNEYNFKKRAELLNKFPHYKTRIQGLDLHFIRVKP-EIKEGVQVLPLLMMHGWPS	156
	β1 β2 β3	
	: ** :: :	
SacF	SVAEFLDVIGPLSDPRAHGGDPADAFHVVVPSMPGYGFS-GPTAEPGWDVRRIARAWAELMNRLGYER-YVAOGGDWGKVV	180
NCSF2	SVAEFLDVIDPLTNPRAHGGDPADAFHLVIPSLPGFGFS-GPTPEPGWNLPRVASAWAELMRRLGVSR-VAVOGGDIGAWT	180
NcsF1	TPVEFLDIIGPLTDPRAHGGDPADAFHLVIPSLPGFGLS-GPLKSAGWELGRIAMAWSKLMASLGYER-YIAOGGDIGAFT	179
Bfuehl	STVEFT DVI DLI VDPTAHCGNAFDAFHUVI PSTPCYCFS-DKPTARCWERNRIGRAWDTI MKSI CYDE-YVAOCCDWCSVI	187
ANEW	STUDE WITH THE PERFORMANCE AND A DESCRIPTION AND A DUNNAL WAS DUNNAL WAS A DUNNAL W	107
Human	SEVERITIES AND A CONTRACTOR AND A CONTRACT AND A CO	221
Human	SFIEFIKIIFLLIDFKNNGLDENVEVICFSIFGIG-EASSKNGFNSVAIAKIFIKLALGFQE-FILGGDWGSLI	231
JHEH	SKEPDKVIPILITPKREINTVFEVVAVDLPGIGFS-EGINKPGLNPVQIGVMMKNLMLKLGFEK-FIIQAGDWGSQC	232
	<u> </u>	
a b		
SQCF	r = r = r = r = r = r = r = r = r = r =	2.40
	SLELGLADPERVAGVHLNMLVTF PPQDAPEAIGKLDESDLGKLANSGEFADTGIGWQKIQATKPQTLA	249
NcsF2	SLEIGLADPEHVAGVHLNMLVIF PPODAPEAIGRLDESDLGKLANSGEFADIGIGWQKIQATRPOTLAY SLTLSGVDHEHVVGTHVNFLITPPSGDPADLAGLGEQDLARLQLLABFGAEGSGYMKIQSTRPQTLSY	249 248
NcsF2 NcsF1	SLELGLADPEHVAGVHLNMLVIF PPQDAPEAIGRLDESDLGKLANSGEFADIGIGWQKIQATRPQTLAY SLTLSGVDHEHVVGTHVNFLITPPSGDPADLAGLGEQDLARLQLLAEFGAEGSGYMKIQSTRPQTLSY SLLLGAIDPSHLAGIHVNLLQTNLSGEPGELETLSDADKARLAVSERFLDDLSGPMKMQSTRPHTIGY	249 248 247
NcsF2 NcsF1 Bfueh1	SLELGLADPERVAGYHLMLVIF PPQDAPEAIGRDESDLGKLABGEFADIGIGWQKIQATRPQTLAY SLTLSGVDHEHVVGTHVNFLITPPSGDPADLAGLGEQDLARLQLAEFGAEGSGYMKIQSTRPHTIGY SLLLGAIDPSHLAGIHVNLLQTNLSGEPGELETLSDADKARLAVSERFLDDLSGPMKMQSTRPHTIGY TTEMGRIKLEGLRAIHVNLPFVAPAQFPEVPTPDEQIAIDQCLRFASDGSFYHHLQTTRPQTIGY	249 248 247 252
NcsF2 NcsF1 Bfueh1 AnEH	SLELGLADFENVAGVHLNMLVIFFPQDAPEAIGRDESDLGKLANSGEFGAIGIGWQKIQATKPQTLAY SLILSGVDHEHVVGTHVNFLITPPSGDPADLAGLGEQDLARLQLLAEFGAEGSGYMKIQSTRPDTLSY SLLLGAIDPSHLAGIHVNLLQTNLSGEPGELETLSDADKARLAVSERFLDDLSGPMKMQSTRPHTIGY TTEMGRIKLEGLRAIHVNLPFVAPAQFPEVPTPDEQIAIDQCLRFASDGSFYHHLQTTRPQTIGY GRLLG-VGFDACKAVHLNLCAMRAPPEGPSIESLSAAEKEGIARMEKFMTDGLAYAMEHSTRPSTIGH	249 248 247 252 264
NcsF2 NcsF1 Bfueh1 AnEH Human	SLELGLADFENVAGVHLNMLVIFFPQDAPEAIGRLDESDLGKLANSGEFGAIGIGWQKIQATKPQTLAY SLTLSGVDHEHVVGTHVNFLITPPSGDPADLAGLGEQDLARLQLLAEFGAEGSGYMKIQSTRPDISY SLLLGAIDPSHLAGIHVNLLQTNLSGEPGELETLSDADKARLAVSERFLDDLSGPMKMQSTRPHTIGY TTEMGRIKLEGLRAIHVNLPFVAPAQFPEVPTPDEQIAIDQCLRFASDGSFYHHLQTTRPQTIGY GRLLG-VGFDACKAVHLNLCAMRAPPEGPSIESLSAAEKEGIARMEKFMTDGLAYAMEHSTRPSTIGH CTNMAQLVPSHVKGLHLNMALVLSNFSTLTLLLGQRFGRFLGLTERDVELLYPVKEKVFYSLMRESGYMHIQCTKPDTVGS	249 248 247 252 264 312
NcsF2 NcsF1 Bfueh1 AnEH Human JHEH	SLELGLADFENVAGVHLNALVIFFPQDAPEAIGRLDESDLGKLANSGEFGAIGIGWQHLQATKPQTLAY SLTLSGVDHEHVVGTHVNFLITPPSGDPADLAGLGEQDLARLQLLAEFGAEGSGYMKIQSTRPQTLSY SLLLGAIDPSHLAGIHVNLLQTNLSGEPGELETDEQIAIDQCLRFADDLSGPMKMQSTRPHTIGY TTEMGRIKLEGLRAIHVNLPFVAPAQFPEVPTPDEQIAIDQCLRFASDGSFYHHLQTTRPQTIGY GRLLG-VGFDACKAVHLNLCAMRAPPEGPSIESLSAAEKEGIARMEKFMTDGLAYAMEHSTRPSTIGH CTNMAQLVPSHVKGLHLNMALVLSNFSTLTLLLGQRFGRFLGLTERDVELLYPVKEKVFYSLMRESGYMHIQCTKPDTVGS ATHMATLFPDQVLGLHTNMPLSSRPLSTVKLFIGALFPSLI-VDAKYMDRIYPLKN-LFSYILRETGYFHIQATKPDTIGV	249 248 247 252 264 312 311
NCSF2 NCSF1 Bfueh1 AnEH Human JHEH	SLELGLADPENVAGNILMELTIF PPQDAPEAIGRLGESDLGKLANSGEFADIGIGWQXIQATKPQTLAY SLTLSGVDHEHVVGTHVNFLITPPSGDPADLAGLGEQDLARLQLLAEFGAEGSGMKIQSTRPHTIGY TTEMGRIKLEGLRAIHVNLLQTNLSGEPGELETDEQIAIDQCLRFASDGSFYHHLQTTRPQTIGY GRLLG-VGFDACKAVHLNLCAMRAPPEGPSIESDEQIAIDQCLRFASDGSFYHHLQTTRPQTIGY GRLLG-VGFDACKAVHLNLCAMRAPPEGPSIES	249 248 247 252 264 312 311
NcsF2 NcsF1 Bfueh1 AnEH Human JHEH	SLELGLADPEHVAGVHLMLVIFPPQDAPEAIGRLGESDLGKLANSGEFGAIGIGWQKIQATKPQTLAY SLILSGVDHEHVVGTHVNFLITPPSGDPADLAGLGEQDLARLQLLAEFGAEGSGYMKIQSTRPDTLSY SLLLGAIDPSHLAGIHVNLLQTNLSGEPGELETLSDADKARLAVSERFLDDLSGPMKNQSTRPHTIGY TTEMGRIKLEGLRAIHVNLPFVAPAQFPEVPTPDEQIAIDQCLRFASDGSFYHHLQTTRPQTIGY GRLLG-VGFDACKAVHLNLCAMRAPPEGPSIESLSAAEKEGIARMEKFMTDGLAYAMEHSTRPSTIGH CTNMAQLVPSHVKGLHLNMALVLSNFSTLTLLLGQRFGRFLGLTERDVELLYPVKEKVFYSLMRESGYMHIQCTKPDTVGS ATHMATLFPDQVLGLHTNMPLSSRPLSTVKLFIGALFPSLI-VDAKYMDRIYPLKN-LFSYILRETGYFHIQATKPDTIGV α3 β6 α4	249 248 247 252 264 312 311
NcsF2 NcsF1 Bfueh1 AnEH Human JHEH	SLELGLADPENVAGVNLNMLVTFPPQDAPEAIGRLGESDLGKLANSGEFGAIGIGWQHQATKPQTLAY SLTLSGVDHEHVVGGTHVNFLITPPSGDPADLAGLGEQDLARLQLLAEFGAEGSGYMKIQSTRPHTIGY TTEMGRIKLEGLRAIHVNLLQTNLSGEPGELETLSDADKARLAVSERFLDDLSGPMKMQSTRPHTIGY GRLLG-VGFDACKAVHLNLCAMRAPPEGPSIESDEQIAIDQCLRFASDGSFYHHLQTTRPQTIGY GRLLG-VGFDACKAVHLNLCAMRAPPEGPSIESLSAAEKEGIARMEKFMTDGLAYAMEHSTRPSTIGH CTNMAQLVPSHVKGLHLNMALVLSNFSTLTLLLGQRFGRFLGLTERDVELLYPVKEKVFYSLMRESGYMHIQCTKPDTVGS ATHMATLFPDQVLGLHTNMPLSSRPLSTVKLFIGALFPSLI-VDAKYMDRIYPLKN-LFSYILRETGYFHIQATKPDTIGV $\alpha 3$ $\beta 6$ $\alpha 4$	249 248 247 252 264 312 311
NcsF2 NcsF1 Bfueh1 AnEH Human JHEH SgcF	$SLELGLADPEHVAGVHLMLVIF PPQDAPEAIGRLGEODLARLOLLAEFGAIGIGWQIQATKPQTLAYSLILGAIDPSHLAGIHVNELITPPSGDPADLAGLGEODLARLOLLAEFGAEGSGYMKIQSTRPTLGYTIEMGRIKLEGLRAIHVNLQTNLSGEPGELETLSDADKARLAVSERFLDDLSGPMKMQSTRPHTIGYGRLLG-VGFDACKAVHLNLCAMRAPPEGPSIESDEQIAIDQCLRFASDGSFYHHLQTTRPQTIGYGRLLG-VGFDACKAVHLNLCAMRAPPEGPSIESLSAAEKEGIARMEKFMTDGLAYAMEHSTRPSTIGHCTNMAQLVPSHVKGLHLNMALVLSNFSTLTLLLGQRFGRFLGLTERDVELLYPVKEKVFYSLMRESGYMHIQCTKPDTVGSATHMATLFPDQVLGLHTNMPLSSRPLSTVKLFIGALFPSLI-VDAKYMDRIYPLKN-LFSYILRETGYFHIQATKPDTIGV\frac{\alpha 3}{\beta 6}$	249 248 247 252 264 312 311 321
NcsF2 NcsF1 Bfueh1 AnEH Human JHEH SgcF NcsF2	SLELGLADPENVAGVNLNMLVTFPPQDAPEAIGRLGEQDLARLOLLAEFADIGIGWQLQATKPQTLAI SLTLSGVDHEHVVAGTNNKLITFPSGDPADLAGLGEQDLARLOLLAEFADIGIGWQLQATKPQTLSY SLLLGAIDPSHLAGIHVNLLQTNLSGEPGELETLGEQDLARLOLLAEFADIGIGWQLQATKPQTLSY TTEMGRIKLEGLRAIHVNLPFVAPAQFPEVPTPDEQIAIDQCLRFASDGSFYHHLQTTRPQTIGY GRLLG-VGFDACKAVHLNLCAMRAPPEGPSIESDEQIAIDQCLRFASDGSFYHHLQTTRPQTIGY GRLLG-VGFDACKAVHLNLCAMRAPPEGPSIES	249 248 247 252 264 312 311 321 321
NcsF2 NcsF1 Bfueh1 AnEH Human JHEH SgcF NcsF2 NcsF1	$SLELGLADPENVAGVHLNMLVTFPPQDAPEAIGRLGEQDLARLOLLAEFGAIGIGWQXIQATKPQTLAYSLTLSGVDHEHVVGTHVNFLITPPSGDPADLAGLGEQDLARLQLLAEFGAEGSGYMKIQSTRPHTIGYTTEMGRIKLEGLRAIHVNLLQTNLSGEPGELETLSDADKARLAVSERFLDDLSGPMKNQSTRPHTIGYGRLLG-VGFDACKAVHLNLCAMRAPPEGPSIESDEQIAIDQCLRFASDGSFYHHLQTTRPQTIGYGRLLG-VGFDACKAVHLNLCAMRAPPEGPSIESLSAAEKEGIARMEKFMTDGLAYAMEHSTRPSTIGHCTNMAQLVPSHVKGLHLNMALVLSNFSTLTLLLGQRFGRFLGLTERDVELLYPVKEKVFYSLMRESGYMHIQCTKPDTIGVATHMATLFPDQVLGLHTNMPLSSRPLSTVKLFIGALFPSLI-VDAKYMDRIYPLKN-LFSYILRETGYFHIQATKPDTIGV\frac{\alpha 3}{\beta 6}$	249 248 247 252 264 312 311 321 321 321 321
NcsF2 NcsF1 Bfueh1 AnEH Human JHEH SgcF NcsF2 NcsF1 Bfueh1	$SLELGLADPENVAGVNLNMLVIF PPQDAPEAIGRLGESDLGKLANSGEFANIGIGWQKIQATKPQTLAYSLTLSGVDHEHVVGGTHVNFLITPPSGDPADLAGLGEQDLARLQLLAEFGAEGSGYMKIQSTRPDTLSYSLLLGAIDPSHLAGIHVNLLQTNLSGEPGELETLSDADKARLAVSERFLDDLSGPMKMQSTRPHTIGYTTEMGRIKLEGLRAIHVNLPFVAPAQFPEVPTPDEQIAIDQCLRFASDGSFYHHLQTTRPQTIGYGRLLG-VGFDACKAVHLNLCAMRAPPEGPSIESLSAAEKEGIARMEKFMTDGLAYAMEHSTRPSTIGHCTNMAQLVPSHVKGLHLNMALVLSNFSTLTLLLGQRFGRFLGLTERDVELLYPVKEKVFYSLMRESGYMHIQCTKPDTVGSATHMATLFPDQVLGLHTNMPLSSRPLSTVKLFIGALFPSLI-VDAKYMDRIYPLKN-LFSYILRETGYFHIQATKPDTIGV\boxed{\alpha3} \qquad \qquad$	249 248 247 252 264 312 311 321 321 321 319 318
NCSF2 NCSF1 Bfueh1 AnEH JHEH SgCF NCSF2 NCSF1 Bfueh1 AnEH	$\label{eq:scalar} SLELGLADPENVAGVNLNMLVTFPPQDAPEAIGRLGEQDLARLOLLAEFGAEGSGYMKIQATKPQTLAY SLTLSGVDHEHVVGGTHVNFLITPPSGDPADLAGLGEQDLARLQLLAEFGAEGSGYMKIQSTRPHTIGY TTEMGRIKLEGLRAIHVNLLQTNLSGEPGELETLSDADKARLAVSERFLDDLSGPMKMQSTRPHTIGY GRLLG-VGFDACKAVHLNLCAMRAPPEGPSIESDEQIAIDQCLRFASDGSFYHHLQTTRPQTIGY GRLLG-VGFDACKAVHLNLCAMRAPPEGPSIESLSAAEKEGIARMEKFMTDGLAYAMEHSTRPSTIGH CTNMAQLVPSHVKGLHLNMALVLSNFSTLTLLLGQRFGRFLGLTERDVELLYPVKEKVFYSLMRESGYMHIQCTKPDTVGS ATHMATLFPDQVLGLHTNMPLSSRPLSTVKLFIGALFPSLI-VDAKYMDRIYPLKN-LFSYILRETGYFHIQATKPDTIGV \frac{\alpha 3}{\beta 6}$	249 248 247 252 264 312 311 321 321 319 318 333
NCSF2 NCSF1 Bfueh1 Human JHEH SgCF NCSF2 NCSF1 Bfueh1 AnEH Human	SLELGLADPENVAGVNLNMLVTFPPQDAPEAIGRLGESDLGKLANSGEFADIGIGWQKLQATKPQTLAY SLTLSGVDHEHVVAGTNUNLVGTIVPFSDPADLAGLGEQDLARLQULAEFGAEGGYMKIQSTRPHTIGY TTEMGRIKLEGLRAIHVNLLQTNLSGEPGELETLSDADKARLAVSERFLDDLSGPMKNQSTRPHTIGY GRLLG-VGFDACKAVHLNLCAMRAPPEGPSIESLSAAEKEGIARMEKFMTDGLAYAMEHSTRPSTIGH CTNMAQLVPSHVKGLHLNMALVLSNFSTLTLLLGQRFGRFLGLTERDVELLYPVKEKVFYSLMRESGYMHIQCTKPDTVGS ATHMATLFPDQVLGLHTNMPLSSRPLSTVKLFIGALFPSLI-VDAKYMDRIYPLKN-LFSYILRETGYFHIQATKPDTIGV α3 β6 *** ::*::: .::*::: ::*::: .::*::: ::*::: .::*::: ::*::: .::*::: ::*::: .::*::: ::*::: .::*::: .::*::: .::*::: .::*::: .::*::: .::*::: .::*::: .::*::: .::*::: .::*::: .::*::: .::*::: .::*::: .::*::: .::*::: .::*::: .::*::: .::*::: .::*:::: .::*::: .::*::::::::::::::::::::::::::::::::::	249 248 247 252 264 312 311 321 321 319 318 333 389
NCSF2 NCSF1 Bfueh1 Human JHEH SgCF NCSF2 NCSF1 Bfueh1 AnEH Human JHEH	$\label{eq:scalar} Sledsdefined with the product of the product o$	249 248 247 252 264 312 311 321 321 319 318 3389 386
NcsF2 NcsF1 Bfueh1 AnEH Human JHEH SgcF NcsF2 NcsF1 Bfueh1 AnEH Human JHEH	SLELGLADPENVAGVNLNMLVTFPPQDAPEAIGRLGESDLGKLANSGEFGAIGIGWQKIQATKPQTLAY SLTLSGVDHEHVVAGTHVNFLITPPSGDPADLAGLGESDLGKLANSGEFGAEGSGYMKIQSTRPDTLSY SLLLGAIDPSHLAGIHVNLLQTNLSGEPGELETLSDADKARLAVSERFLDDLSGPMKNQSTRPHTIGY TTEMGRIKLEGLRAIHVNLPFVAPAQFPEVPTPDEQIAIDQCLRFASDGSFYHHLQTTRPQTIGY GRLLG-VGFDACKAVHLNLCAMRAPPEGPSIESLSAAEKEGIARMEKFMTDGLAYAMEHSTRPSTIGH CTNMAQLVPSHVKGLHLNMALVLSNFSTLTLLLGQRFGRFLGLTERDVELLYPVKEKVFYSLMRESGYMHIQCTKPDTVGS ATHMATLFPDQVLGLHTNMPLSSRPLSTVKLFIGALFPSLI-VDAKYMDRIYPLKN-LFSYILRETGYFHIQATKPDTIGV α3 β6 ****ii: .:i*::**i* GLTDSPVGQLAWILDKFQEWSGGKNVEEAISRDRLLTHVMIYWLTATAGSSAQLYYESARGMADFARTWGGP SLTDSPVGQLAWILDKFQEWSGG	249 248 247 252 264 312 311 321 321 321 319 318 333 389 386
NCSF2 NCSF1 Bfueh1 AnEH JHEH SgCF NCSF2 NCSF1 Bfueh1 AnEH Human JHEH	$ SLELGLADPENVAGVNLNMLVTFPPQDAPEAIGRLGEQDLARLQLLAEFGAEGSGYMKIQATKPQTLAYSLTLSGVDHEHVVAGTHVNFLITPPSGDPADLAGLGEQDLARLQLLAEFGAEGSGYMKIQATKPQTLSYSLLLGAIDPSHLAGIHVNLLQTNLSGEPGELETLSDADKARLAVSERFLDDLSGPMKMQSTRPHTIGYTTEMGRIKLEGLRAIHVNLPFVAPAQFPEVPTPDEQIAIDQCLRFASDGSFYHHLQTTRPQTIGYGRLLG-VGFDACKAVHLNLCAMRAPPEGPSIESLSAAEKEGIARMEKFMTDGLAYAMEHSTRPSTIGHCTNMAQLVPSHVKGLHLNMALVLSNFSTLTLLLGQRFGRFLGLTERDVELLYPVKEKVFYSLMRESGYMHIQCTKPDTVGSATHMATLFPDQVLGLHTNMPLSSRPLSTVKLFIGALFPSLIVDAKYMDIYPLKN-LFSYILRETGYFHIQATKPDTIGV \hline \alpha 3 \qquad \beta 6 \qquad \qquad$	249 248 247 252 312 311 321 321 321 319 318 333 389 386
NCSF2 NCSF1 Bfueh1 Human JHEH SgCF NCSF2 NCSF1 Bfueh1 AnEH Human JHEH	SLELGLADPENVAGVNLNNLVTFPPQDAPEAIGRLGESDLGKLANSGEFADIGIGWQKLQATKPQTLAI SLTLSGVDHEHVVQGTHVNFLITPPSQDPADLAG	249 248 247 252 264 312 311 321 319 318 333 389 386
NCSF2 NCSF1 Bfueh1 Human JHEH SgCF NCSF2 NCSF1 Bfueh1 AnEH Human JHEH	SLELGLADPENVAGVNLNNLVTPPQDAPEAIGKLGEQDLARLQLLAEFGAEGSGYMKIQATKPQTLAY SLLLGAIDPSHLAGIHVNLLQTNLSGEPGELETLGEQDLARLQLLAEFGAEGSGYMKIQSTRPHTIGY TTEMGRIKLEGLRAIHVNLPFVAPAQFPEVPTPDEQIAIDQCLAEFASDGSFYHHLQTTRPQTIGY GRLLG-VGFDACKAVHLNLCAMRAPPEGPSIES	249 248 247 252 264 312 311 321 321 319 318 333 389 386
NCSF2 NCSF1 Bfueh1 AnEH Human JHEH SgCF NCSF2 NCSF1 Bfueh1 AnEH Human JHEH SgCF NCSF2	SLELGLADPENVAGVNLNNLVIFPPQDAPEAIGRLGEQDLARLQLLAEFGAEGSGYMKIQATRPQTLAY SLTLSGVDHEHVVAGTHVNFLITPPSQDPADLAGLGEQDLARLQLLAEFGAEGSGYMKIQATRPQTLSY SLLLGAIDPSHLAGIHVNLLQTNLSGEPGELETLSDADKARLAVSERFLDDLSGPMKNQSTRPHTIGY TTEMGRIKLEGLRAIHVNLPFVAPAQFPEVPTPDEQIAIDQCLRFASDGSFYHHLQTTRPQTIGY GRLLG-VGFDACKAVHLNLCAMRAPPEGPSIESLSAAEKEGIARMEKFMTDGLAYAMEHSTRPSTIGH CTNMAQLVPSHVKGLHLNMALVLSNFSTLTLLLGQRFGRFLGLTERDVELLYPVKEKVFYSLMRESGYMHIQCTKPDTVGS ATHMATLFPDQVLGLHTNMPLSSRPLSTVKLFIGALFPSLIVDAKYMDRIYPLKN-LFSYILRETGYFHIQATKPDTIGV α3 β6 * .** .::* GLTDSPVGQLAWILDKFQEWSGG	249 248 247 252 264 312 311 321 319 318 333 389 386 85 87
NCSF2 NCSF1 Bfueh1 AnEH JHEH SgCF NCSF2 NCSF1 Bfueh1 AnEH Human JHEH SgCF NCSF2 NCSF2 NCSF1	SLELGLADPENVAGVNLNALVIPPPQDAPEAIGRLGEQDLARLQLLAEFGAEGSGYMKIQATKPQTLAY SLILSGVDHEHVVAGTHVNFLITPPSQDPADLAGLGEQDLARLQLLAEFGAEGSGYMKIQATKPQTLSY SLLLGAIDPSHLAGIHVNLLQTNLSGEPGELETLSDADKARLAVSERFLDDLSGPMKNQSTRPHTIGY TTEMGRIKLEGLRAIHVNLPFVAPAQFPEVPTPDEQIAIDQCLRFASDGSFYHHLQTTRPQTIGY GRLLG-VGFDACKAVHLNLCAMRAPPEGPSIESLSAAEKEGIARMEKFMTDGLAYAMEHSTRPSTIGH CTNMAQLVPSHVKGLHLNMALVLSNFSTLTLLLGQRFGRFLGLTERDVELLYPVKEKVFYSLMRESGYMHIQCTKPDTVGS ATHMATLFPDQVLGLHTNMPLSSRPLSTVKLFIGALFPSLI-VDAKYMDRIYPLKN-LFSYILRETGYFHIQATKPDTIGV α3 β6 *.**i:: .::*::*i:* GLTDSPVGQLAWILDKFQEWSGGKNVEEAISRDRLLTHVMIYWLTATAGSSAQLYYESARGMADFARTWGGP SLTDSPVGQLAWILDKFQEWSGG	249 248 247 252 264 312 311 321 321 318 333 389 386 85 87 85
NCSF2 NCSF1 Bfueh1 Human JHEH SgCF NCSF2 NCSF1 Bfueh1 SgCF NCSF2 NCSF1 Bfueh1	SLELGLADPENVAGVNLNNLVTPPPQDAPEAIGKLGESDLGKLANSGEFADIGIGWQKLQATKPQTLAY SLTLSGVDHEHVVGGTNVNFLITPPSQDPADLAG	249 248 247 252 264 312 311 321 319 318 333 389 386 85 85 96
NCSF2 NCSF1 Bfueh1 Human JHEH SgCF NCSF2 NCSF1 Bfueh1 Human JHEH SgCF NCSF2 NCSF1 Bfueh1 Bfueh1	SLELGLADPENVAGVILAMLVTF PPQDAPEATGKLGSDLGRLARSGEFADTGGTGWKTQATKPQTLAT SLTLSGVDHEHVVGTHUNFLITPPSQDPADLAGLGSQDLARLQLAEFGAEGSGYMKTQSTRPTISY SLLLGAIDPSHLAGIHVNLLQTNLSGEPGELETLSDADKARLAVSERFLDDLSGPMKMQSTRPHTIGY TTEMGRIKLEGLRAIHVNLPFVAPAQFFEVPTPDEQIATDQCLRFADGGSYMKTQSTRPTIGY GRLLG-VGFDACKAVHLNLCAMRAPPEGPSIESLSAAEKEGIARMEKFMTDGLAYAMEHSTRPSTIGH CTNNAQLVPSHVKGLHLNMALVLSNFSTITLLLGQRFGRFLGLTERDVELLYPVKEKVFYSLMRESGYMHIQCTKPDTVGS ATHMATLFPDQVLGLHTNMPLSSRPLSTVKLFIGALFPSLI-VDAKYMDRIYPLKN-LFSYILRETGYFHIQATKPDTIGV $\alpha_3 \qquad \beta6 \qquad \qquad \alpha4$ * .**: : : : : : : : : : : : : : : :	249 248 247 252 264 312 311 321 321 319 318 333 389 386 85 87 85 98
NCSF2 NCSF1 Bfueh1 AnEH Human JHEH SgCF NCSF2 NCSF1 Bfueh1 AnEH SgCF1 Bfueh1 AnEH Human	SLELGLADDENVAGVHLANLVIFYPPDDAFERIGKLGEQDLARLQLLASGEFADIGIGWGKIQATKYQILAI SLILGAIDPSHLAGIHVNLLQTNLSGEPGELETLGEQDLARLQLLASGEFADGGSGYKKIQSTRPHTIGY TTEMGRIKLEGLRAIHVNLPFVAPAQFPEVPTPDEQIAIDQCLRFASDGSFYHHLQTTRPQTIGY GRLLG-VGFDACKAVHLNLCAMRAPPEGPSIESDEQIAIDQCLRFASDGSFYHHLQTTRPQTIGY GRLLG-VGFDACKAVHLNLCAMRAPPEGPSIES	249 248 247 252 264 312 311 321 318 333 389 386 85 85 98 55
NCSF2 NCSF1 Bfueh1 AnEH Human JHEH SgCF NCSF2 NCSF1 Bfueh1 AnEH Human JHEH SgCF NCSF1 Bfueh1 AnEH Human	SLELGLAPDEHVAGVHLANLVIFYPPDDAFEATGRLGEVDLARLQLLARSEFADIGIGWQKIQATKPQTLAY SLILGAVDHEHVVCTHVNFLITYPSGDPADLAG SLLLGAIDPSHLAGIHVNLLQTNLSGEPGELETLSAAKARLAVSERFLDDLSGPMKMQSTRPHTIGY TTEMGRIKLEGLRAIHVNLPFVAPAQFPEVPTPDEQIAIDQCLRFASDGSFYHHLQTTRPQTIGY GRLLG-VGFDACKAVHLNLCAMRAPPEGPSIES	249 248 247 252 264 312 311 321 319 318 333 389 386 85 87 85 96 98 55 62
NCSF2 NCSF1 Bfueh1 AnEH JHEH SgCF NCSF2 NCSF1 Bfueh1 AnEH SgCF NCSF2 NCSF1 Bfueh1 Bfueh1 AnEH Human JHEH	SLELGLAPPERVAGVHLMMLVTFPFQDAPELATGR LGEQDLARLOLLAE FADTGTGMQMTQATKPQTLAT SLTLSGVDHEHVVQGTHVNFLITPPGOPADLAG LGEQDLARLQLLAE FGEGSGYMKIQSTRPHTIGY SLLLGAIDPSHLAGIHVNLLQTNLSGEPGELET LGEQDLARLQLLAE FGEGSGYMKIQSTRPHTIGY TTEMGRIKLEGLRAIHVNLPTAPAOFPEVPTP DEQIAIDQCLR FASDGSFYHHLQTTRPQTIGY GRLLG-VGFDACKAVHLNLCAMRAPPEGPSIES LSAAEKEGIARMEK FASDGSFYHHLQTTRPQTIGY GRLG-VGFDACKAVHLNLCAMRAPPEGPSIES LSAAEKEGIARMEK FYSILRETGYFHIQATKPDTIGY GA *.**	249 248 247 252 264 312 311 321 319 318 333 386 85 85 85 96 98 55 62
NCSF2 NCSF1 Bfueh1 Human JHEH SgCF NCSF2 NCSF1 Bfueh1 AnEH SgCF NCSF2 NCSF1 Bfueh1 Bfueh1 AnEH Human JHEH	SLELGLADPENVAGUNENTLIPPSGDAPLAG	249 248 247 252 264 312 311 321 319 318 333 389 386 85 87 85 96 98 55 62

References

- (1) Liu, W.; Nonaka, K.; Nie, L. P.; Zhang, J.; Christenson, S. D.; Bae, J.; Van Lanen, S. G.; Zazopoulos, E.; Farnet, C. M.; Yang, C. F.; Shen, B. *Chem. Biol.* **2005**, *12*, 293-302.
- (2) Lin, S.; Horsman, G. P.; Chen, Y.; Li, W.; Shen, B. J. Am. Chem. Soc. 2009, 131, 16410-16417.
- (3) Doderer, K.; Lutz-Wahl, S.; Hauer, B.; Schmid, R. D. Anal. Biochem. 2003, 321, 131-134.
- (4) Mateo, C.; Archelas, A.; Furstoss, R. Anal. Biochem. 2003, 314, 135-141.