Modular domain swapping among the bacterial cytotoxic necrotizing factor (CNF) family for efficient cargo delivery into mammalian cells

## Elizabeth E. Haywood, Mengfei Ho, and Brenda A. Wilson\*

From the Department of Microbiology, School of Molecular and Cellular Biology, University of Illinois Urbana-Champaign, Illinois 61801

Running title: CNF Modular Domain Swapping

\*To whom correspondence should be addressed: Dr. Brenda A. Wilson, Department of Microbiology, University of Illinois Urbana-Champaign, 601 S. Goodwin Ave Urbana, IL 61801, telephone: (217) 244-9631; fax: (217) 244-6697; E-mail: <u>bawilson@life.illinois.edu</u>

**Keywords**: bacterial toxin, drug delivery system, fusion protein, protein chimera, protein engineering, molecular evolution, small GTPase, protein deamidation, structure-function, protein translocation

**Supplementary Information** 





**Supplementary Figure S1A. Molecular phylogenetic analysis by Maximum Likelihood method.** The evolutionary history was inferred by using the Maximum Likelihood method based on the JTT matrixbased model (1). The tree with the highest log likelihood (-10668.36) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches (Blue). Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site (below the branches). The analysis involved 9 amino acid sequences. All positions containing gaps and missing data were eliminated. There was a total of 1004 positions in the final dataset. Evolutionary analyses were conducted in MEGA7 (2). NCB Accession # WP045110427 CNFm, NCB Accession # WP075533205 CNFm2, NCB Accession # WP05306733 CNFp, NCB Accession # WP079952502 CNFs, NCB Accession # WP024231387 CNF3, NCB Accession # WP012304286 CNFy, NCB Accession # WP059330985 CNFx.

CNFx CNFp CNFm CNFm2 CNFy CNFs CNF3 CNF1 CNF2	1 MT E KWGQQYYL E FNEL I SKFPSPEKTI SNY I KNK FNSEL TWFNWVDPDKLYFVQ FTQNRSNNK 1 MNDQWQK KYL FEYNEL VSKFPSPEKI I SDY I KDK FNTDL HWFHWADPDNLYFVQ FSQSRSNNK 1 MNDQWQQKYLYEYNEL I SQFPSPEKI I SDY I KDRFNTDL HWFNWADPDNLYFI QFSQSRSNHR 1 MNDQWQQKYLYEYNEL I SQFPSPEKI I SDY I KDRFNTDL HWFNWADPDNLYFI QFSQSRSNHR 1 MKNQWQHQYFLSYSEL VANFPSPEKVVSDY I KHKFSTTL PWFGWADPDNLYFI QFSQSRSNHR 1 MNTQWQQKYLLEYNEL VSKFPSPEKVYSDY I KHKFSTTL PWFGWADPDNLYFI QFSQSRSNHR 1 MNTQWQQKYLLEYNEL VSKFPSPEKVYSDY I KHKFSTTL PWFSRI DPDNTYFI QFSQSRSNSK 1 MNTQWQQKYLLEYNEL VSKFPSPEKVYSDY I KHKFKTDL PWFSRI DPDNTYFI QFSQNRSNSR 1 MGNQWQQKYLLEYNEL VSNFPSPERVYSDY I KHKFKTDL PWFSRI DPDNAYFI CFSQNRSNSR 1 MNVQWQQKYLLEYNEL VSNFPSPERVYSDY I KNCFKTDL PWFSRI DPDNAYFI CFSQNRSNSR	63 63 63 63 63 63 63 63
CNFx CNFp CNFm2 CNFy CNFs CNF3 CNF1 CNF2	64 TYTGWDHVGKFSIRSMTLTQAAIINIGRYFEVPDEANAVAGIYKVPKAEMFDEKNEAQMLPSE 64 SYTGWDHLEKYKTNAMTLTQAAMLNMGHRFSTYDDANSVAGIYKTSNANLFDEKNEAKMLPSE 64 SYTGWDHLDEHKTNVMTLTQAAMLNISNRFSTYDDANAVAGIYRTSSATLFDEKNEAKMLPSE 64 SYTGWDHLDEYKTNVMTLTQAAMLNISNRFSTYDDANAVAGIYRTSSATLFDEKNEAKMLPSE 64 SYTGWDHLGKYKTNVMTLTQAAMLNISNRFSTYDDANAVAGIYRTSSATLFDEKNEAKMLPSE 64 SYTGWDHLGKYKTDTLTLTQAAIVNIGSRFDIFDEANSTAGIYKTNNADSFDETNEAKMLPSE 64 SYTGWDHLGKYKTDTLTLTQAAIINIGFRFDVFDDANTTAGIYTTNNADLFNETNEAKILPSE 64 SYTGWDHLGKYKTDALTLTQAAIINIGYRFEVFDEANATAGIYTTNNADLFDETNEAKMLPSE 64 SYTGWDHLGKYKTDALTLTQAALINIGYRFEVFDEANATAGIYTTNNADLFDETNEAKMLPSE 64 SYTGWDHLGKYKTGVLTLTQAALINIGYRFDVFDDANSSTGIYKTSSADVFNEENEEKMLPSE 64 SYTGWDHLGKYKTGVLTLTQAALINIGYRFDVFDDANSSTGIYKTSSADMFNEKNEEKMLPSE	126 126 126 126 126 126 126 126 126
CNFx CNFp CNFm2 CNFy CNFs CNF3 CNF1 CNF2	127 Y I SFLNK CDFSAVYNDEL STFWNDNTSVFKSLLKKYYTSSILYLFKNEQI SRKEFDFAINAI - 127 Y LRFIYN CDFAGLYNKAL SDYWSKYYERFKLLLKNYYVSSILYLYKNNLVSKDEHDFAMAAL - 127 Y LSFIYD CDFAGLYNKAL SDYWSQYYER LKLLLKNYYVSSILYLHKNNLVSKEEHDFAMDAL - 127 Y LSFIYD CDFAGLYNKAL SDYWSQYYER LKLLLKNYYVSSILYLHKNNLVSKEEHDFAMDAL - 127 Y LYFLRD CDFSNLYNKAL SDYWSQYYER LKLLLKNYYVSSILYLHKNNLVSKEEHDFAMDAL - 127 Y LYFLRD CDFSNLYNKAL SDYWSQYYER LKLLLKNYYVSSILYLHKNNLVSKEEHDFAMDAL - 127 Y LYFLRD CDFSNLYNKAL SDYWSENYDKFSTLLQNYY I SSALYLYKDSAI SKDEYEFSIDAI F 127 Y LYFLKN CDFAGLYNKAL SDYWSENYDKFKLLLKNYY I SSSLYLYKNEV I NKDEYEFAMKSL - 127 Y LYFLKN CDFAGLYNKAL SDYWSKNHEKFKLLLKNYY I SSSLYLYKNEV I NKDEYEFAMKSL - 127 Y LYFLKN CDFAGLYNKAL SDYWSKNHEKFKLLLKNYY I SSALYLYKNEV I SKDEHEFTMKAL - 127 Y LYFLKN CDFAGLYNKAL SDYWSKYDKFKLLLKNYY I SSALYLYKNGELDEREYNFSMNAL - 127 Y LYFLKGCDFSGIYGRFLSDYWSKYDKFKLLLKNYY I SSALYLYKNGELDEREYNFSMNAL -	188 188 188 189 188 188 188 188
CNFx CNFp CNFm2 CNFy CNFs CNF3 CNF3 CNF1 CNF2	189 NKKNNIKIYFFDVYGYIASDMFVASNDEMTMLFITGAEYPVMFTKNIDELREKIKEIIS-YSG 189 NRENNISLFFDIYGYYSSDIFLAENKERVMLFIPGATNPFLFSENLNSLRGRLKEIIK-EQN 189 NRDKNISLFFLDVYGYYSSDIFWAENKEKVMLFIPGAKNPFLFSNNRNSLRGRLKELIKEEKD 189 NRDKNISLFFLDVYGYYSSDIFWAENKEKVMLFIPGAKNPFLFSNNRNSLRGRLKELIKEEKD 190 NKKNKILRYYFDVYGYYSSDIFWAENKEKVMLFIPGATNPFIFADNITDLRDKIKALIS-DKN 189 NRENNIVLSFFDIYGYYSSDIFVANNDNKTMLFIPGATNPFIFADNITDLRDKIKALIS-DKN 189 NRENNIVLSFFDIYGYYSSDIFVAKNDDRIMLFIPGATNPFLFAENITRLRTRLKELIKEEKD 189 NRENNIVLSFFDIYGYYSSDIFVAKNDDRIMLFIPGATNPFLFAENITRLRTRLKELIK-END 189 NRSDNISLFFDIYGYYSSDIFVAKNDKVMLFIPGAKKPFLFKKNIADLRLTLKELIK-DSD 189 NRSDNISLFFFDIYGYYSSDIFVAKNNKTMLFIPGAKKPFLFKKNIADLRLTLKELIK-END	250 251 251 251 250 250 250 250
CNFx CNFp CNFm2 CNFy CNFs CNF3 CNF3 CNF1 CNF2	251 NKIAFLOHFSLYD RODG VTYYG VESIFNKISG - EEFNNSYIMYKSOL I TNPDIFSEITSLVKO 251 NTSLLSTHFSLYD RODGTSYSG VNTAL NG I KENKNFDESYFFYSPKKI TEKNVFEAIAIL VKK 252 NASLFATHFSLFD RODGTSYSG VNTVL NG I KKDHGFNESYFFYSPKKI TERNIFEAMAIL VKK 252 NASLFATHFSLFD RODGTSYSG VNTVL NG I KKDHGFNESYFFYSPKKI TERNIFEAMAIL VKK 252 TRELFSKHFSLYD RODGTYYG VNSMLEOI VS - GVVDTNYIMYSNKNI RERNVFESMAFSTRE 251 NRELLSRHFSLYD RODGTTFYG I DSTLOKI VN - GSFDESYFLYSRKNI SERDVFDAIAL SVCK 251 NRELLSRHFSLYD CODGSTFYG VDSVLKEI VN - GNFNESYFMYTYK KFNERDVFDAIAL SVCK 251 KOQLLSOHFSLYSROG VSYAG VNSVLHAIENDGNFNESYFLYSNKTLSNKDVFDAIAL SVKK 251 NKQLLSOHFSLYSROG I TYAG VNSVLNAIENDG VFNESYFLYSNKRINNKDVFDAVAFSVKK	312 313 314 314 313 312 312 313 313
CNFx CNFp CNFm2 CNFy CNFs CNF3 CNF1 CNF2	313 RG I SDGD V I I KSNSESS RDYAL E I I QT LL SLTP VFD VI I PEIS I PLGLG VVASGLG I SFDQL I 314 RSFSDGD VL I TSNSEASKEDALNILQT I LSMAP I FD VI I PEVS VPASLG I LATS VGLSFDQL I 315 RSFSDGD TL I TSDAEALKEDALNMLQT I LSMAP I FD VVLPELSLPLSLG I LSTS VGLSFD KL I 315 RSFSDGD I LI KSDAEALKEDALNMLQT I LSMAP I FD VI LPEVS VPVSLG I LSTS VGLSFD KL I 314 RSFNDGD VI I KSNAE VQRDYALNVLQT I LSLSP I FD I VLPEVS VPVSLG I LSTS VGLSFD KL I 313 RSFSDGD T I I KSNSEAQRDYALT I I QT I VSMAP VFD VI LPEVS VPLSLG I ASSMG I SFDQL I 314 RSFSDGD T I I KSNSEAQRDYALT I I QT I VSMAP VFD VI LPEVS VPLSLG I ASSMG I SFDQL I 314 RSFSDGD T VI KSNSEAQRDYALT I LQT I LSMTP I FD I VVPEVS VPLSMG I I ASSMG I SFDQL I 314 RSFSDGD I VI KSNSEAQRDYALT I LQT I LSMTP I FD V VPEVS VPLGLG I I TSSMG I SFDQL I	375 376 377 377 376 375 375 376 376
CNFx CNFp CNFm2 CNFy CNFs CNF3 CNF3 CNF1 CNF2	376 NGDTYEERRAAIPGIATNSILLGMSFVIPYIINKSKDIYTVLT-LPSEHIPVSNOSAVLSLLK 377 NGDTFEERRSAIPGVVTNVLLLGLSFAIPFIISKASANKEILNEFVSNEDNILNATNTDDFLK 378 NGDTFEERRSAIPGVVTNSVLLGLSFAIPFIISKAVANKNLLGLSVSNKENVLNDKNIDDFLK 378 NSDTFEERRSAIPGVVTNALLLGLSFAIPFIISKAVANKNILGKSVSNEDNILNDKNIDDFLK 377 NGDTYEERRSAIPGVATNAVLLGISFAIPFIISKAVANKNILGKSVSNEDNILNDKNIDDFLK 376 NGDTYEERRSAIPGVATNAVLLGISFAIPFLISKASKNKVILSKTVSNEDTPLNETNIDKFLS 376 NGDTYEERRSAIPGVATNAVLLGISFAIPYLISKASKNKVILSKTVSNEDTPLNETNIDKFLS 377 NGDTYEERRSAIPGVATNAVLLGISFAIPYLISKASENKVILSKTVSNEDSILNETNIDNFLA 377 NGDTYEERRSAIPGVATNAVLLGISFAIPLLISKASINGENVILSKTVSNEDSILNETNIDNFLA 377 NGDTYEERRSAIPGLATNAVLLGLSFAIPLLISKAGINGEVLSSVINNEGRTLNETNIDIFLK	437 439 440 439 438 438 438 439 439

CNFx CNFp CNFm2 CNFm2 CNFy CNFs CNF3 CNF1 CNF2	438 KHHVSIDE I PSDGVLTIELSQFNF - VNIVKLNDED - EFVAIKGSSLSGVYYEVEPETGYEILS 440 EYSINKDDISSTSVLEINIKETEQSVNIVKLSDEDNKIIAVKGSALSGIYYEADIKTGYEIFS 441 GYSINKDEISSTSVLEINIKETEQSVNIVKLSDENNKIIAVKGSALSGIYYEADIKTGYEIFS 441 EYSINKDEISSTRVLEINIKETEQSVNIVKLSDENNKIIAVKGSALSGIYYEADIKTGYEIFS 440 KYNISESDIPENGSLVINLKNTNVPVRLVKLNDEEGEIVAIKGSTLSGIYYEVDTETGYEILS 439 ENGINKENIPESGILEVEIKNTELPVNLVKISDENNQIIAVRGSAQSGIYYEVDIETGYEILS 439 ENGINKEDIPANGILEVDIKKSGIPVNLVKISDEDNQIVAVRGSSQSGIYYEVDIETGYEILS 440 EYGIAEDSISSTNLLDVKLKSSGOHVNIVKLSDEDNQIVAVRGSSLSGIYYEVDIETGYEILS 440 EYGIAEDSISSTNLLDVKLKSSGOHVNIVKLSDEDNQIVAVKGSSLSGIYYEVDIETGYEILS	498 502 503 503 502 501 501 502 502
CNFx CNFp CNFm2 CNFy CNFy CNF3 CNF3 CNF1 CNF2	499 RR VYRTEFDKK I YWTR SGGL KGG L PYNFONLE I PVFIKDK SYAELGEPSELSFINDDSALLYP 503 RR VYRTEYNNE I VWVR SGGLNGG KPFDFTTLELP I FFEDOPYSKLASS - ELSFINDDSPLLYP 504 RR VYRTEYDNK I LWI RGGGLNGG KPFDFNTLELPTFFEDOPYSKLASSNELSFINDDSPLLYP 504 RR VYRTEYNNE I LWI RGGGLNGG KPFDFNTLELPTFFEDOPYSKLASSNELSFINDDSPLLYP 503 RR VFRTEYNEK I YWTRGGGL KGGO PFNFEGLDI PVFID KPYSELASSVELSFVNDDSPLLFP 502 RR VYRTEYNNK I FWTR SGLKGGO PFDFENLDI PTFFVD KPYSELASSOELSFINDDSPLLFP 503 RR VFRTEYNNE I FWI RNGGLKGGO PFDFENLDI PTFFVD KPYSELASSOELSFINDDSPLLFP 504 RR VYRTEYNNE I FWI RNGGLKGGO PFDFESLNI PVFFKDEPYSAVTGS - PLSFINDDSPLLFP 505 RR I YRTEYNNE I LWTRGGGL KGGO SFDFESLKLPIFFKDEPYSAVTGS - SLSFINDDSSLLYP	561 564 566 565 564 564 564 564
CNFx CNFp CNFm2 CNFy CNFs CNF3 CNF1 CNF2	562 K I DSR I P SPTPEYELRY FFT - KD I YKEQL VTLMKGTTEQE AWN I ANYKTAGG VNEKL PE I FPG 565 NVDSR L P A STSEMEMMH FYNDR PKF VEQDL I LMRGTTE EE AWN I ASTKTAGGSNEAL KD I S I E 567 DL DSR L PKPTSEMDMHN FIEDRT QFNEQDL I LMRGTTE EE AWN I ANNKTAGGSDGEL KD I S I D 567 DL DPR L PKPTSEMN I NNFI DDRS QFAE QDL I LMRGTTE ED AWN I ANNKTAGGSDGEL KD I S I E 566 EMDSR L PKPTPEL D I KYYSSNL SSFKEDT VI LMRGTTE EE AWN I ANYKTAGGSNK DLE EN FIE 565 DVDSR L PKPTPE I D I RNYSTHFSR FSENT VTLMRGTTE EE AWN I ANYKTAGGSNKELE EI FIG 565 DVDSR L PKPTSEMD I SYYSSNFSSFAENT VTLMRGATE EE AWN I ASYKTAGGSNKELE EI FIG 565 DTNPKL P QPTSEMD I VNYVKGSG SFGDRF VTLMRGATE EE AWN I ASYHTAGGSTE ELHE I LLG 565 NSTPKL P QPTPEME I VNYVKRAGNFGERL VTLMRGTTE EE AWN I ASYHTAGGSTE ELHE I LLG	623 627 629 629 628 627 627 627 627
CNFx CNFp CNFm2 CNFm2 CNFy CNFs CNF3 CNF1 CNF2	624 E GPOSRL GFTEYTTDINS ADSSSRGHFLVVIK VEVKY - INNDNVFYANYWK I PDD AP VEVVA I 628 GHPOTDVSTTAYTTDFKS ADAASRRHFLVIVK VKLKYVVGNNEVSHVNHWA I IDE AP VEVVA V 630 ANPOEGVSTTVYTTDYKS ADVVSRRHFLVVVKVKLKYVVSNNETSHANHWA I IDE AP VEVLAV 630 ANPOAGVSTTVYTTDYKS ADVASRRHFLVVVKVKLKYVVSNNETSHANHWA I IDE AP VEVLAV 629 AGPOFNLSFSEYTSSINS ADTASRKHFLVVIKVKLKYVVSNNETSHANHWA I PDE AP VEVLAV 628 GGPOANLSFTEYTSNVRS ADAASRGHFLVVVNVKIKY - I SNDNVLYANHWA I PDE AP VEVLAV 628 GGPOANLSFTEYTSNIRS ADAASRRHFLVVINVKIKY - I SNDNVLYANHWA I PDE AP VEVLAV 628 GGPOSLGFTEYTSNVRS ADAASRRHFLVVINVKIKY - I SNDNVLYANHWA I PDE AP VEVLAV 628 QGPOSLGFTEYTSNVRS ADAASRRHFLVVIK VVINVKIKY - I NNNVSYVNHWA I PDE AP VEVLAV	685 690 692 690 689 689 689 689 689
CNFx CNFp CNFm2 CNFw2 CNFs CNFs CNF3 CNF1 CNF2	686 ADRRFLFSEIASKPDMSFFKK - I LOKPAPESIIKKSNSNYKRLSIGDIDVLKGRGAFSSVROR 691 VDRRFLFPEPPSOSELSILOKMLRLRFFRKNIAENASINFOKLAKGDINVLKGRGSISSTROR 693 VDRRFSFPEPPSPLELPILOKFLRIFYKKNIAEKASINFNRLAKGDINVLKGRGNIASTROR 693 VDRRFLFPEPPSPLELSILOKFLROKFYKKNVAEKASINFNRLAKGDINVLKGRGNIASTROR 691 VDRRFIFPEPPVKPKLSFIQK - I ANRFLTENVAE ISSINFRRLNSGNINVLKGRGVFSSRLR 690 VDRRFIFPEPPAPPKLSLIOK - LSRRFFTEDIAETSRINFORLNSANINVLKGRGSLSSKNOR 690 VDRRFIFPEPPTPPKLSLIOK - I SORFFTEDIDETSRINFORLNSGNINVLKGRGSLSSKNOR 690 VDRRFIFPEPPTPPKLSLIOK - I SORFFTEDIDETSRINFORLNSGNINVLKGRGSLSSKNOR 690 VDRRFIFPEPPTPPKLSLIOK - LSRRFFTEDIDETSRINFORLNSGNINVLKGRGSLSSKNOR 690 VDRRFNFPEPSTPPDISTIRKLLSLRYFKESIESTSKSNFOKLSRGNIDVLKGRGSISSTROR 690 VDRRFNFPEPSTPPNISIIHKLLSLRYFKENIESTSRLNLOKLNRGNIDIFKGRGSISSTROR	747 753 755 755 752 751 751 752 752 752
CNFx CNFp CNFm2 CNFm2 CNFy CNFs CNF3 CNF1 CNF2	748 NIYISFQAANSDLLNKSGVYIRTLKLSDIGYDPVFYNYGLGISGSPTLNTYTGEIVTEETLQT 754 AIEPYFISSNADELQ-GGFYTKKDRFNNLGYDTHFYNSSIGSNGAPTLNTYTGEILTDSSSLG 756 TIDLNFISANADELR-TDFYIRKSPLNEAGYDRHFYNNTIGVNGFPTLNTYTGEIMADPSALG 756 TIDLNFISANADELR-TDFYIRKSPLNEAGYDRHFYNNTIGVNGSPTLNTYTGEIMADPSALG 753 EIYLRFDAANADELR-TDFYITKNPINEAGYDRHFYNNTIGVNGSPTLNTYTGEIMADPSVLG 752 SIYLRFDAANADELRPGDVYVKKTKFDSMGYDSHFYNEGIGINGAPTLNTYTGEILTDPSSFG 752 SIYLRFDAVNADALRPDEIYIKNNQFDDGGYDRYFYNNAVGLDGSPTLNTYTGEILTDPSSFG 753 AIYPYFEAANADEQQPLFFYIKKDRFDNHGYDQYFYDNTVGLNGIPTLNTYTGEILSDASSLG 753 AIYPYFESANADEQQPVFFYIKKNRFDDFGYDQYFYNSTVGLNGIPTLNTYTGEILSDASSLG	810 815 817 817 815 814 814 815 815
CNFx CNFp CNFm CNFm2 CNFy CNFs CNF3 CNF1 CNF2	811 TSYWKKYNLTTDTSIIHLSNSKQGANGIKINLDDLMKEKPVVITSGELSGCTSIWARKGNQFY 816 STYWKEHNLTNNTNIIRVSNSARGANGIKIELEQVQANKPVIITSGELSGCTTIVARKGGWLY 818 LTYWKENNLTNNAAIINISNSTRGANGIKIELEAVKVNKPVIITSGELSGCTTILARKGGYLY 818 LTYWKENNLTNNAGIINISNSTRGANGIKIELEAVKVNKPVIITSGELSGCTMILARKGGYLY 816 ATYWLKYNLTNETSIIKVSNSARGANGIKIALEEIEENKPVVITSGTLTGCTVVFARKGEYFY 815 SLYWSKYNLTNKTSIIRVTNSARGANGIRIALKEVQENKPVIITNGNLSGCTTIVARKGEYLY 816 STYWKKYNLTNETSIIRVSNSARGANGIRIALKEVQENKPVIITNGNLSGCTTIVARKGEYLY 816 STYWKKYNLTNETSIIRVSNSARGANGIRIALKEVQENKPVIITSGNLSGCTTIVARKGEYLY	873 878 880 878 877 877 877 878 878

CNFx	874	ΑV	нтα	) T	/EF	<mark>2</mark>	٢N	FΤ	SТ	ТС	) V	I K	ΑI	E١	/ L S	SS	L -	- S	G٧	/ N N	I A I	D	IQ	sv	SN	D	۲L '	۷N	FL	SE	NF	DT	SF	· V A	934
CNFp	879	κv	нтα	) T I	T E F	PV/	٩ <mark>G</mark>	FΤ	SТ	ΤG	€V <mark>(</mark>	ΩK	ΑV	E١	/F	EL	LL	ΚD	GΑ	GF	۶ <mark>Р</mark> ۴	۲V	KG	I M	ΝN	DF	۲ L ۱	۷N	ΥL	AD	NF	DE	S١	. I T	6 941
CNFm	881	QV	нтα	) T S	E I	LL	DG	FΤ	SТ	I G	) V	٢ĸ	ΑI	E١	/F	EL	L -	- E	DT	' T	P۴	۲V	EG	I M	ΝN	DF	۲L,	AN	ΥL	ΑK	NF	DE	S١	. I T	6 941
CNFm2	881	QV	нтα	) T S	E	LL	DG	FΤ	ST	RG	) V	٢ĸ	ΑI	E١	/F <mark>E</mark>	EL	L -	- R	DT	AK	۲ <mark>Р</mark>	۲V	EG	I M	ΝN	DF	EL,	AN	ΥL	ΑK	NF	DE	SL	. I <mark>T</mark> `	6 941
CNFy	879	ΑV	нтα	SN <mark>S</mark>	SE S	S L	I G	FΤ	SΤ	SG	SV/	٩ <mark>K</mark>	ΑI	E١	/ L S	SS	L -	- S	ΕL	. E \	/ <mark>P</mark> /	۱L	PD	۷I	ΝN	INT	۲L '	٧E	ΥL	SD	NF	DS	AL	. I <mark>S`</mark>	939
CNFs	878	ΑV	нтα	) T S	S E F	۲L	۷G	FΤ	SТ	ТС	) V	٢ĸ	ΑV	E١	/ L S	SΤ	L -	- A	GC	ΣE Ι	P٦	ΓL.	AD	ТΕ	ΝN	DF	۲ L ۱	۷ <mark>N</mark>	FL	AD	NF	Dĸ	SL	. I T	938
CNF3	878	ΕV	нтα	) T L	. E F	<sup>2</sup> L I	G	FΤ	SΤ	ТС	) V	٢ĸ	ΑV	E١	/ L S	SΤ	L -	- A	ЕC	ΣE Ι	P۵	SL.	AG	ТΕ	ΝN	DF	۲ L ۱	٧D	FL	AE	NF	Dĸ	SL	. V T `	938
CNF1	879	ĸν	нтα	) T I	٢K	SL/	٩ <mark>G</mark>	FΤ	SТ	ΤG	) V	٢ĸ	ΑV	E١	/L <mark>e</mark>	EL	L -	- T	ΚE	E <mark>P</mark> I	PF	۲V	EG	I M	SN	DF	۲ L ۱	٧D	YL	SE	NF	EC	SL	. I <mark>T</mark> `	6 939
CNF2	879	ΚV	HTG	) T I	C I F	L /	٩ <mark>G</mark>	FΤ	SΤ	ΤG	) V	٢K	ΑV	E١	/F	EL	L -	- T	N٨	1 <mark>5</mark> 1	/ <mark>P</mark> F	۲V	EG	VМ	ΝN	DF	۲L '	۷ <mark>N</mark>	ΥL	AE	SF	DE	SL	. I <mark>T</mark> `	6 939
																													_						
CNFx	935	SS	SEK	K <mark>K</mark> /	۱N	SΚ	ΙT	I N	HS	N١	/F	TΥ	AY	ΥT	DL	_ T	PV	PS	FG	B T S	SV/	۱L	LΤ	KG	DG	G	K	۷ <mark>K</mark>	AL	SE	ΤY	A A	KF	RDG	997
CNFp	942	AS	S D A	۱ <mark>۲</mark> ۵	QG	5 I -	I T	vs	RD	N١	/L	ΤF	ΡY	ΥT	'NE	DN	LL	- G	FG	B T S	S A 1		L A	RV	DN	٦N	۲ L '	٧S	SL	SE	SY	'S L	Τī	DG/	A 1003
CNFm	942	AS	SR/	N <mark>K</mark> F	P N S	S V	I T	vs	ΥN	N١	/s	ΤF	ΡY	ΥT	D	DG	MI	- G	FG	B T S	S A 1		L A	RV	DN	IK I	11	٧ <mark>K</mark>	SL	SE	SY	'S L	SF	GE	1003
CNFm2	942	AS	ss/	N <mark>K</mark> F	s	S V	ΙΤ	٧S	ΥN	N١	/s	ΤF	ΡY	ΥT	D	DG	LI	- G	FG	B T S	S A 1		L A	RV	DN	IK I	11	٧ <mark>K</mark>	SL	SE	SY	'S L	ΚA	GE	1003
CNFy	940	SS	SSL	. <mark>K</mark> F	P N S	SM	I N	I S	RE	N١	/s	ΤF	SY	ΥT	D	ן כ	QL	PS	FG	B T S	SV1		LV	RT	ND	N	۲V	۷R	SL	SE	SY	ΤN	INS	SNS	1002
CNFs	939	SS	SKL	. <mark>K</mark> F	P N S	5 I -	I T	VR	RD	N١	/s	ΤF	ΡY	ΥT	D	NC	IQ	ΡG	FG	B T S	SV1		LV	ĸ۷	DN	I <mark>N</mark> T	۲V	۷R	SL	SE	SY	ST	'NI	DG	1001
CNF3	939	SS	STL	. K F	D	S I	ΙT	I S	RD	N١	/s	ΤF	ΡY	ΥT	D	וכ	ΙH	ΡG	FG	B T S	SV1		LV	RI	DD	N	۲V	٧ <mark>K</mark>	SL	SE	SY	ΊVΤ	'N/		1001
CNF1	940	SS	SEK	( <mark>k</mark> f	D	Q	ΙT	11	RD	N١	/ <mark>S</mark>	۷F	ΡY	FL	DN.	11	ΡE	HG	FG	B T S	S A 1	V	LV	RV	DG	N\	/V'	۷R	SL	SE	SY	'S L	NA		1002
CNF2	940	SS	SEG	2 <mark>K</mark> I	G	SΚ	ΙT	I S	RD	N١	/s	ΤF	ΡY	FL	DN.	11	ΡE	KG	FG	ST S	S V I		LV	RV	DG	N۱	/ I '	۷ <mark>K</mark>	SL	SE	SY	'S L	N۱	EN <mark>S</mark>	1002
		_	_	_		_																													
CNFx	998	<u> </u>	P - F	DL	. L (	R	EL	L -																											1009
CNFp	1004	K۷	S I S	S <mark>K</mark> /	۱L <mark>s</mark>	SKI	۲	ΝV	EQ	LE	s	ΒV	ST	AG	2 <mark>P</mark> E	ΞS	GV	ST	ЕC	2															1037
CNFm	1004	ΚI	S I S	S <mark>K</mark> /	۱L <mark>s</mark>	SKI	EF	SA	SS																										1019
CNFm2	1004	KI	S I S	S <mark>K</mark> /	۱L <mark>د</mark>	SKI	EF	SA	SS																										1019
CNFy	1003	КM	VVF	۱ <u>N</u>	/L	۵ <mark>К</mark>	DF																												1014
CNFs	1002	ΚI	S V F	۲ <mark>۲</mark>	/L <mark>S</mark>	SKI	DF																												1013
CNF3	1002	RI	S V F	۶ <mark>К</mark> \	/L <mark>S</mark>	S K I	DF																												1013
CNF1	1003	ΕI	SVL	. <mark>K</mark> \	/F <mark>s</mark>	SK I	٢																												1014
CNF2	1003	NI	SVL	. H \	/F	SKI	DF																												1014

**Figure S1B: Amino acid alignment of CNF toxin homologs shown in Supplementary Figure S1A**. The alignment was generated using Muscle (39) and visualized using Jalview (40) to color the amino acid residue sequences in Clustal format.

CNFy	1 MKNOWOHOYFLSYSELVANFPSPEKVVSDYIKHKFSTTLPWFGWADPDNLYFIRFTQSRSNNKSYTGWDHLGK	73
CNF3	1 MNTOWQQKYLLEYNDLVSKFPSPEKVTSDYIKHKFKTDLPWFSRVDPDKTYFIQFSQNRSNSRSYTGWDHLGK	73
CNF1	1 MGNOWQQKYLLEYNELVSNFPSPERVVSDYIKNCFKTDLPWFSRIDPDNAYFICFSQNRSNSRSYTGWDHLGK	73
CNF2	1 MNVQWQQKYLLEYNELVSNFPSPERVVSDYIRRCFKTDLPWFSQVDPDNTYFIRFSQSRSNSRSYTGWDHLGK	73
CNFy	74 YA IETLTLTQAA I VNIGSREDIFDEANSTAGIYKTNNADSEDETNEAKMLPSEYLYFLRDCDESNLYNKALSD	146
CNF3	74 YKTDALTLTQAA I INIGYREEVEDEANATAGIYTTNNADLEDETNEAKMLPSEYLYFLKNCDEAGLYNKALSD	146
CNF1	74 YKTEVLTLTQAALINIGYREDVEDDANSSTGIYKTKSADVENEENEEKMLPSEYLHFLOKCDEAGVYGKTLSD	146
CNF2	74 YKTGVLTLTQAALINIGYHEDVEDDANASAGIYKTSSADMENEKNEEKMLPSEYLYFLKGCDESGIYGRELSD	146
CNFy	147 YWAENYEKFSTLLONYY I SSAYYLYKDSA I SKDEYEFS I DA I FNKKNK I LRYYFDVYGYYSSDMFVAMNDNKT	219
CNF3	147 YWSKNHEKFKLLLKNYY I SSSLYLYKNNV I SKDEHEFTMKAL - NRDDN I ELFSFD I YGYYSSD I FGAKNDDR I	218
CNF1	147 YWSKYYDKFKLLLKNYY I SSALYLYKNGELDEREYNFSMNAL - NRSDN I SLLFFD I YGYYASD I FVAKNNDKV	218
CNF2	147 YWSKYYDKFKLLLKNYY I SSALYLYKNGEI DEYEYNFSI SAL - NRDN I SLFFD I YGYYSSDMFVAKNNERV	218
CNFy	220 MLF I PGATNPF I FADN I TDLRDK I KAL I SDKNTRELFSKHFSLYDRODGNTYLGVNSMLEQ I VS - GVVDTNYI	291
CNF3	219 MLF I PGATNPFLFSEN I SHLRTHLKEL I KENDNRELLSRHFSLYDCODGSTFYGVDSVLKE I VN - GNFNESYF	290
CNF1	219 MLF I PGAKKPFLFKKN I ADLRLTLKEL I KDSDKOOLLSOHFSLYSRODGVSYAGVNSVLHA I ENDGNFNESYF	291
CNF2	219 MLF I PGAKKPFLFKKN I ADLRLTLKEL I KDNKOLLSOHFSLYSRODG I TYAGVNSVLNA I ENDGVFNESYF	291
CNFy	292 MYSNKNIRERNYFESMAFSTRERSFNDGDVIIKSNAEVORDYALNVLOTILSLSPIFDIVLPEVSIPISLGIT	364
CNF3	291 MYTYKKFNERDVFDAISFSVOKRSFSDGDTIIKSNSEAORDYALTIIOAIVSMIPVFDIILPEVSVPLSMGII	363
CNF1	292 LYSNKTLSNKDVFDAIAISVKKRSFSDGDIVIKSNSEAORDYALTILOTILSMTPIFDIVVPEVSVPLGLGII	364
CNF2	292 LYSNKRINNKDVFDAVAFSVKKRSFSDGDIVIKSNSEAORDYALTILOTILSMTPIFDVAIPEVSVTLGLGII	364
CNFy	365 ASSVG I SFDEL I NGDTYEERRSA I PGLATNAVLLG I SFA I PFL I SKAEENKLI I I NNLVOSDEN I LNKNNLGDF	437
CNF3	364 ASSMG I SFDOL I NGDTYEERRSA I PGVATNAVLLG I SFALPYL I SKASENKVI LSOTVSNEDS I LNETN I DNF	436
CNF1	365 TSSMG I SFDOL I NGDTYEERRSA I PGLATNAVLLGLSFA I PLL I SKAG I NOEVLSSVI NNEGRT LNETN I DI F	437
CNF2	365 ASSMG I SFDOL I NGDTYEERRSA I PGLATNAALLGLSFA I PFL I SKAG I NOEVLSSVI NNEGRT LNETN I DI F	437
CNFy	438 LEKYNISESDIPEN <mark>GS</mark> LVINLKNTNVPVRLVKLNDEEGEIVAIKGSTLSGIYYEVDTETGYEILSRRVFRTEY	510
CNF3	437 LAENGINKDDIPANGILEVDIKKSGIPVNLVKISDEDNQIVAVRGSSQSGIYYEVDIETGYEILSRRVYRTEY	509
CNF1	438 LKEYGIAEDSISSTNLLDVKLKSSGQHVNIVKLSDEDNQIVAVKGSSLSGIYYEVDIETGYEILSRRIYRTEY	510
CNF2	438 LEEYGINKNSISETKVLEVELKGSGQHVNIVKLSDEDNKIVAVKGNSLSGIYYEVDIETGYEISSRRIYRTEY	510
CNFy	511 NEK I YWTRGGGLKGGOPFNFEGLDI PVYFI DKPYSELASSVELSFVNDDSPLLFPEMDSRLPKPTPELDI KYY	583
CNF3	510 NNE I FWI RNGGLKGGOPFDFENLDI PTFFVDKPYSELASSPELSFI NDDSPLLFPYVDSRLPKPTSEMDI SYY	582
CNF1	511 NNE I LWTRGGGLKGGOPFDFESLNI PVFFKDEPYSAVTGSP-LSFI NDDSSLLYPDTNPKLPOPTSEMDI VNY	582
CNF2	511 NDKI FWTRGGGLKGGOSFDFESLKLPI FFKDEPYSAVPGSS-LSFI NDDSSLLYPNSTPKLPOPTPEMEI VNY	582
CNFy	584 SSNLSSFKEDTVILM <mark>R</mark> GTTEEEAWNIANYKTAGGSNKDLEENFIEAGPOFNLSFSEYTSSINSADTASRKHFL	656
CNF3	583 SSNFSSFAENTVTLMRGATEEEAWNIAYKTAGGSNKELEEIFIGGGPOANLSFTEYTSNIRSADAASRRHFL	655
CNF1	583 VK <mark>GSGSFGD</mark> RFVTLMRGATEEEAWNIASYHTAGGSTEELHEILLGOGPOSSLGFTEYTSNVNSADAASRRHFL	655
CNF2	583 VKRAGNFGERLVTLMRGTTEEEAWNIARYHTAGGSTEELHEILLGOGPOSSLGFTEYTSNINSADAASRRHFL	655
CNFy	657 VIIKVQVKYISNDNVLYANHWAIPDEAPVEVLAVVDRRFIFPEPPVKPKLSFIQK-IANRFLTENVAEISSIN	728
CNF3	656 VVINVKIKYISNDNVLYANHWAIPDEAPVEVLAVVDRRFIFPEPPTPPKLSLIQK-ISQRFFTEDIDETSRIN	727
CNF1	656 VVIKVHVKYITNNNVSYVNHWAIPDEAPVEVLAVVDRRFNFPEPSTPPDISTIRKLLSLRYFKESIESTSKSN	728
CNF2	656 VVIKVQVKYINNNNVSHVNHWAIPDEAPVEVLAVVDRRFNFPEPSTPPNISIIHKLLSLRYFKENIESTSRLN	728
CNFy	729 FR <mark>RLNSGN I NVLKGRGVFSSRRLRE I YLRFDAANADE LRPG</mark> DVYV <mark>KKTKFDSMGYDSHFYNEG I G I NGAPTLN</mark>	801
CNF3	728 F <mark>ORLNSGN I NVLKGRGSLSSKNOR</mark> S I YLRFDAVNADD LRPDE I YVKKDOFDD LGYDRYFYNAVGLDGSPTLN	800
CNF1	729 FOKLSRGN I DVLKGRGS I SSTRORA I YPYFEAANADEQOP LFFY I KKDRFDNHGYDQYFYDNTVGLNG I PTLN	801
CNF2	729 LOKLNRGN I D I FKGRGS I SSTRORA I YPYFESANADEQOP VFFY I KKNRFDDFGYDQYFYNSTVGLNG I PTLN	801
CNFy	802 TYTGEYVADSSSQGATYWL KYNLTNETSI I KVSNSARGANG I KIALEE I EENKPVVI TSGT LTGCTVVFARKG	874
CNF3	801 TYTGEFLTDPSLFGSLYWSKYNLTNKTSI I RVANSARGANG I RIALKEVQENKPVI I TNGNL SGCTT I VARKG	873
CNF1	802 TYTGE I PSDSSSLGSTYWKKYNLTNETSI I RVSNSARGANG I KIALEEVQEGKPVI I TSGNLSGCTT I VARKE	874
CNF2	802 TYTGE I LSDASSLGSTYWKKYNLTNETSI I RVSNSARGANG I KIALEEVQEGKPVI I TSGNLSGCTT I VARKE	874
CNFy	875 EYFYAVHTGNSESLIGFTSTSGVAKAIEVLSSLSELEVPALPDVINNNTLVEYLSDNFDSALISYSSSSLKPN	947
CNF3	874 EYLYEVHTGTLEPLLGFTSTTGVKKAVEVLSTLAEQEIPSLAGTINNDFLVDFLAENFDKSLVTYSSSTLKPD	946
CNF1	875 GYIYKVHTGTTKSLAGFTSTTGVKKAVEVLELLTKEPIPRVEGIMSNDFLVDYLSENFEDSLITYSSSEKKPD	947
CNF2	875 GYLYKVHTGTTIPLAGFTSTTGVKKAVEVFELLTNNPMPRVEGVMNNDFLVNYLAESFDESLITYSSSEQKIG	947
CNFy	948 SMINISRENVSTFSYYTDDIQLPSFGTSVTILV <mark>RT</mark> NDNTVVRSLSESYTMNSNSKMVVFNVLQKDF	1014
CNF3	947 SIITISRDNVSTFPYYTDDIIHPGFGTSVTILVRIDDNTVVKSLSESYVTNADGSRISVFKVLSKDF	1013
CNF1	948 SQITIIRDNVSVFPYFLDNIPEHGFGTSATVLVRVDGNVVVRSLSESYSLNADASEISVLKVFSKKF	1014
CNF2	948 SKITISRDNVSTFPYFLDNIPEKGFGTSVTILVRVDGNVIVKSLSESYSLNVENSNISVLHVFSKDF	1014

**Figure S1C: Amino acid alignment of CNF toxins used in this study**. The amino acid sequences of CNF1, CNF2, and CNF3 from *Escherichia coli* and CNFy from *Yersinia pseudotuberculosis* were aligned using Muscle (39) and visualized using Jalview (40) to color the amino acid residue sequences in Clustal format.



**Supplementary Figure S2. Simultaneous exchange of domains B and B<sub>2</sub> of CNFy with that of CNF1 leads to significantly decreased delivery efficiency of by CNFy domain T**. HEK293-T cells with reporter plasmids were treated with the indicated toxin at the indicated doses and subjected to SRE-luciferase assay to determine the normalized fold activation relative to CNF1 and untreated cells, as described in Figure 3. (A) Dose response curve comparing CNFy cargo delivered by CNFy translocation domain using the CNF1 receptor binding domains (CNF1 green closed triangles, CNFy red closed squares, CNF1y-223 dark blue open squares, CNF1y1y orange open squares). For comparison, the dose response curve for CNF1y1-688 (pale blue open squares) shown in Figure S9A is also included here. (B) Corresponding scatter plot of all data points used to derive the best-fit lines and mean values show in (A).



**Supplementary Figure S3. Corresponding scatter plots of Figure 2.** Corresponding scatter plots of all data points used to derive the best-fit lines and mean values show in Figure 2.



**Supplementary Figure S4. Corresponding scatter plots of Figure 3.** Corresponding scatter plot of all data points used to derive the best-fit lines and mean values show in main text Figure 3.



**Supplementary Information Figure S5. CNF3 cargo is delivered less efficiently by CNF1, CNF2 and CNFy delivery vehicles, regardless of joining site.** HEK293-T cells with reporter plasmids were treated with the indicated toxin at the indicated doses and subjected to SRE-luciferase assay to determine the normalized fold activation relative to CNF1 and untreated cells, as described in Figure 3. (A) Dose

response curves comparing CNF3 cargo delivered by CNF1, CNF2 and CNFy delivery vehicles, joined at amino acid 688 (CNF1 green closed triangles, CNF3 purple closed diamonds, CNF13-688 green open diamonds, CNF23-688 blue open diamonds, CNFy3-688 red open diamonds). (B) Dose response curve comparing CNF3 cargo delivered by CNFy delivery vehicle. Curves are the same as Figure 3D and Supplementary Figure S3A, shown here for comparison. (CNF3 purple closed diamonds, CNFy3-688 black open diamonds, CNFy3-735 red open squares). (C&D) Corresponding scatter plots of all data points used to derive the best-fit lines and mean values show in (A) & (B), respectively.



**Supplementary Figure S6. Corresponding scatter plots of main text Figure 4.** Corresponding scatter plot of all data points used to derive the best-fit lines and mean values show in main text Figure 4.



**Supplementary Figure S7. Corresponding scatter plot of main text Figure 5.** Corresponding scatter plot of all data points used to derive the best-fit lines and mean values show in main text Figure 5.



Supplementary Figure S8. SDS PAGE Gels of Purified CNF toxins. Purified CNF toxins were diluted to 0.1  $\mu$ g/ $\mu$ L in a solution of SDS and bromophenol blue dye, then 10  $\mu$ L of the resulting dilution was loaded and ran on a 10% SDS PAGE gel and visualized by staining with Coomassie Blue. The sizes of the molecular marker bands are listed in kilo-Daltons. Arrows denote the expected protein bands.



**Supplementary Figure S9. CNF2 domain T is able to be swapped into CNF1 without loss of efficiency, but utilizing CNFy domain T results in significant drop in efficiency**. The receptor-binding domain B and catalytic domain A were previously identified in CNF1, so to generate chimeras swapping domain T, we chose to hold these domains B, and A of CNF1 constant, while exchanging domain T using

the joining sites 223 and 688, 720 or 735. HEK293-T cells with reporter plasmids were treated with the indicated toxin at the indicated doses and subjected to SRE-luciferase assay to determine the normalized fold activation relative to CNF1 and untreated cells, as described in Figure 3. (A) Dose response curve comparing CNF1 cargo delivered by CNF2 translocation domain (CNF1 green closed triangles, CNF121-688 blue open circles, CNF121-720 pale green open circles). (B) Dose response curve comparing CNF1 cargo delivered by CNF2 translocation domain (CNF1 green closed triangles, CNF191-688 blue open squares, CNF191-720 pale green open squares, CNF191-735 orange open squares). (C&D) Corresponding scatter plots of all data points used to derive the best-fit lines and mean values show in (A) & (B), respectively.