Supplemental Information

Substrate-mediated fidelity mechanism ensures accurate decoding of proline codons

Byung Ran So^{1,4}, Songon An^{5,‡}, Sandeep Kumar^{1,4}, Mom Das^{1,3,4}, Daniel A. Turner¹, Christopher M. Hadad¹, and Karin Musier-Forsyth^{1,2,3,4}

Departments of ¹Chemistry and ²Biochemistry, ³The Ohio State Biochemistry Program, ⁴Center for RNA Biology, The Ohio State University, Columbus, OH 43210; ⁵Department of Chemistry, University of Minnesota, Minneapolis, MN 55455

Correspondence should be addressed to: Karin Musier-Forsyth, Department of Chemistry, The Ohio State University, 100 West 18th Ave., Columbus, OH 43210, USA Tel.: 614-292-2021; Fax: 614-688-5402; E-mail: musier@chemistry.ohio-state.edu [‡]Current address: Department of Chemistry, The Pennsylvania State University, University Park, PA 16802, USA

SUPPLEMENTAL METHODS

Prote in and tRNA preparation. All tRNAs used in this study were prepared by *in vitro* transcription as described (1,2). His-tagged *E. coli* ProRS (3), CysRS (4), WT AlaRS (5), C666A/Q584H AlaRS (AlaRS-CQ) (5) and *E. coli* tRNA nucleotidyltransferase (NTase) (6) were purified using the Talon cobalt affinity resin (Clontech) as previously described. *H. influenzae* YbaK protein was purified from *E. coli* B834 cells containing plasmid pCYB2_HI1434 using the IMPACTTM I system (New England Biolabs) as described (7). All of the YbaK point mutants were created from plasmid pET15b_HI1434 encoding the WT *H. influenzae* YbaK protein using the QuikChange site-directed mutagenesis kit (Stratagene). All YbaK mutants were overexpressed and purified as described previously (8). The concentrations of *E. coli* AlaRS-CQ, tRNA NTase and all YbaK proteins were determined by the Bradford assay (9). The concentrations of *E. coli* ProRS, AlaRS and CysRS were determined by active site titration (10).

3'-End-modified tRNA preparation. To prepare 3'-deoxy-A76 tRNA variants, the tRNA NTasecatalyzed nucleotide exchange reaction was performed in 20 mM Tris-HCl (pH 8.5), 20 mM MgCl₂, 1 mM sodium pyrophosphate, 5 mM ATP analogs (2'-H-ATP and 3'-H-ATP), 15 μ M *E. coli* tRNA^{Cys}, and 5 μ M tRNA NTase. After a 4 hr incubation at 37 °C, the reactions were subjected to phenol:chloroform extraction and ethanol precipitation. To remove residual amounts of the ATP analogs, dialysis against diethyl pyrocarbonate (DEPC)-treated water was carried out at room temperature. To eliminate the possibility of residual WT tRNA contamination, the sample was oxidized using 4 mM NaIO₄ in 50 mM sodium acetate (pH 5.0) for 1 hr at room temperature as described previously (11). The sample was dialyzed against DEPC-treated water and the final product containing the modified tRNAs was obtained by ethanol precipitation. Aminoacylation reactions were performed as described (2) to compare charging levels of deoxy-tRNAs to those of WT tRNA^{Cys} and to optimize conditions for Cys-tRNA preparation. $3'-[^{32}P]$ -labeled tRNA^{Pro} for aminoacylation assays was prepared as previously described (12).

Visualization of reaction products of YbaK-catalyzed deacylation of [35 S]-Cys-tRNA^{Pro}. To view the reaction products, some [35 S]-Cys-tRNA^{Pro} deacylation assays were analyzed using cellulose TLC according to the published conditions (13). Briefly, reactions containing 0.3 μ M [35 S]-Cys-tRNA^{Pro} in deacylation buffer were initiated with 0.1 μ M WT YbaK. At each time point, 2 μ L of reaction mixture was quenched with 8 μ L of 6 mM dithiothreitol in ethanol at 4 °C. The quenched mixture (1 μ L) was spotted on a cellulose TLC plate pre-run with water, developed in butanol: acetic acid: water (4:1:1 (v/v/v)), and exposed on phosphor-screens for 8-12 hrs. Screens were scanned using a Typhoon phosphorimager (GE healthcare). Deacylation assays for ESI-MS and LC/MS analysis were performed in similar fashion except 50 μ M non-radioactive Cys-tRNA^{Pro} and 10 μ M YbaK were used.

ESI-MS and LC/MS analysis. High-resolution ESI-MS was performed on a Bruker BioTOF II mass spectrometer (University of Minnesota). Crude reaction mixtures were mixed with equal volumes of methanol and data were recorded in the ESI mode. LC/MS was performed using multimode ESI-APCI ionization on an Agilent Technologies 110 series LC equipped with G1956B LC/MSD SL mass selective detector (University of Minnesota). A C8 column (150 mm x 5 μm) was used and the solvent elution conditions were as follows: (a) water/MeOH (0.1% formic acid) (20:80) isocratic elution or (b) water/MeOH (0.1% formic acid) (0-100%) linear gradient over 20 min using a 0.5 mL/min flow rate.

Computational analysis of cysteine and selenocysteine ring strain. The homodesmotic model (14) was utilized to accurately calculate the difference in ring strain between cysteine thiolactone and selenocysteine selenolactone. Unlike isodesmic calculations, the homodesmotic model accounts for the hybridization of each carbon atom on both sides of the reaction, as well as the same number of

hydrogen atoms attached to the carbon. Additionally, the lactone analog was computed to explore various trends in the three group VIA elements. Geometry optimizations were performed on all the homodesmotic model structures using both density functional theory (B3LYP/6-311+G(2df,2pd)) and the Complete Basis Set method (CBS-QB3). The Gaussian 09 (15) software package was chosen for the geometry optimizations. The frequencies of all optimized structures were computed and verified to have no imaginary frequencies.

SUPPLEMENTAL RESULTS

Effect of mutations and non-cognate substrates on binding of YbaK to CCA-Cys

To determine the factors controlling substrate binding by YbaK, we calculated the free energy of binding of YbaK and several mutants to 5'-CCA-Cys (cognate), as well as binding of WT YbaK to non-cognate CCA-Ala, CCA-Pro, and CCA-Ser substrates, employing the molecular mechanics Poisson–Boltzmann surface area (MM-PBSA) method. All *in silico* mutants of YbaK, as well as the non-cognate substrates, were obtained by manually modifying the structure of the YbaK:CCA-Cys complex, followed by energy minimization. These structures were subjected to 15 ns of MD simulation in explicit solvent and 250 snapshots of the protein, ligand, and the complex (without solvent molecules) were extracted at equal intervals. The total energy of each of the components was calculated in implicit solvent using the MM-PBSA module in AMBER 9. The free energy of binding was calculated as the difference between the total energy of the complex and that of the free ligand and the protein.

Our results show that while the binding free energy of the CCA-Pro substrate is moderately reduced compared to the CCA-Cys substrate due to the loss of H-bonding interactions of the amine group, the overall orientation of the CCA-Pro is preserved. In case of CCA-Ala and CCA-Ser, the binding free energy is similar to that of the cognate substrate (**Supplemental Table 2**). This implies that the CCA-Pro substrate can bind YbaK with an orientation similar to that of CCA-Cys, albeit with less affinity, whereas CCA-Ala and CCA-Ser substrates are not discriminated at the level of binding. These results provide further support for the hypothesis that deacylation by YbaK involves cysteine thiol chemistry. We also performed similar MD and MM-PBSA analyses for several YbaK mutants. As expected, the K46A mutation reduced the CCA-Cys binding free energy by 5-fold and perturbed the substrate orientation significantly, whereas CCA-Cys binding to T47A, S129A, and Y20A mutants remained relatively unperturbed (**Supplemental Table 2**). These data support the role of Lys46 in substrate binding and the role of Tyr20, Thr47, and Ser129 in subsequent catalysis.

Computational analysis of cysteine and selenocysteine ring strain

Using the homodesmotic model, the ring strain was calculated for the cysteine thiolactone, selenocysteine selenolactone and lactone analog. For both B3LYP/6-311+G(2df,2pd) and CBS-QB3 calculations, the selenolactone was calculated to have the least amount of ring strain, followed by thiolactone and lactone. Using B3LYP, the difference in ring strain energy for selenocysteine when compared to cysteine was 2.8 kcal/mol and was 3.6 kcal/mol with CBS-QB3 (**Supplemental Table 3**). Analysis of the lactone, thiolactone and selenolactone geometries shows that as the size of the group VIA elements increases, the bond length between the two carbon atoms and the group VIA element increases, which increases the angle opposite to the group VIA element. This increase in angle relieves some ring strain and can explain the difference between the thiolactone and selenolactone

(Supplemental Table 4).

Supplemental Table 1. Relative deacylation activities of YbaK against Ala-, Ser and Pro-tRNA^{Pro} substrates in the presence of high concentrations of *H. influenzae* YbaK.

Substrate	Ala-tRNA ^{Pro}		Ser-tRNA ^{Pro}		Pro-tRNA ^{Pro}				
Mutant	Relative activity	S.D.	Relative activity	S.D.	Relative activity	S.D.			
WT	1.00	0.12	1.00	0.06	1.00	0.03			
Y20A	0.48	0.04	0.26	0.13	0.20	0.13			
F29A	0.09	0.10	0.02	0.08	0.08	0.02			
K46A	0.10	0.12	0.18	0.08	0.17	0.06			
K46R	-0.04	0.05	0.20	0.02	0.27	0.04			
K46I	0.35	0.1	0.41	0.05	0.39	0.07			
T47A	0.12	0.06	0.16	0.09	0.30	0.05			
G101A	0.12	0.05	0.36	0.25	0.21	0.06			
S129A	0.04	0.03	0.05	0.02	0.21	0.03			
G131A	0.49	0.08	0.80	0.28	0.56	0.17			
G134A	1.04	0.22	0.91	0.21	1.10	0.02			
S136A	0.29	0.05	0.20	0.22	0.70	0.32			
S136H	0.30	0.06	1.12	0.11	0.45	0.05			

Ala-, Ser- and Pro-tRNA^{Pro} (1.0 μ M) deacylation by YbaK variants (21 μ M) was performed at room temperature. The k_{obs} (min⁻¹) values were calculated by fitting the percent Cys-tRNA^{Pro} remaining as a function of time to a linear equation, $y = k_{obs} \times t + A$. Relative activity was calculated by comparing the k_{obs} of deacylation with that of WT YbaK, which was set to 1.0 for each substrate tested. Results are the average of three trials with the standard deviation (S.D.) indicated. Negative values indicate substrates were protected from hydrolysis. WT YbaK deacylates Ala-, Ser-, and Pro-tRNA^{Pro} with 240, 572 and 466-fold reduced efficiency relative to Cys-tRNA^{Pro}.

Supplemental Tab	ole 2. Substrate	binding energies	of WT and mutant	t <i>H. influenzae</i> YbaK t	0
11				J	

YbaK	Ligand	Binding energy (kcal/mol)
	CCA-Cys	-47.15 ± 7.73
WT	CCA-Ala	-47.09 ± 6.00
	CCA-Pro	-19.88 ± 10.06
	CCA-Ser	-36.87 ± 6.82
Y20A	CCA-Cys	-43.31 ± 7.48
K46A	CCA-Cys	-9.47 ± 6.37
T47A	CCA-Cys	-41.06 ± 6.27
S129A	CCA-Cys	-39.72 ± 6.36

CCA-AA substrates calculated using MM-PBSA.

All *in silico* mutants of YbaK, as well as the non-cognate substrates, were obtained by manually modifying the structure of the YbaK:CCA-Cys complex, followed by energy minimization. These structures were subjected to 15 ns of MD simulation in explicit solvent and 250 snapshots of the protein, ligand and the complex (without solvent molecules) were extracted at equal intervals. The total energy of each of the components was calculated in implicit solvent using the MM-PBSA module in AMBER 9. The free energy of binding was calculated as the difference between the total energy of the complex and that of the free ligand and the protein.

Supplemental Table 3. Calculated ring strain energy values (in kcal/mol) for cysteine,

selenocysteine and the oxygen analog using the homodesmotic model (shown below, where X=O, S, Se).



Homodesmotic Model Compound	B3LYP/ 6-311+G(2df,2pd)	CBS-QB3
lactone (X=O)	36.4	36.8
cysteine thiolactone (X=S)	20.4	23.6
selenocysteine selenolactone (X=Se)	17.6	20.0

Supplemental Table 4. Geometric comparison of CBS-QB3 optimized NH_3^+ based thio- and selenolactones (bond lengths in Å, angles in degrees).

Parameters	H ₃ N ⁺ ² ₃ S	H ₃ N ⁺ 2 ⁻ Se	H S 3	O H Se	H_3N H_3N H
C3-S(Se)	1.86	2.01	1.82	1.96	
C1-S(Se)	1.78	1.94	1.78	1.94	
C1-S(Se)-C3	77.2	72.6	99.2	95.7	
C1-C2-C3	95.6	99.2			113.0

Supplemental Figure 1. Multiple sequence alignment of YbaK subfamily using CLUSTALW (16)

	10	20	30	40	50	2	60	70	80	9	0	100	110	120	13	0	140	150	160	170	
H. Influenzae	MEPAIDLLKKQKIPFI	LHTY	DHDP NNQHF GDE	AE	KLGID		TLLVAENG	DQKK	LACFVLATA	NXLNLKK	AAKS IGVK	VEXADEDA	QKSTGY	VGGISPLO	-QKKRVK-	TVINS	TALEFETI	VVSGGKRG	LSVEIAPODL	AKVLG-AEFTDI	VDE
P. aeruginosa	MTPAIDLLKKARAAHK	VLSY	SHDPKAPSYGLEA	AE	-KLGLD	- PQRVFK	TLLAATEK-	GE	LLVAVVPVI	GSLDLKA	LAHAAGAK	KADMADPQA	QRATGY	VGGISPLO	-QKKRLR-	TFIDE	SARQYPS II	HV <mark>S</mark> AGRRG	LEVELAAEVL	AQLTQ-GSFAAI	GRE
M. smegmatis	ATPA IAAL VAAG IQHD	VVRY	HHDPRNDSFGEEA	AA	-QLAADGY	VAEQVEK	TLVIALPK	G	LAVAVVPVP	TKLSLKA	AAAAL GVA	RAEMADPAA	AQRSTGY	VRGISPLO	-QRKPLP-	TVIDA	SALRFORVI	LCS AGKRG	LDIVLAPADL	IRATD-AVTAE I	AVG
C. crescentus CB15	STPATIAL TKAGVAFD	L YTY	DY DP GS E R VGL QA	AE	- AL GE D	PRRVLK	TLIVMLD	GK	PACAILPSD	QEVS MKK	LAAALKGK	SAAMAPVPD	AERITGY	(VGGV <mark>S</mark> PFO	-QKKRLP-	TVLEA	AALAEPYVI	FLNGGORG	LOVKLAPADA	VKAL GAVTAA IV	AE -
S. aureus	KTNAMEML DRAK I KYE	VHS FEV	PEEHLSGAEVA	EL IQANVKT-		VF K	TLVLENT	TKHE	HFVFVIPVS	ETLDMKK	AAALVGEK	KLQLMPL DNI	. KNVTGY	RGGCSPVC	MKTLFP		SCENYSHI	SVS GGL RT	MOITIAVEDL	ITITKCKIC	AVI
N. meningitid is Z2491	ITPAVEVLRENGIEFE	PFTY	AYEEHGGTAQF	AR	- L F <mark>G</mark> K D	- E HL V I <mark>K</mark>	TIVL QDE N-	GK	GL I VL MHGD	OKQISTRN	LARHLGAK	HIEPATPAQ	AN <mark>KWTGY</mark> I	VGGTTPFO	- IRTKLD-	IYVEC	SVMDLETI	INGGKRG	FIGIRPGDL	N- IL NPKTIQAA	V
S. pneumoniae R6	KTLVEQILSKAAIPHQ	GIQINA	LEGE-LPQ	GYERDQ-		IFK	TLAL LG	DKTG	PIIGIVPIT	QHLSEKK	LAKISGNKI	KVS MIPQKDI	EKTTGY	HGANNPVO	IROKHN	YP IF IDK	IALDLDRM	IVSAGEVG	HS I IVAP QDL	ASFVKADFV	DIL
B. subtilis	KTNAVRL IEQQNISYE	VL GY KT	DGQPVDGISVA	EKIGYSLKY-		VY <mark>K</mark>	TL VAAA	GT NH	YY VF VV <mark>P</mark> VA	E EL DVKK	AAKVTGEKI	KIEMIAMKEI	LAVTGY	RGGCSPIC	MKKLFP	TY I DA	SAETLDFI	IV <mark>S</mark> AGKIG	MOLKLAPODL	ARSCDAAFA	EIV
L. monocytogenes EGD-e	KTNACE IL DKQKVNYE	LREYAW	SEDSLDALHVA	EETGNNPAQ		IF K	TLVLTG	DK TG	NIVACIPAD	KTL HL KH	LAKISGNKI	KCEL IAVDSI	EKLTGY	RGGCSPIC	MKKLFP	TFIDN	SAESLOTM	LISAGKRG	LQIELAPADL	KKVVRAEFA	MIT
Nostoc sp. PCC 7120	KTNAARLLDKLGITYE	ILNYEV	DPDDLAAESTA	QKVGLPPEQ-		VFK	TLVVRG	DQTG	I CF AVL PGN	IAHL DL KA	LARISGNRI	KIE TVPLKE	QPLTCY	RGGVTALA	SKKSYP	TYVDE	TATLFEKI	TVS GGMRG	ML I L L NPS DY	LSALNGVVG	AIA
C. perfringens	KTNAME IL DS KKVS YE	ML S Y E S	EDGKIDGISVA	HK I GVDE KN-		VF K	TL VAQGI	TS KE	L YVF V I PVA	ELDLKN	AAKIAGEKI	KVEMIAVKD		IRGGCSP10	MKKNYK	TFIHE	SAEDL DF I	IF SAGK IG	HQIKMNPKDL	LSVVEGEFA	FLI
F. nucleatum	KTNAIRELEIHKIEHI	VREYEV	DEEHLDAVSVA	LKTNKDITR-		VFK	TLVLL NE	EKRE	MVVAC I PGM	IEKLDLKK	LGKLSGHK	KLEMLPMKDI	FSMIGY	RGGCSPIC	IKKRHS	IF THE	SAL DNKT II	L VS GGL RG	LQIEIEPQKL	IDYLKITVG	DII
X. campestris	ATRATEAL DAAGVAYV	LHPY	QY QADADAK GL QA	AQ	- ALQLP	- PHQVLK	TL MI WVD	EH/	AICVVIPCD	ORRESEKT	LASACGGK	SARMMEVAD	AERRIGY	VGGISPLO	- QQRPTP-	VLIES	SVLDAPAL	NINAGORG	LLLELAGAET	VHVLGARAYPLC	E
S. coelicolor A3(2)	GTPATVAL TAAGVTYT	VHAY	E H DP A H P S Y GE E A	AE	- AMGVS	- PERVFK	L VADVD	GAL	LTVAVVPVA	GOL DL KA	LAAAVGGK	RAAMADPAL	ERITGY	RGGISPLO	-QRKRLA-		SAS GHAT I	CVSAGRRG	LEVELSPDDL	ARL D-AVLAP	GRA
C. efficiens YS-314	ATPAL TKL TADQTPYE	LDIHGE	DIKSDKG-FGLDA	AE	- TMGVD	PARVEK	IL MAE TOGE	H	-VVAIVPVS	TILNEKA	LAKAGGGK	RAEMMORSRA	AQVVIGY	PGGISPIC	- QK HP HR -	VELDE	SATLQETI	TSAGRRG	WSLIMAPDDV	LRAAG-GQYADI	ADH
5. oneid ensis MK-1	MIPAVQMAKKAKIPFE	VLEY	SHUPHCAATGEEA	AN	-VLGLA	PAQVER	LLVAIDKA	AHAPI	LAVALVPVD	HOLNERA	VARCEGORI	KL QMADADL	AQUSSCI	VGGISPL	-QKKLLP-	IIIDE	SAKOFUKI	TVSAGREG	LEICLIANDL	AKLCK-GSFADT	KIL
L. plantarum WCFS1	KILVE KILDKANTS YQ	QYEFPI	EAEG-DVQQLQ-V	DHL GVDE HH-	14 COD	D V V V C	LALIG	NK1G	PVVGVLPLD	DEHLSYKK	LAKLSGNK	KVGMIPLKD	EKILGYO	HGANTPVC	IWEIKK	FPITLUE	TAKQQGKI	VSSGKIG	RSTETDADDL	KKLVHGIFG	PIT.
v. paranaemoiyucus	MIPAINLAKKKKVPHI	THUY		AL	-VLGQD	PKKVFK	TLUFCLINGE				AAKAAKGK	ADMADPDT	CATTON	RECEPTO	- UKKALP -	IFLHE	SAKE QE I II	LVS AGKRG	LETELAPOUL	VALTR-OUFAPL	CL-
E. laecalis V 5 8 5	KTNAMEMIVE QHK VPTK	ETEFAW	SED FILSAES VA	ESLUTERUR-					PVVAVIPUN	NELDLAR	LANASGINK	VE ML HL KDI	CALLOT	LCC LCP IC	MIK K QF P	TLAE	EAUUTSAT	TVS AGKRG	DUCLAPEAT	ATULO AVEADL	C I I
P. aucreyi	ETRATOL BORKVAFT		PHUPUNHNE GUE A	AP	OLCL D	BHAVVE	TI INF DE C	DQKKI			AAQALAV	NUAR COREV	OPUSCY		TRKRM		SAKDFATT	VICCERC	LUVE LAPRUL	ATVL Q-AVFADT	LD.
B. paraperussis				FOI CEDVRO		- FRAVV	TLVIP				APACONICI	SVE MIDOSE		RCCCCR			TCALERY	VISACVEC	I DI PIAPADI	LANTCAERACLO	CAT
W surcincomes	KTNAAP II DKKCIAVE	LIDVEV		OTI CMDEFE.		iv.		ET CE	VIVICIPAP		LAPNAOCK		L VITCY	RCCCSPLO			PAPEPECV	VSACMPC	KOIPLAPEKI	KEACEAE WAL	DI -
C din https://	ATPALAVI FAACIEHS	VSTE	FCCTDHECDAA	AA	AL DVD	SDRIFK	TIVIDITA	KCPKRO	AVACIPUT	SKISIKK		ATMADOHD	AKSSOV	PCCISPIC	- OK HPL P-		TAILEDTIE	EVSCOKRO	I DVELSCEDI	IRVVD-CSEADL	OAF
L iohnsonii	KTI LEKMI DOOK IPYK	OL OF AT	VOKG-DVKOMD-	SIL DDE FAL	ALDVD		TIVC-FO		PVVGVVPVT	FHISMER	AKASCNK	CELLPLKK	VATTON	HCANTPIO	IVOKHH	FPINIOS	SMKEHDHL	AVSSCEVC	RS IMINPODI	OKVTNATFV	DII
B hacteriovorus	VTI A IRELOKHOVOVT		PYFEKCCTAHS	SK	FI GVP	- FHHVIK	TI IMENEK-	KF	PL VVI MHCD	DVS TKO	ARELOVK	TISPCKPEV	DRHSCY	VOGTSPE	-TKREMP-	VYMEK	TIDDDI	INCOKEC	FLYSLKPODY	ORILERPTIVOVG	VKA
B. Jonaum	STPATVOLEKAGVEEH	VYEVEH	S NDHMDDGY GVE A	AT	KL GED	- EHOVEK	TI MADTG	AF	RVVGVVPVS	GHMDI KA	AAAVGAK	KAS MADPKV	MRESCY	VGGISPIC	-OKTRHK-	TVI DE	SAL OF DO LL	VSGGKRG	SVGVNPODI	KVI N-AVAAPI	GTW
R thur indiensis	- TNAMELL DKEKLEYS	MMTYDP	DDGKIDGVSVA	FKIGREVRE-		VYK	TI LAOGI	NS KN	HVELLEVD		AAKAVSEK	KIEMIPVKD	TKVSCY	RGGCSPV	MKKLES	TCIDA	SAOSI ET L	IVS GCK IG		AKVTRAOFG	FIT
Acinetobacter sp. ADP1	MTPACKLLKTNK IEYS	IHEY	EHDPNAKSEGLEA	AE	-KLNLD	VEEVEK	TLL VTDD	KN	HYVAILPVN		VAGAFOCK	KL HMADPKD	ERL TOY	VGGISPLO	-OKKRLK-	TVIDA	SAESKHKI	VSGGKRG	LDIGVAPOAL	AKILG-AOFVDI	LDO
P. acnes	AS AATKALAKAK IPHV	LHSY	DHDPSNHHFGDEC	AA	-KLGFD	-PSIMLK	TLVVELVPS	5GK	LAVAVVPVS	ROLDLKA	FASAVGAK	KVAMADPAA	ERATGY	VGGISPLO	-OKKRLP-	ICIDE	SVL GOPRVI		LSVELSPSDL	VTVTG-AVTAK I	SC-
D. psychrophila	MTPGVKQVKKAGMSYK	LHEY	KHDVRAES YGLEA	AE	-KLGVA	VERVEK	TL VVE DGD-	GN	MAVAIIPVS	CSLNLKK	MAMAL GVK	KINMADKVR	OKTTOY	LGGVSPL	-OKKRLP-	TVLES	SARE LAT I	VVSAGKRG	LDIELAPDVL	VTLLG-AAFADL	SSS
S. thermophilum	KT IAARML DOMK IPYE	IRTYEW	SEEEMDAVSVA	RKIGLPPEQ-		VF K		DRTG	VLVAS IPGN	AELDLKR	LAAVSGNK	KVE MVHVREI	OGLTCY	RGGVSPLO	MKKPYP	YYLDE	TAQIIEQLA	AVSAGORG	LOLILS GPDL	IRATGAKLA	DIA
I. loihiensis	MTPAIDVAKENE ISFK	VHEY	SHDPASES YGGEA	AE	-KLGVP	- QE QVF K	TL VVSLDG-	KEI	LAVGIIPVL	SRLSLKL	IAKAL GAK	KAAMATGSD	ERTTOY	LGGVSPLO	-OKKQLR-	TIIDT	SATYNPTVE	FISAGREG	VDMELTPEDL	KTLVR-GEFVNI	T
Z. mobilis	KTRGTAFLEKAGIAFT	VHPY	DY DP KAP AAGL QA	AE	- AL QQP	AE IVYK	TL MTEVD	GK	PVCVVVQVN	IHEVS MKK	LAAAAGGK	SANMIKPVD	ERMIGY	IVGG I SPF	-QKKRVP-	VIFDE	SAF QAEK I	FINGGORG	VL VAL APE DA	RRAVDCKIASVA	N
B. clausii	- TNAIRLLEQHKKPYT	LFSYPV	SDGKIDGQSVA	QK INKPEHV-		VYK	TLVAQGH	KSGA	IYVAL IPVN	IEELHLKK	LOKVTAEK	KVNMLLAKQ	HOVTNY	RGGCSPIC	MKKPYP	TFIDS	QAAQL DA IN	VVNAGKVG	LOMELRLSDL	VEAANAQFA	DLT
G. oxyd ans	DTSATLTLDALGISYE	RHVY	DY DP S AGKL GVAA	AA	- AL GL S	- PDHVLK	TLMVQID	GR	PACVVLPAD	DRELDFAR	VAQAF GGR	KAKMLSRPD	EQRICY	RIGGVSPFO	- QRSTVP -	TAFEC	CGMDGPPV	VINGGS QG	LL IELS PQDA	LTATNGITADLV	KPL
V. fischeri ES114	MTPATKLLKKAKIEHQ	VLEY	HHDPKAQAYGLEA	AE	- KL GL D	- QKE VF K	TL VVQVDE -	KE	LVVGIIPVA	E QL S MKL	IAKAVKGK	KAKMADPAV	QKTTGY	LGGVSPL	-QKKRLK-	TV ID I	TAESLES I	YV <mark>S</mark> GGKRG	LDIGLAPKDL	QQVL N-AQFLDI	KAE
B. linens BL2	ATPAVEVLNLAGIEYT	L HS F	DHDPSTRRYGSEA	AD	- KL GVG	- S DQVF K	TLMIQVD	GA	PVTAL VPVS	GOL DL KG	LASARGAK	KAQLAGVAE	ERRTGY	PVGGV <mark>S</mark> PF0	G-QRHAVP-	VVIDR	TALEHSSVI	FVSAGRRG	LEIEIRPEDL	VMLTN-AQVGK I	AAL
E. sibiricum	TTNVARLLKQSDIPFD	LLSYPV	DE Q DL S AQNVA	RKINYPLSS-		IF <mark>K</mark>	TLVLETC	DAHQ	YVFAV I SGE	EEVSLKA	VARAVSAK	KTHL VP MK DI	KQLTGY	IRGGVAP I	AKKP	FPVLLSH	SALNQEF I	IISAGKRG	LQVKLAPDDF	LRFTNGIVF	MNR
E. faecium	QTKVE RYLEEQK ISYQ	PLDFSH	VVEE-SFTEEL	QKRGIDPAL -		ICK	TLVVKG	NK TG	VIIALIPLN	IDHL DYKK	TROVTKDR	KVGFPGMDF	L MHTGY	HGANTP I	IKLAHP-D	YCFLADS	SLKEKEEV	IV <mark>S</mark> SGEIG	RSLRIKVADL	EKIVHPIYA	DL I
C. psychrerythraea 34H	MTPGINAAKKSKISYH	VHEY	S HDAAS GS Y GDE A	AQ	- KLGIS	- SERVFK	TLVVS IDN-	KAI	LVVAIIPVS	AMLSMKL	IAKAHGGKI	KATMALKTD	ERSTOY	L GGVSPL 0	-QKKRLN-	TFIDS	SSVQFTTVF	FVSAGKRG	LEIELSPQDL	KQLTQ-AKLTEL	CQ-
5. saprophyticus	KTNAMEML DRSKIPYE	VNTYEV	SEIHMDGESVA	EKVGVDPGL-		VYK	TLVLENI	TNHA	HEVEVIPVR	ESLOMKA	AAHAVGEK		KKVTGY	RGGCSPIC	MKRAFP	TIIDC	QAMDIDTL	VSGGERG	TOIKIDVRDL	INVTKAKVV	EVI
C. aurantiacus	KLNS MRLLERHHTAY T	VH	EYEYNDAYDAQT	AR	- MIGVP	- AS QVFK	L VVGGAG	VR	PILAMVPAD	DAQLSLKR	LAALIGVK	SLOLLPRAD	ERLIGLO	21GG IGAL	ALTAKGWI-	SYLDE	SALQYPVV	VNAGKRG	ILLGLAPTEL	VRVVG-AKVGAT	AER
P. halop lanktis	MIPATELLKKQKTGFE	LLSY	EHDPNIINFGLEA	AQ	-KLNLN	RDQVFK	TVLE NOD-	GQ	LIVAI	QUVNLKL	LASECGAK	KVAMAAAKK	QASTGY	LGGVSPLO	-QKKCLA-	IFLHS	SAAQFICI	VSAGKRG	LDIAINPHEL	AKLTA-AVLGEF	
D. desulturicans	MIPA IAAAKKAK IWFA	VHEY	AUDDREESFUKEA	AE	-KLEID	PERVEK	L VVAVS	ERD	LAVMVVPVL	HUL DL KL	AAKAL GVK	AAMADVKQ	ERATOY	L'GGVSPLC	- QKKKLP-		SAAL HAS VI	VSGGKRG	LUTELAPSUL	CALIG-GVF APV	AK-
H. Chejuensis	MIPAINQAKKAKIITQ	VHEY	AHDPRSESTGEER	AQ	AMGAP	PERVER	TLLVSLIG	AQK	LAVGIVPVS	GREDERA	VANALGAK	KAE MADPKE	RATGY	VGGISPLO	-QKKKLP-		SLNYAT H	VSAGRRG	LETELAPODL	LSLIG-GVSAAT	UVS
Janib acter sp	GIPATVAL GNAGIATA	VKAT	EHDPAAESFULEA	AC	ALGVE	DEEVEN			LAVGVVPVD		MAS AL GVK	VAMADPAA	CHE S S G T	VCCISPIC	- UKRALU-		SAHE HE IVI	LVS GGRRG	LDLELSPSDL	LKACCRACEADI	AR-
Marmononas sp	MERACINEL KKKK IPF3			VE						TVISEVV	MAKALCVK	VE MACANA			OKKRLA		SAESI AIM	UNS ACKRO		ATL TC AVE ADL	SVC
Reinekea sn	MERALDEL TRECIPEK		TVSOPTCHYCLDA	AS		PORVER	TLLVSIDA	DDD	AVCIL		AAKAHCAK	AOMADRTV	FKI TCY	IVCC IS PEC		CVIDE		TSCCPPN	LEVEENPODY	IPCI N-VKTAPI	CP-
M actinob acterium	- TPATVAL DAL TITEV		DHDS GT LEE GVE A	AT	AL CVD	ADOVEK	TI MTMVD		VVAVVPVS			AVMADESV	ORRICY	VCCISPIC	-OKTAHR-		TVELEETVE	EVSCCKRC	EDIEL APKDI	VRVTE-ATLAAL	ARS
P tunicata	MTPAVI FAKKNNIFFT	HOY	EHDPHHOSE CLEA	AF	KINIA	ACEVEK	TI VIEL DN-	HO	AVAILEVT	HOLNIKI	VAKALGMK	AOMAEPOK	FRITCY	I GOVSPI O	-OKKRIS-		SAOOLESI	SCGRRG	E LOLAPNOL	OOVI N-AOFAPI	CO-
Bacillus sp	- TNAMELL DAAKLEYN	SISYDS	KDGKIDGVSVA	CK ICROPEL -	RENEA		TI VAO AC		VVVFIIPVF	AFLDIKK	AAKAAGEKI	VEMIPVKD	OKL TCY	RGGCSPVC	MKKLYP	TE IDS	OGEOLESI	IVS CCK IC		KELLEAFWK	DIV
D. hafniense	KTNAAR IL DOAG LAYE	IPYEV	DEADLSATTVA	OK VNMPPEL -		IYK		DK TG	IL LACIPAH	IYELNIKA	LAOVSCNK	MEVVPLKE	OPL TOY	RGGVSPLO	TKKKYP	VYEOE	DML OHEK LA	AVSAGTRG	I OILALSODI		PIA
Psychromonas sp	MERAIRFLKDONVDEK	VHOY	SHOKKTKTYGEEA	LE	-KLGVG	SELVER	TL VVCL DE -	HE	LVVAVLPVL	EKLSMKK	VAHEFEAK	AHMAEKNA	MRSTGY	LGGVSPL	-OKKRLR-	TL IHO	SAFNVACIE	VSAGKRG	MD IELS AL OL	OTLVG-ASOADI	CV-
yen on one op							0000												and the state of the	10010	
																					2
conconvotion																					
conservation		-																			
															and the second	and the second					÷
	1+38324753149272	2503	21322003106	32		97*	++7500-	022	258688353	22976+2	7+374439	571671232	3329++4	3++5977+	-64222	56753	871343182	2788 34+	3382825237	23653-207	10-

	10	20	30	40	50	60	70	80	90	and the second	100	110	120	130	1.2	140	150	16	0	170
L. salivarius	KTNEE RVL DOHNIE YE	ETSENV	VIEOG-KDALKE	- AAEL GVEPES		IL KT IVL OG	KCDPK	DEEVVCLPIVH	EMDLKLV	AOOLGKK	VHLADNKI		IHGANTP IG IH	AFKGF	PITEDE	RIKPEEI	SVSAGEV	GRSVRLKC	DEDLVKLVI	
C. salexigens	MEPAIRLLERHATPHE	VLOY	PHOPKNSAYGE	EAAH	-ALGLAPES	VENTLLAOL	DDG	RLVVAL VPVTA	TLDLKRL	ASAAGTR	AKMADPTI	AERTTG	VVGGISPLG-0	RKRLP		ADTLAAN		GLEIRLAP	SALIGLU	E-ARTAPLARH
Oceanob acter sp	MEPAINLL IKONVGHT	IHEY	NHDARAESEGL	EAAD	-KLNVDASR	VENTLYVET	DNK	ELAVAIVEVNR		AKALKYK		VERSTO	VLGGVSPLG-0	KKRLR	- TVIDE	TTSLEDT	YVSAGKR	GLE IEL PA	TALAKLL	N-ADFANIATE
C. Koribacter	KTNAGE IL DOLK LOYE	REVEN	DPDDLAAFS	VAFKIGEPPEO		VENTI VAR-	- GDR N	GVCEAVVPGNY	EL DI KAL	ADATODR	LAL VPL KI	VOPL TO	IRGGVTAL GAR	KDYP		AFIEDVI	SI SAGVR	GTOIL LAP	ADYIRAVI	NATVAAIA
D. acetoxid ans	TTOAVROL KOHKVRYT	P 01	AYEEKGGTR	VSAR	- FI GVE EHA	VINTIVMED	EH0	SPL LU MHGDC	EVSTREN	ARTINVK	LAPCSPD	AHKHTG	OVGGTSPEG-T	KPLP	-VYVER		YINGGER	GEL VSLAP	EVIVNVI	DATEVEVAL
C. n hytofer mentans	ITNAMEMI TOAGLEYK		EEDHLAGVH	VAKOLOMPVEO		VENTI VAH-	- GEKK	CYLVEC I PVAE		ANLIGOK	IFLAPVKI		IRGGCSP IGME	KKEP		TAILEDEI	TVSACVR	CCOLL IPR	FEL IFEL	DATICDLT
P atlantica	MEPAINIL IKOKVAHE	V1 0	EHDASAESECI	FAAF	-KINIAPNI	VYKTIVVKI	DS 0	TI CVALL PVFR	MISMKIL	AKAAGAK	AOMADKT	VERVIC	VI COVSPI C-C	KRLT		AKAL TNA	WVSACKR	GLEINIAP	OAL TOLL	S-AREAAL CO-
A horkumensis	MERAVNAAKRAKVAFC	I HSY	EHDPKAPS VCI	FAAO	- AL NI PVER	VENTLIALV	DG	APVVAMVPVAH		ARAHCCK	AOMCPAD	AORI TO	VI COISPIC-C	KRIP		AAFOAF	EMSACRE	GLE LAMAP	RDLLALT	- CK MVAL VRD
H aurantiacus	RTRATOVICEVTSDEE		SEEA-TEETAE	FAAF	-KISIPINM	FKTLLAKG	EKK	CYVLAVVPS DH	RISIKKI	ADAL COK	VEMAPLE	FI TRI TC	KCCVSPLC-T	RRAMP	- VVI DE	AFOHAR	SI SACOR	GLOMI VAP	SHL ODAA	- AL VADISON
S frigid imarina	MERALELAKKHKIDEE		HHOTNSESVEL	E AAO	-KI OVPTEO	VENTIVVSI	DN0	AL VVCVVPVSS	MISMEL	AKAACAK	AHMAFAL	VERSSON	VI COVSPIC-C	KEPIP	TVIDI	ACOHOR	VISACKE	CVDISISE	ADLOWIL	-ATEADICAO
aln ha n roteob acterium	MERAINAANKAKVTET	T1 SY	EHDS NAOS EGL	FAAO	-KLOLDPIS	VENTIVIOC	DK	OL VVAL IPVND	TISEKKI	AKVVNAK	ARMADPH	VTNKTC	IMCGVSPLC-C	KTSL 0	- TILAK	ATLINI	HVSACKR	GLEIGISP	HDISOLL	R-AOFADITA-
0. oeni	KTOVE OLL DKNK IPYO	SLELN	LNOS-STELKK	NISKAGIPRES	L	IYKTLAS	NGDKS	GPV LAVL PVTK	HLDEKKL	AAVSONK	TEMIPLKI	OKTTO	IHGANNP VG IV	WHNHHGK F	PITEDK	TSLSDEL	YVSAGEL	NRS DL INA	KOVAEL	AS FADLL
L. casei	KTI VEKMI DKAGLAYV	PYFFPT	FEVG-DVAOLT	- IDHI DVPEHO		IXKTIVI	TONKT	CPVVGVVPI DR	HVDYKKI	SKVSGNR	VGMVPLK		EHGANTPIGIE	ERHKY	PIEISO	FAFDOAT	IVSACKV	GRSVGIDE	K IL AOF VI	DAOFADIA
L. delbrueckii	KTLVEKILDOHKIAYE	OHISE	SEDH-GVAOVG	- TGKPVLDGHP		VYKTLAV	NGNKT	GP IVGVVPLSG	HLDLKK	AKASGNK	CEMIPLK	DLEKTTO	VHGANTP IG I	HSKKF	PIFLDE	VLKEDG	WVS S GEV	GRSVMVNP	LDLOKLV	KATVADLO
L. mesenteroid es	KTI PEOVMOKHKVKYE	PLELN	LDKT-PVORDA	ILE TE HVKHDD		IYKTLAA	HGDKT	GP I VAVL P I TK	HISLKKI	AAASGNK	VAMERIKI	LORTTO	IHGANNPVG IV	NKHF	PITEDL	AAEOPE	L VS GGEL	SRSDKIN	ROVASLV	AEEVDLL
A. hvd rop hila	MTPAINLLKKLKIPHK	LYPYEC	EAHDDEGK	HAAT	- OL GL PEAO	VENTL LAHH	DK	OAVVAIVPSSG	LCSLKOL	AKATGLK	VEMMKPA	AEKLTG	KVGGISPLA-0	KKLLP	- TVL DD	SAMOF DE	LVSGGKR	GLSVGVAP	ADMLRLM	O-WIAAPIGE-
C. aggregans	KLNSMRVLEROOIPYT	V	DYDP-TLPDAV	AVAE	-AIGIPAAH	VENTL VAKS	D	RPLLAMVPAHC		AVLAGVK	VEMAPHVI	EAERLTG	RVGGIGALAL	AKRWP	- SYL DE	SALVPSKI	YVNAGOR	GVMIGIAP	SDLVRVV	- ATVGA ITAC
P. propionicus	TTOAVRHLROEHVVFT	DHLY	SYVDKGGTT	VSAR	-ELGVDEHI	VINTLYMED	DR O		OVSTREL	ARIVGVK	TPCTPE	TAHRHTG	LVGGTSPFG-T	RRKMP	- VYMEE	LAALPE	LINGGKR	GYL VR MOP	AELIRVL	OPTLVRVGI
C. cellulolyticum	KTNVMR IL DSAK INYN	MYTYDS	KGGAIDGVS	VAEKIGOPVER		VYKTLVTR-	-GNS K	NEEVEVIPVSK	ELNLKAA	AKAVGEK	IEMIRVD		IRGGCSPIGM	KDFK	-TVIDS	TCKNLAT	IF SGGK	GF OVE VNP	KELVNE I	KAGCG
S. amazonensis	MEPAIKOAEKAGIPES	VLEY	E HDS RAPS YGM	EAAD	-KLGLAPEL	VENTLMVAL	DEKSA	P LAVAL VPVAK	OLNEKLA	AKALGOK	LVMANPT	AAERSSG	LLGGISPLG-0	KKPLP	-TLIDI	AEALAON	HVSAGRR	GLE IALAP	OSLAGLC	R-GVEVALTG-
P. ingrahamii	MEPAVOLLNKHKMKFK	IHOY	QHDKNCDAY GL	E AAQ	-KMGIDGQL	VF KTL VVNV	DNE	NLLVCILPVTK	MLNLKK	AKVCTAK	VVLADKO	AVLRSTG	VLGGVSPLG-0	KKRLK	- TVL DN	SALAL OCL	YVSAGKR	GLE IAL NP	TDLINLT	N-ALNADICQ-
T. carb oxyd ivor ans	KTNAARILESLKISYE	LRKYEV	DEADLSAEA	VAAKLGLPPOO		VFKTLVAR-	- GDKT	GVLLACVPGAA	ELDLKAL	AAVSGNK	VELVPLKI	EVOPLTO	IRGGVSPLGT	KKYP	- VFLDE	AROWERV	SLSAGVR	GCLIILAP	HDLARAV	DATFCAIA
M. marina	KTNAMRLLDKHKAAYO	TVTYEY	DPENL DVAK	IAVDNQLELTS		IYKTLVVK-	- GDK T	GVWAVVPGNK	NLSLKAN	AQAS NNKI	VAMVPVKI	E I QGL TG	IRGGCSPLGM	KDYP	-VVIDTI	NAQVLPTI	FVNAGOR	GILVGTSP	IDLOKIT	AVWAAIS
Bacillus sp	KTNAIRLLEQQKIKFD	VIEYEI	GDGQVDGVS	VAEKIGOPVAR		VFKTLVAR-	-ASAH	KLFVFVIPVAD	ELDLKAA	AKVVGEK	IDMLPVK	DLLSYTC	VRGGCSPVGM	KLYP	- TV I DE	TAQQQGS I	IVSAGKI	GMQMHVQL	NDLKNIT	KAQLAPIT
Marinob acter sp	MTPAIDVAREAGIEHQ	LHQY	QR DP AS AGY GT	EAAE	-KLGLNPAC	VFKTLVVEV	DAK	TLVVAIVPVNA	MLS MKL	AKAASGKI	AIMADKQ	QVORSSG1	IL GGVSPLG-0	KRPLT	- TFIDE	SAQLFERI	HVSAGRR	GLEIELAP	EDLATL	Q-GRFVAL RQE
S. loihica	MTPATKML DKAK IPYR	LHEY	QHDANAGAY GL	E AAE	-KLQLPLEW	VFKTLVAEL	DNG	TLVVAIIPVDK	KL NL KQL	AKAAKAKI	AAMAAPEI	VORSTG	VL GGVSPIG-0	KRLA	- TFIDE	AQQLAQI	YVS GGRR	GLDIELTP	TALQDVT	Q-ARFVSLV
C. aerofaciens	KTNAMRELESAGISYV	LHTYEC	DODGDE S VGL GVA	ISEQLGEEPGQ		GF KTL VCV-	-TPSN	DYVVCCIPVAD	ELDLKSA	ARAAGEKS	LAMMHVK	DLLAATG	VR GGCS PVGM	KRYR	- TL IDE	TCVL WDT I	FISGGKR	GYQLELAP	DDLIAFC	GATVAAIT
E. coli	X TPAVKLLEKNK ISFQ	(I HT)	E HDP AE TNF GD	EVVK	-KLGLNPDQ	VYKTLLVAV	NGDXK	HLAVAVTPVAG	QL DL KKV	AKALGAKI	VEXADPX	VAORSTG	LVGGISPLG-0	KRLP	-TIIDA	PAQEFATI	YVSGGKR	GLDIELAA	GDLAKIL	D-AKFADIARR
M. gryp hiswald ense	KTPAT MAMEKSGKPFA	LLEYQY	DP DAQS IGL	HAAQ	-ALGLPAFQ	VF KTL MVQA	GD	EAVVAIVPSDR	ELNLKAL	AAAAGKK	ATMMKGPI	DAERITG	KIGGISPLG-0	KKRLR	- TFVDD	SALALDFI	VVNGGOR	GLQIKAIA	ADLMAAT	G-AVAAK LATS
S. tropica	VTPATTLLNKRKVRYR	T HP Y	DVPADAANYGA	LVAA	-TLGLPADR	VFKSLVAVV	DD	ALTIVVVPVSG	ELGLKAL	AAAVGAKI	AVLADRT	TAERVTG	VRGGISPLG-0	RNLP	- TVL DE	SALHHETI	YVSAGRR	GLQLELAP	QDL VAL	N-ARMARVAS D
N. aromaticivorans	ATRATKALEQAGIPFE	L RS Y	DY DP DADS IGL	AAAQ	- AL GVD PGQ	VFKTLMILA	DG	KPACVIAPS DG	EVSLKRA	AAALGAK	AEMMRPAI	EAERIAG	KVGGISPFG-0	2MRKVP	- VVIDE	TALLWDAI	FINGGOR	GL QVL VAC	EAAAGLL	GAVVADIATG-
Moritella sp	MTPAVKLAKKAKVAHK	T HE Y	KHDS SAGS YGL	EAAE	-KMAVAAER	IFKTLVVDV	GDK	KLVVAVVPV TA	MLNLKAI	AKVAKAK	AVMADKN	DVMRSTG	VLGGVSPLG-0	KKRLL		SAQTHETI	YVSAGRR	GLEIELSP	LDLKQLT	N-AEFAAISSD
P. pacifica	S TPAVELLS DE GVEFG	VHHY	DYVDRGGTQ	A <mark>SS</mark> A	- AL GVD E HA	VIKTLIFED	DS R	NPFVVLMHGDR	EVSAKQL	ARVLRVR	KPCEPA	VAEKHS G	RVGGTSPFG-1	RKALP	- VYAEA	SIFELPKV	Y INGGS R	GLLVS IDP	KVLRELA	KARPVRVAR
V. vad ensis	KTNVMR IL DQAK IPYS	HYTYE	REGE-AVDGLT	VAASMNQPPER		VFKTLVTR-	-GAGR	DFFVFVIPVAC	ELDLKAA	AHAVGEK	VEMIHVA	EINKVTG	IRGGCSPIGM	KQFR		SALEQRTI	IF SAGR I	GF QVE VAP	ADLEKLI	CSFAGIV
R. obeum	KTNAMRML DRQKVKYE	AFPYEC	DEFIDGLH	SADL IGAPYEQ		SFKTLVME -	- GKS G	GYYVFVVP IEK	EVDRKAA	AKAVGEK	AVDMIHVK	DILKITG	VRGGCSPLGM	KPYP	- VVF DA	SAGGFEEI	YVS GGRV	GLTLKVPL	AALLKVT	NGKLADIT
R. torques	KTNVMR IL DQKK IAYE	SYSY-V	DTGAVSGIE	VAE VL GE NP DM		VFKTLVTV-	- GKS K	TNYVFVIPVNC	ELDLKKA	AHAVKEK	VEMLKSKI	ELLPLTG	VHGGCSPIGM	KQFT	- TVIHK	TAKNFETN	IF SGGK	GYQVELSL	EDLLKVI	AKLEDVT
C. hom inis	KTNAARFLDELGINYE	LKSYKV	DPERIDAVH	I AE DAGL DITQ		VYKTIVCK-	- ADN	EFVVACLQGDL	SLNLKEL	AKIAGAKI	CELIAIK	DLPKITG	IRGGCSPLAM	RKFR	- TF I DN	KALLNDKI	YISAGIR	GLQIFLSS	KDLAKAT	JAL FCD IS
E. ventriosum	KTNVMRVL DS KK I KYN	EYYY-C	DTE AVS GVD	VAKAL GQNPEN		VF KTL VTT-	-GKSG	E HYVF MVP VAE	ELDLKKA	AKAVGEKS	TAMLKSKI	ELLGLTG	IHGGCSP IGM	KLFK	-TTTHK	TAP DF DT I	LFSGGKI	GF QVE IPF	EDFKKVI	NVNTADII
B. capillosus	KTNVMR IL DQKK IRYT	PHTYPC	PDGAVDGAT	VAGLIGVEPER		VFKTLVTR-	-GASK	QNYVF VVPVL G	SLNLKAA	AKAVGEKS	IEMIHQAI	ELLPLTG	VHGGCSPVGM	KQFK		AQDLEAN	IVVSAGK I	GFQVELSP	ADLAGLV	RARFAPIS
R. gnavus	KTNAME IL DRSKISYE	YQTYEC	DNFTDGTS	TADTLOLPHER		VYKILVII-	-GKSG	SHYVEVIPIEA	ELDLKKA	ARAVGEK	VEMLPVRI	ETTVIG	VRGGCTAIGM	KAFP	- TIQEI	NAQTLDY	YVSGGKL	GMQLKLSP	DDLKNVT	JAQYADVT
A. od ontolyticus	PTRATEAL TAAGVAFT	VHEY	EHDPAARSEGE	ETVE	-KLGIDPTQ	AFKTLMVRL	EPTG	EFVVGCVPARA	HLSMKL	AKAVGAK	AAMADPA	VAORREGI	VVGGISPLG-0	2TTSHR	-VEIDS	ACLOHETN	AT ISGGRR	GLSVELSP	LDLVELT	J-AEVTDLAAQ
Clostria ium sp	KINAMETL DRMKTPYT	LNQYQU	DEF IDGVH	VADELGOSYDE		SF STIVII-	-GKSR	SYYVFALPIDK	ETDLKKA	AKVVGEKI	NLELLPVK	DINKVIG	IRGGCTP IGM	KQYK	- VIHE	AKLLPQN	IN SGGKL	GEQILLKP	DULLRAV	NGVYADII
C. Iep um	K INAMEVL DWAK IPTE	TSTE	EDGKIDGVS	VAUKLUQPVEQ		VF NIL VIR-	-GIDK	NETVEVIPVAK	ELNERAA	ARAVGOK	VEMIHVDI	EIKNAIG	TRUCCEPTUM	KQYK	- VIDE	AURLSF	IVSGGKI	GUQIQISP	LMLCDLI	AQFASVA
A. Dutzieri	MIPAINLLKKNKCDFK	IHKY	E HUS DCT NEGD	EAVE	- KLGLDANQ	VYKILLVEL	CNT P	ELVVCVLPVAN	QUSLKEV	ATTEGVE	ALMASKD	AUSVIG	LLGGISPLG-G	KKLLK	- VL DE	KSFEI	FISGGER	GLDTEVKP	KDLEVLLI	C-AKIGKITA-
E. a olichum	KINAMETLUKARLETE	VLITER	QUE AVDULU	VAAKLHQQPQQ		VFNILVLI-	-5N1R	HTTVF MVP VDC	ELDERKC	ARCVINEN	LEMINVK	NLLSISU	IRGGCSP IGMP	KATP	- IVL HE	ALKFUK	TFSGGNI	GIGIAMB	VLLLEFI	AR ADIV
C. Doneae	KTNAMELL DAAR IVIE		DENDLSUSH	AADLING VDHDR			-GERK	NEVEN INVAA	EL DL KAA	ARAAGEN			VD CCC DVCM	KOVP		MAQL TOOK	WIV SAGOR	CTOVCCAR		
r. prausniizii			ELGV-AVDGVI	VAAS TUEDPAC			-USSK	ANVEYIDVE		AKAVCEKS				VEEP	TTIDA		NUSACHU	CYOLE LAR	ADL IRAA	
c. euacus	N THE REAL PORT IN	Incr-y	CIGCINGER	AT UQUPAT				ANT VEVIEVKE	LULIKA	AVUCN	TOWL KOK		Inducar I GM			AUN LOK N		GIQICIAP	NOLERVII	
						-												-		
								1 M M				- -								
				-																
conservation														1						
ochool valion											1 A 1									
							0 0							4222						
	1+38324753149272	2503	5213220031	0032		97 ++750-	0	22386883532	2976+27	+374439	16/123	273329+-	43++59/7+-6	4222	- 56/53	87134318	2/88 34	+3382825	2372365	3-20/10-

The abbreviations used are:

H. Influenzae Hi1434, Haemophilus influenzae; P. aeruginosa, Pseudomonas aeruginosa; M. smegmatis, Mycobacterium smegmatis; C. crescentus CB15, Caulobacter crescentus CB15; S. aureus, Staphylococcus aureus; N. meningitidis Z2491, Neisseria meningitidis Z2491; S. pneumoniae R6, Streptococcus pneumoniae R6; B. subtilis, Bacillus subtilis; L. monocytogenes EGD-e, Listeria monocytogenes EGD-e; C. perfringens, Clostridium perfringens str. 13; F. nucleatum, Fusobacterium nucleatum; X. campestris, Xanthomonas campestris; S. coelicolor A3(2), Streptomyces coelicolor A3(2); C. efficiens YS-314, Corynebacterium efficiens YS-314; S. oneidensis MR-1, Shewanella oneidensis MR-1; L. plantarum WCFS1, Lactobacillus plantarum WCFS1; V. parahaemolyticus, Vibrio parahaemolyticus; E. faecalis V583, Enterococcus faecalis V583; H. ducrevi, Haemophilus ducrevi; B. parapertussis, Bordetella parapertussis; Porphyromonas gingivalis W83, Porphyromonas gingivalis W83; W. succinogenes, Wolinella succinogenes DSM 1740; C. diphtheriae, Corynebacterium diphtheriae; L. johnsonii, Lactobacillus johnsonii; B. bacteriovorus, Bdellovibrio bacteriovorus; B. longum, Bifidobacterium longum; B. thuringiensis, Bacillus thuringiensis; P. acnes, Propionibacterium acnes; D. psychrophila, Desulfotalea psychrophila; S. thermophilum, Symbiobacterium thermophilum; I. loihiensis, Idiomarina loihiensis; Z. mobilis, Zvmomonas mobilis; B. clausii, Bacillus clausii; G. oxvdans, Gluconobacter oxvdans; V. fischeri ES114, Vibrio fischeri ES114; B. linens BL2, Brevibacterium linens BL2; E. sibiricum, Exiguobacterium sibiricum; E. faecium, Enterococcus faecium; C. psychrerythraea 34H, Colwellia psychrerythraea 34H; S. saprophyticus, Staphylococcus saprophyticus; C. aurantiacus, Chloroflexus aurantiacus; P. haloplanktis, Pseudoalteromonas haloplanktis; D. desulfuricans, Desulfovibrio desulfuricans; H. chejuensis, Hahella chejuensis; A. macleodii, Alteromonas macleodii; M. actinobacterium, Marine actinobacterium; P. tunicata, Pseudoalteromonas tunicata; D. hafniense, Desulfitobacterium hafniense; L. salivarius, Lactobacillus salivarius; C. Koribacter, Candidatus Koribacter: D. acetoxidans, Desulfuromonas acetoxidans; C. phytofermentans, Clostridium phytofermentans; P. atlantica, Pseudoalteromonas atlantica T6c; A. borkumensis, Alcanivorax borkumensis SK2; H. aurantiacus, Herpetosiphon aurantiacus; S. frigidimarina, Shewanella frigidimarina; O. oeni, Oenococcus oeni; L. casei, Lactobacillus casei; L. mesenteroides, Leuconostoc mesenteroides; A. hydrophila, Aeromonas hydrophila; C. aggregans, Chloroflexus aggregans; P. propionicus, Pelobacter propionicus; C. cellulolvticum, Clostridium cellulolyticum; S. amazonensis, Shewanella amazonensis; P. ingrahamii, Psychromonas ingrahamii; T. carboxydivorans, Thermosinus carboxydivorans Nor1; M. marina, Microscilla marina; S. loihica, Shewanella loihica; C. aerofaciens, Collinsella aerofaciens; E. coli,

Escherichia coli; M. gryphiswaldense, Magnetospirillum gryphiswaldense; S. tropica, Salinispora tropica; N. aromaticivorans, Novosphingobium aromaticivorans; P. pacifica, Plesiocystis pacifica; V. vadensis, Victivallis vadensis; R. obeum, Ruminococcus obeum; R. torques, Ruminococcus torques; C. hominis, Campylobacter hominis; B. capillosus, Bacteroides capillosus; R. gnavus, Ruminococcus gnavus; A. odontolyticus, Actinomyces odontolyticus; Clostridium sp, Clostridium sp. L2-50; C. leptum, Clostridium leptum; A. butzleri, Arcobacter butzleri RM4018; E. dolichum, Eubacterium dolichum; F. prausnitzii, Faecalibacterium prausnitzii; C. eutactus, Coprococcus eutactus Supplemental Figure 2. Relative Cys-tRNA^{Pro} editing activity of *H. influenzae* YbaK mutants. Cys-tRNA^{Pro} (1.0 μ M) deacylation by YbaK variants (1.0 μ M) was performed at room temperature. Relative activity was calculated by comparing the deacylation level with that of WT YbaK at 30 min. KPO₄ refers to deacylation performed in the presence of 150 mM KPO₄, pH 7.0. The negative level of relative deacylation activity by the K46A variant was due to higher levels of Cys-tRNA^{Pro} remaining relative to buffer hydrolysis. This is likely due to protection of the substrate by K46A YbaK.



Supplemental Figure 3. Cys-tRNA^{Pro} deacylation by YbaK in the presence of EDTA or ZnCl₂. Cys-tRNA^{Pro} (0.5 μ M) deacylation by WT YbaK (0.1 μ M) in presence of EDTA or ZnCl₂ (5mM) was performed at 37°C. The plots represent the average of three trials with standard deviation indicated.



Supplemental Figure 4. ESI-MS spectrum of crude reaction mixture following YbaK-

catalyzed Cys-tRNA^{Pro} **de acylation.** The data were collected in positive ionization mode. Methanol was used as the carrier solvent and PEG-300 was used as a reference for accurate mass calibration. The peak identified as cysteine methyl ester is highlighted in red and peaks due to the PEG-300 reference are marked as with an asterisk (*).



Supplemental Figure 5. Homology model of the editing (INS) domain of *E. faecalis* ProRS

bound to CCA-Ala. The model was generated by first superimposing the *E. faecalis* ProRS structure (2J3L) (17) on our model of *H. influenzae* YbaK bound to CCA-Cys. Next, substrate cysteine was mutated to alanine, followed by energy minimization using AMBER9. The hydrophobicity surface shown is colored blue for hydrophilic residues and orange for hydrophobic residues. The methyl sidechain of substrate alanine binds in a small hydrophobic pocket that would not accommodate larger amino acids such as cysteine or proline.



SUPPLEMENTAL REFERENCES

Additional References for Computational Methods Described in the Manuscript:

CLUSTALW:

Thompson, J.D., Higgins, D.G. & Gibson, T.J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673-4680 (1994).

AUTODOCK 4:

Huey, R., Morris, G.M., Olson, A.J. & Goodsell, D.S. A semiempirical free energy force field with charge-based desolvation. *J. Comput. Chem.* **28**, 1145-1152 (2007).

AMBER 9:

Case, D.A. et al. AMBER 9 (University of California, San Francisco, 2006).

ArchPRED (loop structure prediction server):

Fernandez-Fuentes, N., Zhai, J. & Fiser, A. ArchPRED: a template based loop structure prediction server. *Nucleic Acids Res.* **34**, W173-176 (2006).

PropKa (Prediction of protonation states of residues based on protein structure): Jensen, J.H., Li, H., Robertson, A.D. & Molina, P.A. Prediction and rationalization of protein pKa values using QM and QM/MM methods. *J. Phys. Chem. A* **109**, 6634-6643 (2005).

TIP3P (3-site model for water molecule):

Jorgensen, W.L. *et al.* Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926-935 (1983).

Adaptive Poisson-Boltzmann Solver (APBS):

Baker, N.A. *et al.* Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. USA* **98**, 10037-10041 (2001).

1. Beuning, P. J., and Musier-Forsyth, K. (2000) Proc. Natl. Acad. Sci. USA 97, 8916-8920

- 2. Shitive lband, S., and Hou, Y. M. (2005) J. Mol. Biol. 348, 513-521
- 3. Wong, F. C., Beuning, P. J., Nagan, M., Shiba, K., and Musier-Forsyth, K. (2002) *Biochemistry* **41**, 7108-7115
- 4. Newberry, K. J., Hou, Y. M., and Perona, J. J. (2002) *EMBO J.* **21**, 2778-2787
- 5. Beebe, K., Ribas De Pouplana, L., and Schimmel, P. (2003) *EMBO J.* 22, 668-675
- 6. Nordin, B. E., and Schimmel, P. (2002) J. Biol. Chem. 277, 20510-20517
- 7. Zhang, H., Huang, K., Li, Z., Banerjei, L., Fisher, K. E., Grishin, N. V., Eisenstein, E., and Herzberg, O. (2000) *Proteins: Struct. Funct. Bioinform.* **40**, 86-97
- 8. An, S., and Musier-Forsyth, K. (2004) J. Biol. Chem. 279, 42359-42362
- 9. Bradford, M. M. (1976) Anal. Biochem. 72, 248-254
- 10. Fersht, A. R., Ashford, J. S., Bruton, C. J., Jakes, R., Koch, G. L., and Hartley, B. S. (1975) *Biochemistry* 14, 1-4
- 11. Stathopoulos, C., Jacquin-Becker, C., Becker, H. D., Li, T., Ambrogelly, A., Longman, R., and Söll, D. (2001) *Biochemistry* **40**, 46-52
- 12. Wolfson, A. D., and Uhlenbeck, O. C. (2002) Proc. Natl. Acad. Sci. USA 99, 5965-5970
- 13. Jakubowski, H. (1994) Nucleic Acids Res. 22, 1155-1160
- 14. George, P., Trachtman, M., Bock, C. W., and Brett, A. M. (1975) *Theoretical Chemistry Accounts: Theory, Computation, and Modeling (Theoretica Chimica Acta)* **38**, 121-129
- Frisch, M. J. T., G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, N. J.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. (2009) Gaussian, Inc., Gaussian 09, Revision A.01 Ed., Wallingford, CT
- 16. Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994) *Nucleic Acids Res.* **22**, 4673-4680
- 17. Crepin, T., Yaremchuk, A., Tukalo, M., and Cusack, S. (2006) Structure 14, 1511-1525