Bioinformatics and Biology Insights



OPEN ACCESS Full open access to this and thousands of other papers at http://www.la-press.com.

ORIGINAL RESEARCH

Domain Organization of Long Autotransporter Signal Sequences

Jan A. Hiss and Gisbert Schneider

Johann Wolfgang Goethe-University, Chair for Chem- and Bioinformatics, Centre for Membrane Proteomics, Siesmayerstr. 70, D-60323 Frankfurt am Main, Germany. Email: hiss@bioinformatik.uni-frankfurt.de

Abstract: Bacterial autotransporters represent a diverse family of proteins that autonomously translocate across the inner membrane of Gram-negative bacteria *via* the Sec complex and across the outer bacterial membrane. They often possess exceptionally long N-terminal signal sequences. We analyzed 90 long signal sequences of bacterial autotransporters and members of the two-partner secretion pathway *in silico* and describe common domain organization found in 79 of these sequences. The domains are in agreement with previously published experimental data. Our algorithmic approach allows for the systematic identification of functionally different domains in long signal sequences.

Keywords: bacterial autotransporter, sequence analysis, pattern, protein targeting, signal peptide, protein trafficking

Bioinformatics and Biology Insights 2009:3 189-204

This article is available from http://www.la-press.com.

© the authors, licensee Libertas Academica Ltd.

This is an open access article distributed under the terms of the Creative Commons Attribution License (http://www.creativecommons.org/licenses/by/2.0) which permits unrestricted use, distribution and reproduction provided the original work is properly cited.

Introduction

Bacterial autotransporters translocate via the Sec complex across the inner membrane of Gram-negative bacteria and translocate themselves across the outer membrane.^{1,2} This is accomplished by a translocator domain at the C-terminus of the autotransporter which adopts a β -barrel fold within the outer membrane³ resembling a porin-like domain.¹ The trimeric autotransporter consist of an N-terminal signal sequence, a central "passenger domain", and a β-barrel forming translocation unit.³ The β -barrel domain is necessary for the secretion of the passenger domain and connected via an α-helical linker region.⁴ Bacterial autotransporters have been found in many Gram-negative bacteria and are often associated with virulence factors such as adhesion, biofilm formation, aggregation, invasion, and toxicity.5 For translocation across the inner bacterial membrane autotransporters possess N-terminal signal sequences.² In 2007 Dautin and Bernstein reported around 10% of the know autotransporters to contain a signal sequence with more than 50 residues. These N-terminal signal sequences exhibit a tripartite organization (n, h, c) as described by von Heijne.⁶ According to this nomenclature, "n" refers to an N-terminal region of the signal peptide which varies in length and often contains charged residues. The "h" or core region is a hydrophobic stretch required for the interaction between the signal peptide and SRP.7 "c" refers to the signal peptidase cleavage site. Additionally they can be roughly divided into two domains: i) an N-terminal extension of about 25 residues, ii) a C-terminal part that resembles a signal peptide.³ This division is in two domains, where one is like a functional signal peptide and is strikingly similar to the "NtraC model" which has recently been introduced by the writers as a general model of long eukaryotic signal peptides.8

Henderson et al reported at least 80 proteobacterial autotransporters with a signal sequence of at least 40 residues and published a list containing 46 sequences.¹ The authors propose four different regions based on hydrophobic and charged residue distribution (N1,H1,N2,H2) and a C region (cleavage site) following the standard *n*, *h*, *c* organization of export signals according to von Heijne.⁶ Desvaux et al continued this approach and termed the N2 and H2 region the "extended signal peptide region" (ESPR).⁹ They propose that the ESPR may be important for



additional functions besides targeting. In this report, we extend and formalize this approach by proposing a dual domain organization proposed by our algorithm.

Materials and Methods

We analyzed 16 autotransporters and two-partner secretion sequences published by Szabady et al¹⁰ and 35 further long signal sequences of bacterial autotransporters taken from Henderson et al.¹ Two-partner secreted proteins are known to possess an N-terminal conserved region important for their secretion.¹¹ Additionally we performed a sequence database search in UniProtKB/SwissProt Release 14.712 using the sequence retrieval system (SRS, Release 7.1.3).¹³ We searched for proteobacterial sequences with an annotated similarity to an autotransporter domain and a signal sequence of at least 40 residues, resulting in 56 sequences. Of those 56 sequences 39 were not considered in the work of Henderson et al¹ and Szabady et al.¹⁰ From the sequences considered suitable by the work of Henderson et al¹ Szabady et al¹⁰ and our own database search we assembled a dataset of 90 sequences. The signal peptidase cleavage sites were used as suggested in Henderson et al¹ Szabady et al¹⁰ and for the 39 sequences retrieved via SRS as annotated in SwissProt UniProtKB/SwissProt Release 14.7, respectively. The SwissProt database entries contain sequences with predicted or putative signal sequences.

The following sequences were omitted from our analysis due to minor sequence aberrations between the publications and the UniProtKB database entry: O32591, Q47692, Q54151 and Q8VSL2.

The following to YP_001161762 orthologous sequences were omitted since they possess an identical signal sequence: YP_001719317, YP_001874066, Q1C309, Q1CMJ2 and Q665P5. When two database entries contained the identical sequences one entry was omitted and both accession numbers are given.

In total, 90 signal sequences encompassing more than 40 residues from bacterial autotransporters were analyzed in this study and in regards to their possible internal domain organization.

The 28 long signal sequences not associated with autotransporters were retrived from the UniProtKB/ SwissProt Release 14.7¹² using the Sequence Retrieval System (SRS, Release 7.1.3).¹³ We searched for



"non-potential" bacterial signal sequences with evidence at protein level and a length of at least 40 residues. All retrived sequences contain the twinarginine (TAT)¹⁴ signal which leads to export to the periplasm or extracellular space (Suppl. Table 1).

The 228 short bacterial signal peptides associated with autotransporters were retrieved from the UniProtKB/SwissProt Release 14.7¹² using SRS (Release 7.1.3).¹³ We searched for proteobacterial signal sequences with less than 40 residues that contain annotated similarities to known autotransporters (Suppl. Table 2).

The detection of the domains was performed using the NtraC algorithm,⁸ an algorithmic approach to identifying domains in long eukaryotic signal peptides based on secondary structure aspects. The NtraC model proposes one domain to be essential and sufficient for targeting while rendering the other domain free for additional functions. Here, "N" and "C" denote two potential domains: an N-terminal "N-domain" and a C-terminal "C-domain" predicted by the algorithm. The transition area between both domains is refered to as "tra". The algorithm works on the complete signal peptide sequence and suggests the domain positions. The N- and C-domains contain targeting signals that are not detectable when the whole signal sequence is regarded as an entity as performed by current prediction software. Until recently six predicted domains have already been tested experimentally in vitro, from which five exhibit the predicted targeting function8 (Resch and Hiss in preparation).

Results and Discussion

We analyzed 90 long signal sequences of bacterial autotransporters and the two-partner secretion pathway in regards to their potential two-domain (NtraC) organization.

Of the 16 signal sequences collected in Szabady et al¹⁰ 14 are predicted to have a two-domain organization (Table 1). Of the 46 autotransporter signal sequences collected in Henderson et al¹ 35 are not listed in Szabady et al. Of those 35 sequences 32 are predicted to have a two-domain organization (Table 1). Of the 39 sequences we found via SRS and which are not described in Henderson et al¹⁰ or Szabady et al 2005, 31 are predicted to have a two-domain organization. In total, from 90 long signal sequences considered in this study 77 (86%) are predicted, by our algorithm, to be organized in two domains.

For two additional sequences (Q2J0N4, CAR56027) an NtraC organization is predicted which in the context of this work could be regarded as a false-positive: No C-domain with a targeting capacity was detected. For Q2J0N4 an N-terminal mTP is predicted by TargetP^{15,16} and for CAR56027 a signal anchor by SignalP.¹⁷ If these two sequences are included a total of 79 of 90 (88%) signal sequences are predicted to be organized in two domains.

The two-domain organization proposed by the algorithm is in agreement with the ESPR of Desvaux et al⁹ and the conservation of the "N-terminal extension" reported by Szabady et al¹⁰ within a margin of ± 5 residues.

Szabaday et al¹⁰ further reported a conserved sequence pattern in the N-terminal extension of autotransporter and two-partner secretion systems signal peptides.¹⁰ This conservation is also present in 43 of 46 sequences compiled by Henderson et al.¹ For the long signals sequences extracted via SRS, the conserved pattern is only present in three sequences (Table 1, no. 52-54). One of these three (no. 54, Q8CWC7) is a Pic variant of a different *E.coli* strain (Table 1, Nr. 4). The remaining 36 sequences found via SRS do not show the conserved sequence motif reported by Szabady et al¹⁰ although they are annotated to contain a domain similar to autotransporters. This might argue for the sequences of Szabady et al¹⁰ to form a group. Nevertheless, we found that the SRS sequences have the same domain structure in their long signal peptide as the sequences reported by Szabady et al¹⁰ and Henderson et al.¹ The N-terminal region of this group of autotransporters may have a different function not requiring this conserved motif pattern.

We want to highlight the case of the long signal peptide of EspP. EspP is an extracellular serine protease of *E. coli* which is divided into four subtypes α , β , γ and δ of which α and γ are proteolytically active.^{18,19} The long signal peptide of subtype EspP α contains the conserved sequence pattern reported by Szabady et al¹⁰ and for which experimental results were published by Peterson et al.²⁰ These authors showed that residues 23–55 can act as an independent targeting signal. In 2006 Peterson proposed the N-terminal extension of the signal

Nr.	Accession number ¹	SP length²	NtraC ³	N-domain ⁴	C-domain	Predicted C-domain targeting	Organism
-	NP_052685	55	yes	1–26	27–55	gram-5	E.colř
2	AAC44731/ AAG37043	53	yes	1–17(1–35)	18–53(36–53)	gram–	E.coli ^p
ო	CAA11507	52	yes	1–17	17–52	gram-	E.coli ⁸
4	AAD23953	55	yes	1-15(16)	17-55	SP/SA ⁹	E.coli ¹⁰
ß	CAA88252	56	ОП	I	I	I	S.flexneri
9	CA46156	42	yes	1–26	27(28, 29)–42	gram-	E.coli ¹¹
7	CAC14227	75	yes	1-40	41–75	gram-	Y.pestis
8	AAC43721	50	no	I	I	I	H.influenzae
6	CAC14202	62	yes	1–27(–44)	28–62	gram-	P.multocida
10	AAK68872/ AAK09243	51	yes	1–26	27–51	gram–	N.meningitidis
1	AAG01335	66	yes	1–27	28–66	gram-	X.oryzae
12	CAD12824/ AAA22974	71	yes	1–28(–44)	29–71	gram–	B.pertussis
13	CAI77662	70	yes	1–27	28–70	gram-	H.influenzae
14	AAM88788	62	yes	1–38	39–62	gram-	P.luminescens
15	NP_253231	53	yes	1–24	25–53	mtp, SP, gram+	P.aeruginosa
16	NP_252771	52	yes	1–23	24–52	gram-	P.aeruginosa
17	AAA20524	68	yes	1–26(mTP); 1–35	35–68	gram-	H.influenza
18	AAB96359	48	yes	1-15(1-25)	16—48(26—48)	gram-	M.catarrhalis
19	AAF40927	78	yes	1—43(1—50)	44-78(51-78)	gram-	N.meningitis
20	AAF67320	54	yes	1–30	31–54	gram-	S.flexneri
21	AAG30168	49	yes	1–29	30–49	gram-	E.coli ^{†2}
22	AAK00474	52	yes	1–33(1–36) (mTP,SP)	37–52	Ι	S.flexneri
23	AAK77860	69	yes	1–43	44–69	gram-	Y.enterocolitica
24	AAL18821	52	yes	1-15(1-29)	16–52(30–52)	gram-	E.coli ⁷³
25	AAL78284	67	yes	1–30(39)	31(40)–67	gram-	M.catarrhalis
26	AAQ22366	56	no	Ι	I	Ι	A.actinomycetemcomitans
27	CAA88252	56	ou	Ι	I	Ι	S.flexneri
28	CAC14203	71	yes	1–26	27-71	gram-	P.multicoda
29	CAC14218	84	yes	1–27 (48)	28(49)–81	SP	A.ferrooxidans
30	CAC39286	52	yes	1–16 (29)	17(30)–52	gram-	E.coli ¹⁴

Table 1. NtraC analysis of 90 long bacterial signal peptides.



			_	-	-														ca					~		is		ip. 638	· sp. PRwf-1					(Continued)
Y.pestis	N.meningitis	E.coli ¹⁶	R. solanacearu	R. solanacearu	R. solanacearu	X.campestris	S.enterica	B.xenovarans	B.fungorum	X.campestris	A.vinelandii	H.somnus	H.somnus	H.somnus	E.coli ¹⁷	E.coli ¹⁸	E.coli ¹⁹	E.coli ²⁰	B.bronchisepti	E.coli ²¹	E.coli ²²	S.flexneri	E.coli ²³	B.cenocepacia	P.denitrificans	B.vietnamiens.	Y. pestis	Enterobacter s	Psychrobacter	O.anthropi	O.anthropi	O.anthropi	B.multivorans	
gram-	gram-	mTP ¹⁵ , SA, SP	gram-	gram-	gram-	gram-	SP, mTP	gram-	gram-	SP	gram-	gram-	SP	gram-	Ι	gram-	gram-	gram-	gram-	gramp+, SP	gram+	gram+, SP	SA	SP	gram-, SP	gram-, SP, gram+	gram-, SP, gram+	I	I	I	SP, gram-	gram–, SP, gram+	gram–, SP, gram+	
38(42)–69	49(50)–80	16(25)(39)-48	33(-40)-72	31(–39)–60	34(35, 42)–66	27(28)(31)(44)–66	29(39)–50	30(-38)-57	29(30)–57	30(33, 35)–67	18(29, 32, 33)–49	41(43, 48, 52, 53)–77	38(39, 44, 49)–69	41(42, 43)–69	I	30(41, 42)–61	16(28, 29, 30)–54	33–52	23(29, 30, 39)–65	17–52	17–56	18–52	17–55	25-41	21-40	20–47	16(32)–50	I	I	I	26–45	21–45	25-41	
1-37(41)	1-48(49)mTP	1-15(24)(38)	1–36(37)(39)	1–30(–38)	1–33(34)(41)	1-43(55,56,58)	1–32(39)	1–37(mTP)	1–28(29)	1-58(mTP)	1–32(mTP)	1-40	1–37	1-40	I	1–29(40, mTP)	1–15(27, 28, 29)	1–32	1–38 (mTP)	1–16	1–16	1-16(17)	1–16	1–24(SA, mTP)	1–19(20)	1–18(19)	1–15(28)	I	I	I	1–25(mTP)	1–20	1–18(24)	
yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	ou	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	ou	ou	ou	yes	yes	yes	
69	80	48	72	60	66	66	50	57	57	67	49	77	69	69	52	61	54	52	65	52	56	52	55	41	40	47	50	48	41	40	45	45	41	
CAC92482	NP_274768	NP_308389	NP_519008	NP_519896	NP_522634	NP_636050	NP_807449	YP_553065	ZP_00033562	ZP_00041732	ZP_00088699	ZP_00122019	YP_719000	ZP_00132251	AAC26634	AAC74583	AAD41751	AAF43424	AAG53941	AAP33781	Q84GK0	Q7BCK4	Q8CWC7	YP_840389	YP_916859	YP_001110606	YP_001161762	YP_001165485	YP_001281217	YP_001371305	YP_001371348	YP_001373319	YP_001948780	
31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	



P



Nr.	Accession number ¹	SP length²	NtraC ³	N-domain ⁴	C-domain	Predicted C-domain targeting	Organism
65	YP_001683300	42	yes	1-13(23)	24-42	gram-, SP, gram+	Caulobacter sp. K31
66	YP_001766852	47	yes	1–28(mTP)	29–47	gran negative SP, SP	M.radiotolerans
67	ZP_02906872	42	no	I	I	I	B.ambifaria
68	YP_001811545	42	no	Ι	Ι	I	B.ambifaria
69	YP_001990882	50	yes	1-18(-23)	24–50	gram-, SP, gram+	R.palustris
70	YP_001993037	45	yes	1–19	20–45	gram-, SP, gram+	R.palustris
71	YP_001993508	46	yes	1–19	20-46	gram-, SP, gram+	R.palustris
72	YP_002027122	40	ou	Ι	Ι	I	S.maltophilia
73	Q0AL61	45	yes	1-13(29,35)mTP	14(30, 36)–45	gram-, SP, gram+	M.maris
74	Q0B7F7	47	yes	1-17(1-33 mTP)	18–47	gram-, SP, gram+	B.cepacia
75	Q0PAN9	40	yes	1–17	18-40	gram-, SP, gram+	C.jejuni
76	Q1MKD9	40	yes	1–27(29) (mTP)	30-40	I	R.leguminosarum
77	Q20XG2	49	yes	1-15(-20)	21–49	gram-, SP, gram+	R.palustris
78	Q214C9	50	yes	1–19	20-50	gram-, SP, gram+, mTP	R.palustris
79	Q216F9	42	yes	1-14(-18)	19–42	gram-, SP, gram+	R.palustris
80	Q2J0N4	44	(yes)	1–37(mTP)	38-44	I	R.palustris
81	Q48237	48	yes	1–25	26-48	gram-, SP, gram+	H.mustelae
82	CAR56027	41	(yes)	1-30(31)(SA)	32-40	I	B.cenocepacia
83	Q132B9	44	yes	1–17	18-44	gram-, SP, gram+	R.palustris
84	Q137J7	43	yes	1–26(27)	28-43	gram–, SP	R.palustris
85	Q2KVL9	47	yes	1–15	16-47	gram negatice SP	B.avium
86	Q4ZQ08	43	yes	1–29(mTP)	30-43	SP	P.syringae
87	Q6N2N9	46	yes	1–19	20-46	gram-, SP, gram+	R.palustris
88	Q6N3Z4	46	yes	1-17(-21)	22-46	gram-, SP, gram+	R.palustris
89	Q6N964	46	yes	1–18(23)	24-46	gram-, SP, gram+	R.palustris
06	P39180	52	yes	1–35(u.–29 mTP)	36–52	SP, gram– (30–52)	E.coli ²⁴
¹ Nr. 1- public of the or Tary ¹¹ strair	-16 taken from Szabad ations or annotated in Ur domains. A targeting at $getP. ^{5-/+:}$ gram negativ $h = K-12. ^{12}$ strain = CFT0 $h = APEC13. ^{22}$ strain = O	y et al. ¹⁰ ; Nr. niProtKB v. 14. bbreviation in t <i>i</i> e or gram po.)73. ¹³ strain = E)78:H11. ²³ strain	17–51 taken fr 7. ³ NtraC orgal brackets mean sitive signal pe EH41. ⁴ strain = in = O6:H1. ²⁴ S	om Henderson et al ¹⁰ ; Nr. 52– nization of the sequence. ⁴ Leng s: only that length combination sptide. ⁶ strain = Sakai. ⁷ strain = ±4797/97. ¹⁵ mitochondrial targe train = K12.	90 retrieved via SRS. ² gives the th of the predicted N- or C-domal leads to the targeting function = E2348/69. *strain = EB1. *SP: ting peptide. ¹⁶ strain = Sakai. ¹⁷ st	e position of the last residue of i in. Numbers in brackets refer to al in brackets. Predicted C-doma eukaryotic signal peptide, SA: s train = 042. ¹⁶ strain = K-12. ¹⁹ strai	the signal peptides as used in the thernative possibilities for truncation in targeting: predicted by SignalP signal anchor gram. ¹⁶ strain = 042. in = H10407. ²⁰ strain = ML308–225.





sequences to mediate an interaction with an unknown cytosolic factor or to induce an unusual signal peptide conformation prior to protein translocation.²¹ Notably, the analysis of the 55 residue signal sequence of EspP by our algorithm identified a two-domain (NtraC) organization:

- N-Domain (residues 1–26): unknown function,
- C-domain (residues 27–55): predicted secretion signal for Gram-negative bacteria.

The algorithm thereby proposed the same functional domain Peterson et al described experimentally.

We would like to stress that the NtraC algorithm is based on sequence information only and not influenced by the existing proposed fragmentation of long signal peptides. Our prediction method is therefore unbiased for the analysis of new sequences.

A further surprising result is the prediction of mitochondrial targeting peptides (mTP) for the proposed N-domains of the long bacterial signal peptides. In 17 of 90 (19%) cases the N-domain of a bacterial signal sequence is predicted as mTP (Table 1). Short bacterial signal peptides associated with autotransporters are in 29 of 228 (13%) cases predicted as mTP. As the presence of arginine is a typical feature for mTPs^{15,16} this could, in our case, lead to a prediction of a sequence as mTP if arginine residues are abundant. The positive charged residues are thought to form amphiphilic α -helices.²² This high abundance of positive charges (Table 2) has also been observed in the extended N-region of bacterial autotransporter signal sequences by Peterson et al.²¹ They reported a high net positive charge to be common in the N-terminus of serine protease autotransporters.

The automatic assignment "mTP" should thus not to be regarded as a perfect functional prediction but as the detection of a feature, namely the high abundance of charged residues. In 1994 Izard and Kendall reported that although a positive charge in the N-terminus may not be absolutely required for secretion²³ a net negative charge or zero charge could result in considerably decreased rates of export.^{24–26} While Dierstein and Wickner reported that the N-terminal regions is not strictly required for processing by signal peptidase,²⁷ Peterson et al demonstrated that the positive charges in the N-terminal part of the bacterial signal sequences may influence SRP recognition.²⁰

To investigate the role of charged residues *in silico* in the context of long signal sequences of autotransporters and their potential domain organization we counted the occurrence of charged residues in the N- and C-domain of all 79 autotransporter sequences predicted to be two-domain organized (Table 2). The border between the N-domain and the C-domain (transition area, "tra") often contains charged residues. To take this into account the border between both domains was alternating, and included (+tra) or excluded (-tra) from the domains for the calculation (Table 2).

If the border between the domains was regarded as part of the C-domain(+*tra*), positively charged residues (His, Lys, Arg) occur approximately 1.6 times more often in the N-domain compared to the C-domain. Negatively charged residues (Asp, Glu) occur 2.3 times more often in the N-domain(-*tra*) compared to the C-domain(+*tra*). This difference becomes even more prominent if the border between both domains is counted as part of the N-domain(+*tra*) leading to 2.8 times higher occurrence of positively charged residues and 4.2 times higher occurrence of negatively charged residues in the N-domain compared to the C-domain.

 Table 2. Mean occurrence of charged residues given in one letter code in 79 NtraC-organized autotransporter signal sequences the length of the transition area (tra) is up to eight residues.

	His	Lys	Arg	Asp	Glu
N-domain ²⁵ (– <i>tra</i>) ²⁶	0.5 (±0.8) ²⁷	1.9 (±1.5)	2.3 (±1.7)	0.4 (±0.6)	0.8 (±0.7)
C-domain (+ <i>tra</i>)	0.3 (±0.5)	1.1 (±1.4)	1.5 (±1.4)	0.2 (±04)	0.3 (±0.6)
N-domain (+ <i>tra</i>)	0.6 (±0.8)	2.3 (±1.8)	2.6 (±1.8)	0.4 (±0.7)	0.9 (±0.7)
C-domain (<i>–tra</i>)	0.2 (±0.5)	0.8 (±1.0)	1.1 (±1.2)	0.1 (±0.4)	0.2 (±0.5)

²⁵Domains of the NtraC algorithm.⁸ ²⁶Excluding or including the transition area (domain border) predicted by the NtraC model. ²⁷Standard deviation in brackets.





Figure 1. Cartoon of a possible membrane orientation of the signal peptide of EspP. **A**) Potential orientation of the N- and C-domain of the long signal peptide of EspP. **B**) N- and C-domain of the signal peptide of EspP. **Blue:** positively charged residues. **Underlined:** predicted β -turns.

This charge bias is an argument that charged residues may represent an inherent difference between the N- and the C-domain. The nearly three-fold increase in deviance between the N-(+tra) and C-domain(-tra) indicates that not only the presence of charged residues is of importance but also their position, favoring the N-terminal domain or between the two domains. The relative position curtly before the targeting signal in the C-domain could represent a characteristic feature.

The observed abundance of charged residues in the N-domain was also reported for the long signal peptides from vertebrata analyzed by us previously.^{8,28} The authors therefore propose that a potential additional function of the N-domain in long signal peptides is related to the abundance of positively charged residues in Gram-negative bacteria as well as in vertebrata. This is in agreement with the observation made by Peterson et al^{20,21} regarding a high net positive charge of the N-terminal part of the signal sequence and its potential influence on SRP recruitment. The NtraC algorithmic approach can be used to check individual observations, and pinpoint the sequence part that might be relevant for such an SRP interaction.

A further hint towards a mechanistic aspect arises from to the secondary structure aspect of the NtraC model. As the C-domain of the signal peptide with its hydrophobic core is embedded in the membrane or the Sec complex during translocation, the N-domain may be kept in a defined angle to the membrane due to a predicted β -turn in the border between the N- and C-domain. The positive net charged of the N-domain could have the effect of keeping it outside and on top of the membrane. This might provide the means for the recruitment of other proteins (Fig. 1). We further report priliminary in silico results that 43 out of the 90 (48%) long signal peptide sequences of long autransporters and 21 out of 28 (75%) long bacterial signal peptides not associated with autotransporters could form an amphipathic helix. We compared this to short bacterial signal peptides associated with autotransporters and found that 34 out of 228 (15%) could form an amphipathic helix. The requirement to form an amphipathic helix was to possess nine adjacent amino acids in a helix in a window of 18 residues. In a second approach we allowed the adjacent nine e.g. polar residues to be interrupted by one e.g. nonpolar residue and vice versa. Now 68 out of 90 (76%), 25 out of 28 (89%) and 58 out of 228 (25%) could form an amphipathic helix (Fig. 2). While one must keep in mind that short sequences in general provide less amino acids to form an amphipatic helix at all, we still report a tendancy of long singal peptides to form alpha helices.

Conclusion

We present an extensive analysis of 90 long bacterial autotransporter signal sequences predicting in 86% of the sequences, a common two-domain organization. The described organization is in agreement with published experimental data and allows the identification of potential new domains *in silico* in long signal sequences. We corroborate the importance of charged residues in bacterial signal sequences and emphasize their position near the N-terminus as possible regularity. The approach highlights the relevance of charged residues in long signal sequences.



NP_052685	MNKIYSLKYSHITGGLIAVS	LSGRVSSRATG <mark>KKKHKR</mark> ILAL	CFLGLLQSSYSFA
AAC44731	MNKIYALXYCHATGGLIAVS	LASEVMERAAEGSLLALFN	LSLYGAFLSASQA
CAA11507	MNRIYSLRYSAVARGFIAVS	FARKCVH-	<mark>KSVRR</mark> LCFPVLLLIPVLFSAGSLA
AAP33781	MNRIYSLRYSAVARGFIAVS	FARKCVH-	<mark>KSVRR</mark> LCFPVLLLIPVLFSAGSLA
AAD23953	MNRVYSLRYCPVTGGLIAVS	LARRVIKETCRELT	PAICLCYSQISQA
CAA46156	MNT PUTINACI NUMUUNC		CMUCHI PUPNI LAI CI CETURI COCDURA
CAC14227			SHVSALF ANNLLALSLGSIVELSIGPVER
AAK68872	MNKIYRIIWNSALNAWVVVS	LTRN TX	RASATVATAVLATLLFATVOASA
AAG01335	MNOIYRAVWNKSLGVWAVAS	LSSGDSPGAVASASFIDRR	HRLALTAAIALALGGAGFATPLPANA
CAD12824	MNTNLYRLVFSHVRGMLVPVS	HCTVGNTFCGTT GQARSGARATSLSVAP-	RALAWALMLACTGLPLVT
CAI77662	MNKIYELKFSKELNALVAVS	LTRGCDHSTERVSERPVRTK	VRHLALKPLSAILLSLGIVSIPQSVLASG
AAM88788	MNKQCYCLIYSRTHGELRVVS	LARGCNTTAGQC	GVSRLWVTVRRAVWLLGMALFTGQASA
NP_253231	MNKSYTLVWNQATGCWNVAS	GTRRRSKSG	<mark>RG</mark> KALVVAGASLLGLFCQAPAFA
NP_252771	MNCYALVWNVSQGCWNVVS	GSRRRGRPAGARAAIASVL	ALLGATALAPAY
AAA20524	MNKIYRLKFSKHLNALVAVS	LARGC DHSTENGSEN PARMK	VRHLALKPLSAMLLSLGVTSIPQSVLA
AAB96359	MNXIY VXXNAAGH-LVACS	FARGETERAVLGSLLIVGA	
AAF 40927		TARAGANTATIQAVGILPNUIAGPAG-	FISISVISFSLSLLLGSALILTSSSATA
AAF 07520	MNRTVSLEVSAATGGLTAVS	LAKEVSCETNEELVATMI.SLA	SIIGSIG
AAK00474	MKRHLNTCYRLVWNHITGAFVVAS	LARAOGERGG	VAVALSLAAVTSLPVLA
AAK77860	MNSKLYKLIFCRRLGCLIAVG	FTRTYGRSFSSFGKKIINDNHTRAGKLSH-	LAILTGLALGTLPLLVFA
AAL18821	MNKIYSLKYSSLTGGLIAVS	LSKRVKGRTGRKLMTASVALS	VSLSALPVEA
AAL78284	MNHIYKVIFNKATGTFMAVA	ARSISTGGSCATGQVG-	SVRTLSFARVAALAVLVIGATLNGSAYAQ
CAC14203	MNKVYKVIWSHVTNTFIAVS	LATSKGKVKSFSAISSNPQPKLNSSI-	PATFLISAIALVSILAFAPSQVLA
CAC14218	MNATYELIFNEALGCLQVAS	LAE TGGGAAGGVVGAGV GAPAVDENQVPA	LGLLLRQILAVLQPAVPLLMVGAG-VLAPGLT
CAC39286	MNXIYALXYSSLTGGLIAVS	LSKRVTGRTGRRLMTVSLVLS	VTLSALPG <mark>K</mark> A
CAC92482	MNSKLYKLIFCRRLGCLIAVG	FTRSYGRAFSSKGGQAGNNQRRAVGILSR-	LAMMTGLALGIFPLLVLA
NP_274768	MNKTLYRVIFNRKRGAVVAVA	TTKREGKSCADSDSGSAHVKSVPFGTTHA-	PVC SNIFSFSLLGFSLCLAVGTANIAFA
NP_308389	MNKIYRLNWNRSRNCWSVCS	LGSRVNGRKSRAVLI	SAISLYSSLVFA
NP_519008	MNACY TVFNAVRGMLVAVE	SARSTG GROSGGOAGATAPA-	SAASAARFAVLPVVFGAWCALGLPYAVQA
NP_519896	MNAACINTVFNAARGMLVAV	SANSTGAGAGAGSGASKKAAS-	
NP_522034			
NP_807449	MNRT FULWNAATGTFTVTS	TAK SEGKK SGPRK	LAVSALVGLSSTMVSA
YP 553065	MLNSYNTVWN TTRTYAAAS	VT SRGA GAS	VRGSLVAASAGLLGALAFSOPAAA
ZP 00033562	MNATY SVWNESTGTWVAAS	HASARGKKSSAK	TSSTRAVVGALGLAAGLYGADAFA
ZP_00041732	MNKDLYRLIYNRALRLWQVAS	LATAPGGTPGPS <mark>PTAQ</mark>	RPARACLHPIPFALWLSLGWVSITGMATA
ZP_00088699	MNRIFNIVWNRSLGGWTVAS	HARQRGRPGGACRALASVV	<mark>ALAPA</mark> CAFA
ZP 00122019	MNKIFKTKYDVTTGETKVVS	LA <mark>K</mark> NCPAASGVSCAS <mark>SVGVGQPK</mark> CGVFFG-	GMLGAFKILPLALLISGVLSPLGYAAA
-			
ZP_00123697	MNKIFKTKYDVTTGQTKVVS	LANNRQVASRVE-GASVGVGQPKCGVF	LGMFKVLPLALLMSGLLSSAAY
ZP_00123697 ZP_00132251	MNKIFKTKYDVTTGQTKVVS MNKIFKTKYDVTTGQCKAVS	LANNRQVASKVE-GASVGVGOPNCGVF LASNRQIASSSEKKPRCGANLK-	LGMFKVLPLALLMSGLLSSAAY
ZP_00123697 ZP_00132251 AAC74583	MN IF TTY DVTTGQT VVS MN IFTTY DVTTGQC AVS MN IY VIWNCTLQVFQACS	LANNRQVASEVE-GASVGVGOPRCGVF LASNRQTASSSKEPECGANLE- LTERACETSTVNLERSSGLTT	EGMFKVLPLALLMSGLLSSAAY RTSLSENILFNMLISGLVUFAYPAWA RFSELTLGVLLALSGBASE
ZP_00123697 ZP_00132251 AAC74583 AAD41751 AAF43424	MN IF T Y VTTGOT VVS MN IF T Y VTTGOC AVS MN IY VIWNCLQVFQACS MN VYNTWN STGTWVTS	LANN QVAS VE-GASVGVGOPNCGVF- LASN QIASSSSKFPCGANL LTRAGETSTVNLKASSGLT LTRAGE PE	EGMFKVLPLALLMSGLLSSAAY KTSLSENILFNMLISGLVLFAYPAWA KFSKLTLGVLLALSGASAG QIKHTVLAGLIAGLLMPSMPALA
ZP_00123697 ZP_00132251 AAC74583 AAD41751 AAF43424 AAG53941	MNKIFITY VUTTGOTYVS MNKIFITY VUTTGOCHAVS MNKIY VIWNCTLQVFQACS MNKVYNTVWNKISTGTWVVTS MT IWNTSYLLVWNHITGTLVVAS	LANN QUAS VE-GASVGUGOPN CGVF- LASN QIASSSEKFPC CGANLE- LTREAGE TSTVNLERSSGLTT LTRE GGLEPE LAESE GEETG	EGMFKVLPLALLMSGLLSSAAY KTSLSENILFNMLISGLVLFAYPAWA KFSLTLGVLLALSGSASG QIKHTVLAGLIAGLLMPSMPALA
ZP_00123697 ZP_00132251 AAC74583 AAD41751 AAF43424 AAG53941 Q84GK0	MNKIFITY VUTTGOTYVS MNKIFITY VUTTGOCHAVS MNRIY VIWNCTLQVFQACS MNKVYNTVWNKSTGTWVVTS MTIWNTSYRLVWNHITGTLVVAS MNKNIY VVWSLVRGAWVAG MNKVFSLYYFRAKGFIAVS	LANN QUAS VE-GASVGUGOPN CGVF- LASN QIASSSEKFPC CGANLE- LTREAGE TSTVNLERSSGLTT LTREGLEPE LASSGKETG	EGMFKVLPLALLMSGLLSSAAY KTSLSENILFNMLISGLVLFAYPAWA KTSLSENILFNMLISGLVLALSGSASG QIKHTVLAGLIAGLLMPSMPALA
ZP_00123697 ZP_00132251 AAC74583 AAD41751 AAF43424 AAG53941 Q84GK0 Q7BCK4	MNKIFITY VUTTGOTYVS MNKIFITY VUTTGOCHAVS MNRIYUWNCTLQVFQACS MNKVYNTVWNSTGTWVTS MTIWNSYRLVWNHITGTLVVAS MNKNIYVVVSLVRGAWVAG MNKVFILYSFLAKGFLAVS MNQIHIFFCNMTQCSQGGAG	LANN QVAS VE-GASVGVGOPN CGVF- LASN QIASSSEKFPCCGANLE- LTRAGE TSTVNLERSSGLTT LTRAGE TSTVNLERSSGLTT LARS GERTG	
ZP_00123697 ZP_00132251 AAC74583 AAD41751 AAF43424 AAG53941 Q84GK0 Q7BCK4 Q8CWC7	MNKIFITKY VTTGOTYVS MNKIFITYV VTTGOCHAVS MNKIY VIWNCTLQVFQACS MNKVYNTVWN STGTWVTS MTIUNTSYLJWWNI ITGTLVAS MNKNIYVVVSLVGAWVVAG MNKVFSLYSFLAKGFIAVS MNQIHFFCNMTQCSQGAG MNKVYSLYYPVTGGLIVVS	LANN QUAS VS -GASVGVGOPK CEVF LASN QIASSSKFP CGANL LTR GGL P LARS GC P LARS GC F LARS GC CL SASSI LAR VSV GL SASSIIISP LAR VSV GL SASSIISP LARK VSV GL SSFFVVGAS LASK VIK TC R LT ILLAGI	
2P_00123697 2P_00132251 AAC74583 AAD41751 AAF43424 AAG53941 Q846K0 Q7BCK4 Q8CWC7 YP_840389	MNN IF T Y Y VTTGOTYVS MNN IF T Y Y VTTGOCAVS MNN IY VIWNCTLOYCASS MALEN VYNTWN STGTWVVS MIN IN VYNTWN STGTWVVS MNNN VYNTWN STGTWVVS MNNN Y VYNSLY GAWVAG MNNFSL YBFLAX GFIAVS MNN VFSL YBFLAX GFIAVS MNN VYSL YEFUAGAUVACG MNN VYSL YEFUAGAUVACG MNN WALL	LANN QVAS VS -GASVGVGOPK CEVF- LASN QIASSSKFP CGANL TH AGTSTVNL KSSGLT LARS GC F WASACK SSSPRS QNRQ LARK VSVK CL SASSIIISP LPTVK KTC LSFSPFVVGAS LASKVIK TC RLM ILLAGI	
2P_00123697 2P_00132251 AAC74583 AAD41751 AAF43424 AAG53941 Q84GK0 Q7BCK4 Q8CWC7 YP_840389 YP_916859	MNN IF T Y VUTTGOTYVS MNK IF T Y VUTTGQC AVS MNK IY VUWNCTLQVFQACS MNK STGTWVNS MIN VFSL YELL GFLAVS MIN VSL YCNTQCSQGAG MIN VYSL YCNTQCSQGAG MIN VSL MMM	LANN QVAS VI -GASVGVGOPR CEVF- LASN QIASSS KRP CGANL TH AGTSTVNL KNSSGLT LARS GC P MARACHSSSPR QNRQ LARS VSV CL SASSIIISP LPTV KTCL SFSPFVGAS LASRVIK TC RLFSPFVVGAS	
ZP_00123697 ZP_00132251 AAC74583 AAD41751 AAF43424 AAG53941 Q84GK0 Q7BCK4 Q8CWC7 YP_840389 YP_916859 YP_001110606	MNKIFIT Y UVTGOTVVS MNKIFIT Y UVTGOCAVS MNKIY UIWNCTLQVFQACS MN VYNTWNKSTGTWVVS MNKNIYNVVSLVRGAWVAG MNKVFSL YSFLAKGFIAVS MNKVFSL YSFLAKGFIAVS MNKVYSL YCPVGGLIVVS MNKVYSL YCPVGGLIVVS MNKVYSL YCPVGGLIVVS MVT	LANN QVAS VE -GASVGVGOPR CGVF LASN QIASSS KRP CGANL LTH AGTSTVNL KNSSGLTT LARS GGL P LARS GGL T LARS VSVC LLSASSIIISP	
ZP_00123697 ZP_00132251 AAC74583 AAD41751 AAF43424 AAG53941 Q84GK0 Q7BCK4 Q8CWC7 YP_840389 YP_916859 YP_001110606 YP_001161762	MNN IF T Y Y UTTGOTYVS MNK IF T Y Y UTTGOCAVS MNK IY VIWNCTLQVFQACS MNFN YYNTWN STGTWVVTS MIN WITSYRLYWN STGTWVVTS MNKNIY VVWSLVRGAWVAG MNK VFSL YSPLAKGFIAVS MNK VFSL YSPLAKGFIAVS MNK VFSL YSPLAKGFIAVS MNK VYSL YCVTGGLIVVS MNK VYSL YCVTGGLIVS MNF VYSL YCVTGGLIVS MVWRM	LANN QVAS VE -GASVGVGOPR CGVF LASN QIASSSKRP CGANL TH AGTSTVNLK SSGLT LARS GGLP	
2P_00123697 2P_00132251 AAC74583 AAD41751 AAF43424 AAG33941 Q84GK0 Q7BCK4 Q8CWC7 YP_840389 YP_916859 YP_001110606 YP_001161762 YP_001371348	MNN IF TY Y UTTGOTYVS MNK IF TY Y UTTGOCAVS MNK IY UIWNCTLQVFQACS MT IWNTSY LVWN STGTWVVS MT IWNTSY LVWN ITGTLVVS MNKNIY EVWNLVRGAWVAG MNK VFSL YSFLANGFIAVS MNQ IHIFFCNMTQCSQGGAG MNK VYSL YCPVTGGLIVVS MMKM	LANN QVAS VE -GASVGVGOPR CGVF LASN QIASSS KRP CGANL LTN AGTSTVNLKNSSGLT LTN GGL P LASS GC TG WARACRISSSPREQNEQ	
2P_00123697 2P_00132251 AAC74583 AAD41751 AAF43424 AAG53941 Q84GK0 Q7BCK4 Q84GK0 Q7BCK4 Q8400 Q7BCK4 Q8007 YP_840389 YP_916859 YP_00110606 YP_001161762 YP_001371348 YP_001373319	MNKIFITY VUTTGOTYVVS MNKIFITY VUTTGOCAVS MNKIY VUWNCTLOVFOACS MNVYNTVNN STGTWVVTS MTIWNTSYLVWNITGTLVVS MNKNIYSVVWNLVRGAWVAG MNKVFSLIYSFLANGFIAVS MNQIHIFFCNMTQCSQGGAG MNKVYSLYCFVTGGLIVVS M MKMS MKNS	LANN QUAS VE GASVGUGOPA CCVF LASN QIASSS	
2P_00123697 2P_00132251 AAC74583 AAD41751 AAF43424 AAG53941 Q840K0 Q7BCK4 Q8CWC7 YP_840389 YP_916859 YP_001161762 YP_001371348 YP_001373319 YP_001683300	MNN IF T Y V VTTGOT VVS MNK IF T Y V VTTGOC AVS MNK IF T Y V VTTGOC AVS MNK IF T Y V VTTGOC AVS MNK JF T Y V WNCTLOVCAC MNK VSL YSTGA GRIVVAG MNK VFSL YSFLAX GRIVVAG MNK VFSL YSFLAX GRIVVS MNK VFSL YSFLAX GRIVVS MNK VFSL YSFLAX GRIVVS MNK VFSL YSFLAX GRIVVS MNK V	LANN QVAS VS -GASVGVGOPK CEVF- LASN QIASSS	
2P_00123697 2P_00132251 AAC74583 AAD41751 AAF43424 AAG53941 Q846K0 Q7BCK4 Q8CWC7 YP_840389 YP_916859 YP_001161762 YP_001371348 YP_001373319 YP_001683300 YP_001766852	MNN IF T Y V VTTGOT VVS MNN IF T Y V VTTGOC AVS MNN IY V WNCTLQVFQACS MN VYNTWN STGTWVTS MNNN VYNTWN STGTWVVS MNNNIY VVWSLVEGAWVVAG MNYFSL YBFLAX GFIAVS MNY VFSL YSFLAX GFIAVS MNY VFSL YSFLAX GFIAVS MNY VFSL YSFLAX GFIAVS MNY VRNFFCNMYQCSQGAG MNY VRNF MNS SERVICE MNS V	LANN QVAS VS -GASVGVGOPACGVF LASN QIASSS	
2P_00123697 2P_00132251 AAC74583 AAD41751 AAF43424 AAG53941 Q84GK0 Q7BCK4 Q8CWC7 YP_840389 YP_916859 YP_001110606 YP_001371348 YP_001371348 YP_001373319 YP_001948780 YP_001766852 YP_001990882	MNN IF T Y Y UTTGOTYVS MNK IF T Y Y UTTGQC AVS MNK IY VUWNCTLQVFQACS MNS STGTWVNS MNKN VYNTWN STGTWVVS MNKNY VYNTUNS MNKNY VYNTUNS MNKNY VYNTUNS MNKSIL YSFLALGFIAVS MNK VFSL YSFLALGFIAVS MNK VYSL YCPVTGGLIVVS MNK VYSL YCPVTGGLIVS MNK VYSL YCPVTGGLIVS MNSNKSP	LANN QVAS VS -GASVGVGOPA CEVF- LASN QIASSS	
2P_00123697 2P_00132251 AAC74583 AAD41751 AAF43424 AAG53941 Q84GK0 Q7BCK4 Q8CWC7 YP_840389 YP_916859 YP_00116606 YP_001371349 YP_001371349 YP_001373319 YP_001948780 YP_001683300 YP_001990882 YP_001993037	MNN IF T Y VUTTGOTYVS MNK IF T Y VUTTGQCAVS MNK IY VUWNCLQYPQACS MNF JY VUWNTQVNS STGTWVTS MNKN VYNTWN STGTWVTS MNKNY VUNTQURALY MNKNY VUNTUN STGTWVTS MNKNY VUNTUN STGTWVTS MNKNY VUNTUNS MNK VFSL YSFLARGFIAVS MNK VYSL YCONGGLIVVS MNK VYSL YCONGGLIVS MNKSPK MK SNKSPK MK S	LANN QVAS VS -GASVGVGOPR CEVF- LASN QIASSS	
2P_00123697 2P_00132251 AAC74583 AAD41751 AAF43424 AAG53941 Q84GK0 Q7BCK4 Q8CWC7 YP_840389 YP_916859 YP_001110606 YP_001371349 YP_001373319 YP_001948780 YP_001683300 YP_001766852 YP_001993037 YP_001993303	MNN IF T Y VUTTGOT VVS MNK IF T Y VUTTGQC AVS MNK IY VUWNCTLQVFQACS MNKN VYNTWN STGTWVTS MNKN VYNTWN STGTWVTS MNKNIYRVVWSLVRGAWV4G MNKFSL YSFLAKGFIAVS MNK VFSL YSFLAKGFIAVS MNK SNKSFR MKSNKSFR MKS	LANN QVAS VS -GASVGVGOPR COVF LASN QIASSS	
ZP_00123697 ZP_00132251 AAC74583 AAD41751 AAF43424 AAG53941 Q84GK0 Q7BCK4 Q8CWC7 YP_840389 YP_916859 YP_001110606 YP_001371349 YP_001373349 YP_001948780 YP_001683300 YP_001948780 YP_001993037 YP_001993508 Q0AL61	MNN IF T Y VUTTGOT VVS MNK IF T Y VUTTGQC AVS MNK IY VUWNCTLQVPQACS MN WITSYRUWN STGTWVTS MNKN VYNTWN STGTWVTS MNKNIY VVWSLVRGWVVAG MNY VFSL YSPLANGFIAVS MNQ IHIFFCNNTQCSQGAG MNY VYSL YCVTGLIVVS MNG IHIFFCNNTQCSQGAG MNYYSL YCVTGLIVS MASNSPF MAS	LANN QUAS VS -GASVGUGOPA CCVF- LASN QIASSS KP CGANL LTH GGL P LARS CGL P LARS CGL P LARS CGL P LARS CGL P LARS CGL P LARS VSV GL SASSIIISP- LARS VSV GL SASSIISP- LARS VSV GL SASSIISP- SASSIISP GL SASSIISP- SC SASSIISP GL SASSIISP -SSG WA MIGMANN SSR G -SSG VA MIGMANN SSR G -SSG VA MIGMANN SSR G -NS AT AGGA ARR ILLSS- 	
2P_00123697 2P_00132251 AAC74583 AAD41751 AAF43424 AAG53941 Q840K0 Q7BCK4 Q8CWC7 YP_840389 YP_001161762 YP_001371348 YP_001746852 YP_001948780 YP_001968320 YP_001990882 YP_001990882 YP_001993037 YP_001993082 Q0AL61 Q0B7F7	MNN IF T Y V VTTGOT VVS MNK IF T Y V VTTGOC AVS MNK IF T Y V VTTGOC AVS MNK IY V WNCTLQVCQCAVS MNK JN VYWVNN STGTWVVTS MNK VFSL YBELAN GEIAVS MNK VYSL YCPVTGGLIVVS MNK VYSL YTCPTGGLIVVS MNK VYSL YCPTGGLIVVS MKS S	LANN QUAS VS -GASVGUGOPA CEVF- LASN QIASSS	
2P_00123697 2P_00132251 AAC74583 AAD41751 AAF43424 AAG53941 Q840K0 Q7BCK4 Q8CWC7 YP_840389 YP_916859 YP_001161762 YP_001371348 YP_001371348 YP_001643300 YP_001766852 YP_001990882 YP_001990882 YP_001993037 YP_0019930882 YP_001994882 YP_001994882 YP_00199488 YP_00199488 YP_00199488 YP_00199488 YP_00199488 YP_00199488 YP_00199488 YP_0019948 YP_00199488 YP_0019948	MNN IF TY Y UTTGOTYVS MNK IF TY Y UTTGOCAVS MNK IF TY Y UTTGOCAVS MNK IY VIWNCTLQYPQACS MNK JN VYNTWN STGTWVVS MNK VFSL YSFLAX GRIAVS MNK S	LANN QUAS VS -GASVGVGOPR CEVF- LASN QIASSS	
ZP_00123697 ZP_00132251 AAC74583 AAD41751 AAF43424 AAG53941 Q846K0 Q7BCK4 Q8CWC7 YP_840389 YP_916859 YP_001161762 YP_001371348 YP_001371348 YP_00168320 YP_0016852 YP_001990882 YP_001990882 YP_001993037 YP_001993037 YP_001993508 Q0AL61 Q0B7F7 Q0PAN9 Q1MKD9	MNN IF T Y V VTTGOT VVS MNK IF T Y V VTTGOC AVS MNK IF T Y V VTTGOC AVS MNK IF T Y V VTTGOC AVS MNK IY V VIWNCTLQYPQACS MNK VYSL Y STGTWVNS MNK VFSL YBFLAK GFIAVS MNK S	LANN QVAS VS -GASVGVGOPY CCVF LASN QIASSS	
ZP_00123697 ZP_00132251 AAC74583 AAD41751 AAF43424 AAG53941 Q84GK0 Q7BCK4 Q8CWC7 YP_840389 YP_916859 YP_001110606 YP_001371348 YP_001371348 YP_001371348 YP_001373319 YP_001948780 YP_0017668300 YP_0017668300 YP_001990882 YP_001990882 YP_001990882 YP_0019903037 YP_001993508 Q0AL61 Q0B7F7 Q0PAN9 Q1MKD9 Q21460	MNN IF T Y V VTTGOT VVS MNK IF T Y V VTTGQC AVS MNK IF T Y V VTTGQC AVS MNK IF T Y V VTTGQC AVS MNK IY V UWNCTLQYPQACS MNK VVYNVN STGTWVVTS MNK VYSLYSTAN GRIAVS MNK VFSLYSFLALGEIAVS MNK VFSLYSFLALGEIAVS MNK VFSLYSFLALGEIAVS MNK VFSLYSFLALGEIAVS MNK V	LANN QVAS VS -GASVGVGOPY COVF LASN QIASSS	
ZP_00123697 ZP_00132251 AAC74583 AAD41751 AAF43424 AAG53941 Q84GK0 Q7BCK4 Q8CWC7 YP_840389 YP_916859 YP_001110606 YP_00171348 YP_001371349 YP_001371349 YP_001371349 YP_001948780 YP_001948780 YP_001990882 YP_001990882 YP_001990882 YP_001990877 Q019903508 Q0AL61 Q0B7F7 Q0PAN9 Q1MKD9 Q20XG2 Q214C9 Q214C9 Q214C9	MNN IF T Y VUTTGOT VVS MNN IF T Y VUTTGOC AVS MNN VY VUNN STGTWVTS MNN VSI Y SFLAY GFLAVS MNN VFSI YSFLAY GFLAVS MNN VFSI YSFLAY GFLAVS MNN VFSI YSFLAY GFLAVS MNN VFSI YSFLAY GFLAVS MNN VYSI YCVTGGLIVVS MMNM MNN VYSI YCVTGGLIVVS M	LANN QVAS VS -GASVGVGOP COVF LASN QIASSS	
2P_00123697 2P_00132251 AAC74583 AAD41751 AAF43424 AAG53941 Q84GK0 Q7BCK4 Q8CWC7 YP_840389 YP_916859 YP_001110606 YP_001371348 YP_001371348 YP_001373319 YP_001948780 YP_001948780 YP_001948780 YP_001993337 YP_001993508 Q0AL61 Q0B7F7 Q0PAN9 Q1MKD9 Q20XG2 Q214C9 Q216F9 O2JI0W4	MNN I F T Y V VTTGOT VVS MNN I F T Y V VTTGQC AVS MNN I Y V WNCTLQVFQACS MNN I Y V WNN STGTWVVS MNN V Y V WNN STGTWVVS MNN V SL YSFLARGFIAVS MNN V FSL YSFLARGFIAVS MNN V FSL YSFLARGFIAVS MNN V WNFW-WSLVRGAWVAG MNN V FSL YSFLARGFIAVS MNN V SL YSFLARGFIAVS MNN V WNFW-GOLIVVS MMN V	LANN QUAS VS -GASVGUGOP CCVF LASN QIASSS	
2P_00123697 2P_00132251 AAC74583 AAD41751 AAF43424 AAG53941 Q840K0 Q7BCK4 Q8CWC7 YP_840389 YP_001161762 YP_001371348 YP_001940780 YP_0019483300 YP_001990882 YP_001990882 YP_001990882 YP_001990882 YP_001993037 YP_001993037 YP_001993082 Q0AL61 Q0B7F7 Q0PAN9 Q1MKD9 Q20X62 Q214C9 Q216F9 Q210K4 O48237	MNN IF T Y VUTTGOT VVS MNK IF T Y VUTTGOC AVS MNK IF T Y VUTTGOC AVS MNK IF T Y VUTTGOC AVS MNK IY VUNNCLOYPOACS MNK VUTVNN STGTWVVS MNK VVYNTVNN STGTWVVS MNKNY VUNVVSLVRGAVVAG MNY VFSL YSPLAKGFIAVS MNV VFSL YSPLAKGFIAVS MNV VFSL YSPLAKGFIAVS MNV VYSL YCOVGCSQGAG MNV V	LANN QUAS VE -GASVGUGOPA CEVE- LASN QIASSS	
2P_00123697 2P_00132251 AAC74583 AAD41751 AAF43424 AAG53941 Q840K0 Q7BCK4 Q8CWC7 YP_840389 YP_916859 YP_001161762 YP_001371348 YP_001373319 YP_001948780 YP_001948780 YP_001990882 YP_001990882 YP_001990882 YP_001993037 YP_001993037 YP_001993082 YP_001993082 YP_001993082 Q0AL61 Q0B7F7 Q0PAN9 Q1MKD9 Q20XG2 Q214C9 Q216F9 Q2J0N4 Q48237 CAR56027	MNN IF T Y Y UTTGOT VVS MNN IF T Y Y UTTGOC AVS MNN IF T Y Y UTTGOC AVS MNN IF T Y Y UTTGOC AVS MNN IY VIWNCTLQVFQCS MNS VSL YSTGWNN STGTWVVS MNYFSL YSFLAN GFIAVS MNYFSL YSFLAN GFIAVS MNYMRM	LANN QUAS VS -GASVGUGOPA CEVE- LASN QIASSS	
2P_00123697 2P_00132251 AAC74583 AAD41751 AAF43424 AAG53941 Q840K0 Q7BCK4 Q8CWC7 YP_840389 YP_916859 YP_001161762 YP_001371348 YP_00163703 YP_00163730 YP_001766852 YP_001990882 YP_001990882 YP_001993037 YP_001993037 YP_001993037 YP_001993037 YP_001993037 YP_001993037 YP_001993037 YP_001993037 YP_001993037 YP_001993037 YP_001993508 Q0AL61 Q0B7F7 Q0PAN9 Q1MKD9 Q216F9 Q216F9 Q210N4 Q48237 CAR56027 Q132B9	MNN IF T Y Y UTTGOT VVS MNN IF T Y Y UTTGOC AVS MNN IF T Y Y UTTGOC AVS MNN IY V UWNCTLQVPQACS MNN JN Y UWNN STGTWVVS MIN J WINTSY LVWNN STGTWVVS MNNNIY VVWSLVRGAWVVAG MNYFSL YSFLAX GFIAVS MNY	LANN QVAS VS -GASVGVGOPY CCVF- LASN QIASSS	
2P_00123697 2P_00132251 AAC74583 AAD41751 AAF43424 AAG53941 Q84GK0 Q7BCK4 Q8CWC7 YP_840389 YP_916859 YP_001161762 YP_001371348 YP_001371348 YP_00168320 YP_001990882 YP_001990882 YP_001990882 YP_001993037 YP_001993037 YP_001993508 Q0AL61 Q0B7F7 Q0PAN9 Q1MKD9 Q21AC9 Q214C9 Q21F9 Q210N4 Q48237 CAR56027 Q132B9 Q137J7	MNN I F T Y Y UTTGOT VVS MNN I F T Y Y UTTGOC AVS MNN I Y VIWNCTLQYPQACS MNN J Y VIWNCTLQYPQACS MNN J Y VIWNN STGTWVVS MIN J WINTYY LVWNN STGTWVVS MNN NIY VVWSLVEGAWVVAG MNN VFSL YBELAL GEIAVS MNN VFSL YBELAL GEIAVS MNN VFSL YBELAL GEIAVS MNN VFSL YBELAL GEIAVS MNN V	LANN QVAS VS -GASVGVGOPY CCVF- LASN QIASSS	
2P_00123697 2P_00132251 AAC74583 AAD41751 AAF43424 AAG53941 Q84GK0 Q7BCK4 Q8CWC7 YP_840389 YP_916859 YP_001110606 YP_001371348 YP_001371348 YP_001371348 YP_001948780 YP_0017668320 YP_001990882 YP_001990882 YP_001990882 YP_001990882 YP_001990882 YP_001990882 YP_001990882 YP_001993508 Q0AL61 Q0B7F7 Q0PAN9 Q1MKD9 Q20X62 Q214C9 Q216F9 Q220N4 Q48237 CAR56027 Q132B9 Q13737 Q2KVL9	MNN I FT Y Y UTTGOT VVS MNN I FT Y Y UTTGOC AVS MNN I Y VIWNCTLQVFQACS MNN I Y VIWNN STGTWVVS MIN I Y VIWNN STGTWVVS MNN NIY VVWN STGTWVVS MNN FSL YELAN GRIAVS MNN V Y Y TY MNN Y Y Y TY MNN Y Y TT MNN S NN SAP MANS NN SAP MAY TY ST IQVLLAGGS VC MQG	LANN QVAS VE -GASVGVGOPY CEVE- LASN QIASSS	
2P_00123697 2P_00132251 AAC74583 AAD41751 AAF43424 AAG53941 Q84GK0 Q7BCK4 Q8CWC7 YP_840389 YP_916859 YP_001110606 YP_00171348 YP_00171348 YP_00171348 YP_00173319 YP_001948780 YP_001948780 YP_00199882 YP_001990882 YP_001990882 YP_001990882 YP_001990882 YP_001990882 YP_001990882 YP_001990882 YP_001990882 YP_001990852 Q0DAF61 Q0B7F7 Q0PAN9 Q1MKD9 Q20XC2 Q216F9 Q2J0N4 Q48237 CAR56027 Q137J7 Q2KVL9 Q4ZQ08	MNN IF T Y VUTTGOT VVS MNN IF T Y VUTTGOC AVS MNN IF T Y VUTTGOC AVS MNN IF T Y VUTTGOC AVS MNN IY VUNNCLOYPOACS MNN VYSL YSGTWVTS MNN VFSL YSFLAK GFLAVS MNN VFSL YSFLAK GFLAVS MNN VFSL YSFLAK GFLAVS MNN VYL YSL YCPVGGGLIVVS MNN VYL YSL YCPVGGGLIVS MMMM MNN VYL YSL YCPVGGGLIVS M	LANN QUAS VI -GASVGUOPY CCVF LASN QIASSS	
2P_00123697 2P_00132251 AAC74583 AAD41751 AAF43424 AAG53941 Q84GK0 Q7BCK4 Q8CWC7 YP_840389 YP_916859 YP_001110606 YP_001371348 YP_001371319 YP_001948780 YP_00171648300 YP_00176852 YP_001990852 YP_001990852 YP_001990852 VP_001990852 VP_001990852 VP_001993508 Q0AL61 Q0B7F7 Q0PAN9 Q1MKD9 Q20XG2 Q214C9 Q22004 Q48237 CAR56027 Q132B9 Q137J7 Q2KVL9 Q42Q08 Q6N2N9	MNN I F T Y V VTTGOT VVS MNN I F T Y V VTTGQC AVS MNN I Y V WNCTLQVFQACS MNN I Y V WNCTLQVFQACS MNN V Y V WNN STGTWVVS MNN V SI Y SFLAY GFLAVS MNN V FSL Y SFLAY GFLAVS MNN V Y V TY VYSL Y COVEGGLAVS MNN V Y V TY FRAFC MNN S MNSF MNN S MNSF MNN S MNSF MNN S MNN SF MNN S MNN SF MASS	LANN QUAS VE -GASVGUGOP CCVF LASN QIASSS	
2P_00123697 2P_00132251 AAC74583 AAD41751 AAF43424 AAG53941 Q840K0 Q7BCK4 Q8CWC7 YP_840389 YP_001161762 YP_001371348 YP_001371348 YP_001373319 YP_001948320 YP_001968320 YP_001990882 YP_001990882 YP_001990882 YP_001990882 YP_001990882 YP_001990882 YP_001990882 YP_001993037 YP_001993037 YP_001993037 YP_001993082 Q0AL61 Q0B7F7 Q0PAN9 Q1MKD9 Q20Xc2 Q214C9 Q216F9 Q216F9 Q216F9 Q216F9 Q216F9 Q216F9 Q216F9 Q216F9 Q216F9 Q216F9 Q216F9 Q216F9 Q216F9 Q216F9 Q216F9 Q216F9 Q216F9 Q216F9 Q2177 Q132B9 Q13777 Q2KVL9 Q4ZQ08 Q6N274 Q6N274 Q6N274 Q6N274 Q6N275 Q2175 Q2075 Q2175 Q	MNN IF T Y Y UTTGOT VVS MNN IF T Y Y UTTGOC AVS MNN IF T Y Y UTTGOC AVS MNN IF T Y Y UTTGOC AVS MNN IY VUWNCTLOYCACS MNN YFSL YSETAN GENAVYS MNYFSL YSELAN GENAVS MNYFSL YSELAN GENAVS MNY	LANN QUAS VI -GASVGUGOP CCVF LASN QIASSS	
2P_00123697 2P_00132251 AAC74583 AAD41751 AAF43424 AAG53941 Q840K0 Q7BCK4 Q8CWC7 YP_840389 YP_916859 YP_001161762 YP_001371348 YP_001373319 YP_001948780 YP_001948780 YP_001990882 YP_001990882 YP_001990882 YP_001990882 YP_001993037 YP_001993037 YP_001993082 YP_001993082 YP_001993082 YP_001993082 Q0AL61 Q0B7F7 Q0PAN9 Q1MKD9 Q20XG2 Q214C9 Q216F9 Q2J0N4 Q48237 CAR56027 Q132B9 Q13737 Q2KVL9 Q42Q08 Q6N224 Q6N224 Q6N324 Q6N324 Q6N324	MNN IF T Y V VTTGOT VVS MNN IF T Y V VTTGOC AVS MNN IF T Y V WTTGOC AVS MNN IF T Y V WTTGOC AVS MNN IY V WINCTLOYCACS MNN VYNTWN STGTWVTS MNNNIY VVWSLVFGAWVAG MNYFSL YBELAL GFIAVS MNY	LANN QUAS VI -GASVGUGOP CEVE- LASN QIASSS	

Figure 2. Multiple sequence alignment of 79 NtraC-organized autotransporter signal sequences (Matlab R2009a, Bioinformatics Toolbox Version 3.3). Red: negatively charged residues. Blue: positivly charged residues. Green: residues potentially part of an amphipatic helix.

Acknowledgement

We are most grateful to Dr. H. Bernstein for critical and constructive comments on the manuscript.

Disclosures

The authors report no conflicts of interest.

References

- Henderson IR, Navarro-Garcia F, Desvaux M, Fernandez RC, Ala'Aldeen D. Type V protein secretion pathway: the autotransporter story. *Microbiol Mol Biol Rev.* 2004;68:692–744.
- Yen YT, Stathopoulos C. Identification of autotransporter proteins secreted by type V secretion systems in gram-negative bacteria. *Methods Mol Biol.* 2007;390:33–46.
- Dautin N, Bernstein HD. Protein secretion in gram-negative bacteria via the autotransporter pathway. *Annu Rev Microbiol*. 2007;61:89–112.
- Jacob-Dubuisson F, Fernandez R, Coutte L. Protein secretion through autotransporter and two-partner pathways. *Biochim Biophys Acta*. 2004;1694: 235–57.
- Wells TJ, Tree JJ, Ulett GC, Schembri MA. Autotransporter proteins: novel targets at the bacterial cell surface. *FEMS Microbiol Lett.* 2007;274: 163–72.
- von Heijne G. Signal sequences. The limits of variation. J Mol Biol. 1985; 184:99–105.
- Clérico EM, Szymańska A, Gierasch LM. Exploring the interactions between signal sequences and E. coli SRP by two distinct and complementary crosslinking methods. *Biopolymers*. 2009;92:201–11.
- Hiss JA, Resch E, Schreiner A, Meissner M, Starzinski-Powitz A, Schneider G. Domain organization of long signal peptides of single-pass integral membrane proteins reveals multiple functional capacity. *PLoS ONE*. 2008;3:e2767.
- Desvaux M, Cooper LM, Filenko NA, et al. The unusual extended signal peptide region of the type V secretion system is phylogenetically restricted. *FEMS Microbiol Lett.* 2006;264:22–30.
- Szabady RL, Peterson JH, Skillman KM, Bernstein HD. An unusual signal peptide facilitates late steps in the biogenesis of a bacterial autotransporter. *Proc Natl Acad Sci U S A*. 2005;102:221–6.
- Jacob-Dubuisson F, Locht C, Antoine R. Two-partner secretion in Gramnegative bacteria: a thrifty, specific pathway for large virulence proteins. *Mol Microbiol*. 2001;40:306–13.
- Wu CH, Apweiler R, Bairoch A, et al. The Universal Protein Resource (Uni-ProtKB): an expanding universe of protein information. *Nucleic Acids Res.* 2006;34:D187–91.



- Bendtsen JD, Nielsen H, Widdick D, Palmer T, Brunak S. Prediction of twin-arginine signal peptides. *BMC Bioinformatics*. 2005;6:167–76.
- Emanuelsson O, Nielsen H, Brunak S, von Heijne G. Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J Mol Biol.* 2000;300:1005–16.
- Emanuelsson O, von Heijne G, Schneider G. Analysis and prediction of mitochondrial targeting peptides. Meth. *Cell Biol.* 2001;65:175–87.
- Bendtsen JD, Nielsen H, von Heijne G, Brunak S. Improved prediction of signal peptides: SignalP 3.0. J Mol Biol. 2004;340:783–95.
- Brunder W, Schmidt H, Karch H. EspP, a novel extracellular serine protease of enterohaemorrhagic Escherichia coli O157:H7 cleaves human coagulation factor V. *Mol Microbiol.* 1997;24:767–78.
- Brockmeyer J, Spelten S, Kuczius T, Bielaszewska M, Karch H. Structure and function relationship of the autotransport and proteolytic activity of EspP from Shiga toxin-producing Escherichia coli. *PLoS One.* 2009; 4:e6100.
- Peterson JH, Woolhead CA, Bernstein HD. Basic amino acids in a distinct subset of signal peptides promote interaction with the signal recognition particle. *J Biol Chem.* 2003;278:46155–62.
- Peterson JH, Szabady RL, Bernstein HD. An unusual signal peptide extension inhibits the binding of bacterial presecretory proteins to the signal recognition particle, trigger factor, and the SecYEG complex. *J Biol Chem.* 2006;281:9038–48.
- Hammen PK, Weiner H. Mitochondrial leader sequences: structural similarities and sequence differences. J Exp Zool. 1998;282:280–3.
- Izard JW, Kendall DA. Signal peptides: exquisitely designed transport promoters. *Mol Microbiol*. 1994;13:765–73.
- 24. Inouye S, Soberon X, Franceschini T, Nakamura K, Itakura K, Inouye M. Role of positive charge on the amino-terminal region of the signal peptide in protein secretion across the membrane. *Proc Natl Acad Sci U S A*. 1982;79:3438–41.
- Vlasuk GP, Inouye S, Ito H, Itakura K, Inouye M. Effects of the complete removal of basic amino acid residues from the signal peptide on secretion of lipoprotein in Escherichia coli. *J Biol Chem.* 1983;258:7141–8.
- Puziss JW, Fikes JD, Bassford PJ Jr. Analysis of mutational alterations in the hydrophilic segment of the maltose-binding protein signal peptide. *J Bacteriol.* 1989;171:2303–11.
- Dierstein R, Wickner W. Requirements for substrate recognition by bacterial leader peptidase. *EMBO J.* 1986;5:427–31.
- Hiss JA. Domänen Architektur langer Signalpeptide in silico und in vitro. PhD thesis, University of Frankfurt, Germany, 2008.



Table S1. 28 long signal sequences not associated with autotransporters.

>uniprot_features AAUA_ALCFA_1 SIGNAL: Tat-type signal. MRWLDKFGESLSRSVAHKTSRRSVLRSVGKLMVGSAFVLPVLPVARA >uniprot features ABF2 STRCX 1 SIGNAL: Tat-type signal. MCTREAVRMSREHDLPEIPSRRLLLKGAAAAGALTAVPGVAHA >uniprot_features CHOD_BREST_1 SIGNAL: Tat-type signal. MTDSRANRADATRGVASVSRRRFLAGAGLTAGAIALSSMSTSASA >uniprot_features CHOD_STRS0_1 SIGNAL: Tat-type signal. MTAQQHLSRRRMLGMAAFGAAALAGGTTIAAPRAAAAAKSAA >uniprot_features CPRA_DESHA_1 SIGNAL: Tat-type signal. MENNQKRQQSGMSRRSFLKVGAAATTMGVIGAIKAPAKVANA >uniprot_features DHAQ_ACEPO_1 SIGNAL: Tat-type signal. MGRLNRFRLGKDGRREQASLSRRGFLVTSLGAGVMFGFARPSSA >uniprot_features DHML_METEX_1 SIGNAL: Tat-type signal. MLGKSQFDDLFEKMSRKVAGHTSRRGFIGRVGTAVAGVALVPLLPVDRRGRVSRANA >uniprot_features DHML_PARDE_1 SIGNAL: Tat-type signal MLGNFRFDDMVEKLSRRVAGQTSRRSVIGKLGTAMLGIGLVPLLPVDRRGRVSRANA >uniprot_features|DHML_PARVE_1 SIGNAL: Tat-type signal. MLGNFRFDDMVEKLSRRVAGRTSRRGAIGRLGTVLAGAALVPLLPVDRRGRVSRANA >uniprot_features DMSA_ECOLI_1 SIGNAL: Tat-type signal. MKTKIPDAVLAAEVSRRGLVKTTAIGGLAMASSALTLPFSRIAHA >uniprot_features DMSA_RHOCA_1 SIGNAL: Tat-type signal. MTKFSGNELRAELYRRAFLSYSVAPGALGMFGRSLLAKGARA >uniprot features DMSA RHOSH 1 SIGNAL: Tat-type signal. MTKLSGQELHAELSRRAFLSYTAAVGALGLCGTSLLAQGARA >uniprot_features GADH3_PECCY_1 SIGNAL: Tat-type signal. MSEHKNGHTRRDFLLRTITLAPAMAVGSTAMGALVAPMAAGA >uniprot_features GF0_ZYMM0_1 SIGNAL: Tat-type signal. MTNKISSSDNLSNAVSATDDNASRTPNLTRRALVGGGVGLAAAGALASGLQA >uniprot_features MBHS_AZOVI_1 SIGNAL: Tat-type signal. MSRLETFYDVMRRQGITRRSFLKYCSLTAAALGLGPAFAPRIAHA >uniprot_features MBHS_ECOLI_1 SIGNAL: Tat-type signal. MNNEETFYQAMRRQGVTRRSFLKYCSLAATSLGLGAGMAPKIAWA >uniprot_features MBHS_OLICO_1 SIGNAL: Tat-type signal. MTPTETFYEVMRRQGVTRRSFLKFCSLTATALGLGPAYTSEIAHA >uniprot_features | MBHS_RALEH_1 SIGNAL: Tat-type signal. MVETFYEVMRRQGISRRSFLKYCSLTATSLGLGPSFLPQIAHA >uniprot_features NOSZ_ACHCY_1 SIGNAL: Tat-type signal. MESKEHKGLSRRALFSATAGSAILAGTVGPAALSLGAAGLATPARA >uniprot_features NOSZ_PARDE_1 SIGNAL: Tat-type signal. MESKQEKGLSRRALLGATAGGAAVAGAFGGRLALGPAALGLGTAGVATVAGSGAALA >uniprot_features|NOSZ_PSEST_1 SIGNAL: Tat-type signal MSDKDSKNTPQVPEKLGLSRRGFLGASAVTGAAVAATALGGAVMTRESWAQA >uniprot_features PHNS1_DESVH_1 SIGNAL: Tat-type signal MRFSVGLGKEGAEERLARRGVSRRDFLKFCTAIAVTMGMGPAFAPEVAR >uniprot_features PHNS_DESFR_1 SIGNAL: Tat-type signal. MNFSVGLGRDDAEKRLVQNGVSRRDFMKFCATVAAAMGMGPAFAPKVAE >uniprot_features PHNS_DESVM_1 SIGNAL: Tat-type signal. MKISIGLGKEGVEERLAERGVSRRDFLKFCTAIAVTMGMGPAFAPEVARA >uniprot_features XYNC_STRLI_1 SIGNAL: Tat-type signal. MQQDGTQQDRIKQSPAPLNGMSRRGFLGGAGTLALATASGLLLPGTAHA >uniprot_features | YAGT_ECOLI_1 SIGNAL: Tat-type signal MSNQGEYPEDNRVGKHEPHDLSLTRRDLIKVSAATAATAVVYPHSTLAASVPA >uniprot_features YEDY_ECOLI_1 SIGNAL: Tat-type signal MKKNQFLKESDVTAESVFFMKRRQVLKALGISATALSLPHAAHA >uniprot features | YNFE ECOLI 1 SIGNAL: Tat-type signal MSKNERMVGISRRTLVKSTAIGSLALAAGGFSLPFTLRNAAAA



Table S2. 228 short bacterial signal peptides associated with autransporters.

>BIGA_SALTY_1 SIGNAL: Potential. MNPMQKKKLISIAIALTLQSYYIPAIA >ESTA_PSEAE_1 SIGNAL: Potential. MIRMALKPLVAACLLASLSTAPQA >ESTA PSEPU 1 SIGNAL: Potential. MIRMALKPLVAACLLASLSTAPOA >HAP1 HAEIN 1 SIGNAL: Potential. MKKTVFRLNFLTACVSLGIASOAWA >HAP2 HAEIN 1 SIGNAL: Potential. MKKTVFRLNFLTACISLGIVSOAWA >IGA0 HAEIN 1 SIGNAL: Potential. MLNKKFKLNFIALTVAYALTPYTEA >IGA1_HAEIN_1 SIGNAL: Potential. MLNKKFKLNFIALTVAYALTPYTEA >IGA2 HAEIN 1 SIGNAL: Potential. MLNKKFKLNFIALTVAYALTPYTEA >IGA3_HAEIN_1 SIGNAL: Potential. MLNKKFKLNFIALTVAYALTPYTEA >IGA4_HAEIN_1 SIGNAL: Potential. MLNKKFKLNFIALTVAYALTPYTEA >IGA NEIGO 1 SIGNAL: MKAKRFKINAISLSIFLAYALTPYSEA >LIP1 PHOLU 1 SIGNAL: MKRSFIFAPGMLALSISAISNAHA >OMPA_RICCN_1 SIGNAL: Potential. MANISPKLFQKAIQQGLKAALFTTSTAAIMLSSSGALG >OMPA_RICRI_1 SIGNAL: Potential. MANISPKLFKKAIQQGLKAALFTTSTAA >PERT_BORBR_1 SIGNAL: MNMSLSRIVKAAPLRRTTLAMALGALGALGAAPAAHA >PERT_BORPA_1 SIGNAL: Potential. MNMSLSRIVKAAPLRRTTLAMALGALGAAPAAYA >PERT BORPE 1 SIGNAL: MNMSLSRIVKAAPLRRTTLAMALGALGAAPAAHA >PRTS_SERMA_1 SIGNAL: MILNKRLKLAYCVFLGCYGLSIHSSLA >PRTT SERMA 1 SIGNAL: By similarity. MILNKKLKLAYCVFLGCYGLSLHSSLA >SCA1_RICCN_1 SIGNAL: Potential. MNKLTEQHLLKKSRFLKYSLLASISVGA >SCA1_RICFE_1 SIGNAL: Potential. MNKLTEQNLLKKSRFLKYSLLASISVGA >SCA2_RICCN_1 SIGNAL: Potential. MNLQNSHSKKYVLTFFMSTCLLTSSFLSTSARA >SCA2_RICFE_1 SIGNAL: Potential. MSLQNSHSKKYVLTFFMSTCLLTSSFLSTSARA >SCA2_RICSI_1 SIGNAL: Potential. MSTCLLTSSFLSTSARA >SSA1_PASHA_1 SIGNAL: Potential. MYKIKHSFNKTLIAISISSFLSIA >VACA1_HELPY_1 SIGNAL: Potential. MEIQQTHRKINRPLVSLALVGALVSITPQQSHA >VACA2_HELPY_1 SIGNAL: MEIQQTHRKINRPLVSLALVGALVSITPQQSHA >VACA3_HELPY_1 SIGNAL: Potential. MEIQQTHRKINRPIISLALVGVLMGTELGA

Table S2. (Continued)

>VACA4 HELPY 1 SIGNAL: Potential. MEIQOTHRKINRPLVSLALVGALVSITPOOSHA >VACA HELPJ 1 SIGNAL: Potential. MEIQQTHRKINRPLVSLVLAGALISAIPQESHA >VACA_HELPY_1 SIGNAL: Potential. MEIQOTHRKINRPLVSLALVGALVSITPOOSHA >YADA1_YEREN_1 SIGNAL: MTKDFKISVSAALISALFSSPYAFA >YADA2_YEREN_1 SIGNAL: MTKDFKISVSAALISALFSSPYAFA >YADA YERE8 1 SIGNAL: By similarity. MTKDFKISVSAALISALFSSPYAFA >YADA YERPS 1 SIGNAL: By similarity. MTKDFKISVSAALISALFSSPYAFA >YAIT ECOLI 1 SIGNAL: Potential. MHSWKKKLVVSQLALACTLAITSQANA >YAIT_SALTY_1 SIGNAL: Potential. MHSWKKKLVVSQLALACTLAITSQANA >YFAL_ECOLI_1 SIGNAL: Potential. MRIIFLRKEYLSLLPSMIASLFS >YPJA_ECOLI_1 SIGNAL: Potential. MNRTSPYYCRRSVLSLLISALIYAPPGMA >YTRP_PSEPU_1 SIGNAL: Potential. MRKAPLLRFTLASLALACSQAFA >YUAO_ECOLI_1 SIGNAL: Potential. MCFFLGSRLAYA >B0J5Y7_RHILT_1 SIGNAL: Potential. MLRLTGLASTAALVLAVGPGWAQ >B5ZHX8_GLUDA_1 SIGNAL: Potential. MAGIFRLVLIASPFTAVTSVSFAQ >B5ZKP4_GLUDA_1 SIGNAL: Potential. MRVSVSSLAILCALRLALPHQASAQ >069257_BORPE_1 SIGNAL: Potential. MNMSLSRIVKAAPLRRTTLAMALGALGAAPAAHA >069259_BORPE_1 SIGNAL: Potential. MNMSLSRIVKAAPLRRTTLAMALGALGAAPAAHA >088143_BORPE_1 SIGNAL: Potential. MNMSLSRIVKAAPLRRTTLAMALGALGAAPAAHA >Q0P6P2 PSEPU 1 SIGNAL: Potential. MRKAPLLRFTLASLALACSQALA >Q1JYT2_DESAC_1 SIGNAL: Potential. MTRYFLLAVLCVAILFAQPLQAS >Q1K1B2_DESAC_1 SIGNAL: Potential. MKKLYIIIPFLMMGLFPPAPCHAN >Q1K3G2_DESAC_1 SIGNAL: Potential. MRSLKMACPTLCVMLLTLCWSGLAAAY >Q3R162_XYLFA_1 SIGNAL: Potential. MQKIKNKFIVRTILATTVTTVLSACGG >Q3R2M7_XYLFA_1 SIGNAL: Potential. MERKNHKKTTLATLISVLLMGSAGATYAN >Q3R323_XYLFA_1 SIGNAL: Potential. **MTSNFTRSLLAFAITLTTTOGIAK** >Q3R8A8_XYLFA_1 SIGNAL: Potential. MERKNHKKTTLATLISVLLMGGAGATYAN >Q3R9L6_XYLFA_1 SIGNAL: Potential. MKLKFQKRKFLTVVIVFSMCGGSVVYAN >Q3RA15_XYLFA_1 SIGNAL: Potential.

(Continued)



MTSNFTRSLLAFAITLTTTOGIAK >Q3RAY9 XYLFA 1 SIGNAL: Potential. MKTIFAARTILASALAAALSACGD >Q3RB92_XYLFA_1 SIGNAL: Potential. MKNMKTTFFPGSILVLTLVAFLSACGG >Q3RE35_XYLFA_1 SIGNAL: Potential. MQKIKNKFIVRTILATTVTTVLSACGG >Q3REZ6_XYLFA_1 SIGNAL: Potential. MKTIFAARTILASALAAALSACGD >Q3RGD6_XYLFA_1 SIGNAL: Potential. MTSNFTRSLLAFAITLTTTQGIAK >Q3RGT2 XYLFA 1 SIGNAL: Potential. MERKNHKKTTLATLISVLLMGGAGATYAN >Q546U4 BORPE 1 SIGNAL: Potential. MNMSLSRIVKAAPLRRTTLAMALGALGAAPAAHA >Q5FY73_SALBN_1 SIGNAL: Potential. MNKIYALKYSVRQGALVPV >Q83YP8_ACTAC_1 SIGNAL: Potential. MKILLKPFRYSVIATTIALVFNQPAFA >Q9S3M8_BORPE_1 SIGNAL: Potential. MNMSLSRIVKAAPLRRTTLAMALGALGAAPAAHA >Q9S6M9_BORPE_1 SIGNAL: Potential. MNMSLSRIVKAAPLRRTTLAMALGALGAAPAAHA >A0B1Z1_BURCH_1 SIGNAL: Potential. MFLALSGAGIVPAHATCSTAG >A0LDC0_MAGSM_1 SIGNAL: Potential. MKRSNLSLLTLSLATTGLLLMTQPAVAF >A1JMW6_YERE8_1 SIGNAL: Potential. MHKIYRFIKTFMVSCPVLLGGFSVVDA >A1JSQ7_YERE8_1 SIGNAL: Potential. MNINNIARLPCFRKTLLASLLVPLLTPLYSWA >A4JIS0_BURVG_1 SIGNAL: Potential. MNHRSFPLSRTRTGRRLAHSALIAGAVMPWASSAQ >A4THP4_YERPP_1 SIGNAL: Potential. MNNHKIWRLSAVAVALLISGNSYAD >A4TMU3 YERPP 1 SIGNAL: Potential. MKNSNTLNTRLLPLSILISSLVSGGAMAV >A4TMU4 YERPP 1 SIGNAL: Potential. MKSRHHLNTRLLPLSILISALIPAAVLAA >A4TP40_YERPP_1 SIGNAL: Potential. MDKTLLAGAISLSLVILPVQVLAF >A4TQZ3_YERPP_1 SIGNAL: Potential. MKTNRSTLSPCFRKTMIASLLVPLCSPLYSWAV >A4W658_ENT38_1 SIGNAL: Potential. MKTTFRLTQVATSISLLLGSSVVIPGTALAN >A4W9S4_ENT38_1 SIGNAL: Potential. MIGGVVSGFGILASPAALAA >A4WEB5_ENT38_1 SIGNAL: Potential. MKKKYLSQLISLLVASTAAQGLLTTHALAV >A4WED8_ENT38_1 SIGNAL: Potential. MRKLLKRSLLSQCVLMSLTSLSAFAA >A4WGQ4_ENT38_1 SIGNAL: Potential. MSKNITNPTAIDRRKVLGLSIGSAIALLSSAE >A4XFF0_NOVAD_1 SIGNAL: Potential. MKTNKSRLALGAASAAVAVGLAAQVQAA >A4XXQ8_PSEMY_1 SIGNAL: Potential. MALRVALIGLGLLQVATGVVAG

Table S2. (Continued)

>A5FVP6 ACICJ 1 SIGNAL: Potential. MKISRRFSLLAATALSLSACSG >A5FX02 ACICJ 1 SIGNAL: Potential. MGSEMTFETCVRRLAIGALAAGFIGLAPARAQ >A5V6F0 SPHWW 1 SIGNAL: Potential. MRYLLASTCLAAIAAVPVHAE >A5VXL2_PSEP1_1 SIGNAL: Potential. MRKAPLLRFTLASLALACSQALAG >A5W1C9 PSEP1 1 SIGNAL: Potential. MKSTSNPLRFDSIFYAVSTSLLLATPVETIAY >A5W3S9 PSEP1 1 SIGNAL: Potential. MRLRLMLTLGSLPLLGMVTPAQAN >A5W741 PSEP1 1 SIGNAL: Potential. MGIVKKORGGPLVRAKVVMSALMMLSPIAOAL >A6UG19 SINMW 1 SIGNAL: Potential. MVVAKVRKCFSTVLSGALLAGLFCIVGSGEASAA >A6X2Y0_OCHA4_1 SIGNAL: Potential. MTGVLRHKSMLLMTTAALGFYVTTARGA >A6X4K3_OCHA4_1 SIGNAL: Potential. MHCRNIGSRAIRFLSSTALISLGTVLLQSTPGMAA >A6X7J0_OCHA4_1 SIGNAL: Potential. MKKLWLASTAIISASLFTSAAWSA >A6X8H2_OCHA4_1 SIGNAL: Potential. MLKRLNGKNVLFLRFLFLSAGTALAMTPVLAQ >A7HZ72_PARL1_1 SIGNAL: Potential. MTARNRTAAARRRHIAALMLGTALAALPHSGASAD >A7IHM7 XANP2 1 SIGNAL: Potential. MTVQSLFHGVSRTRALGLVAIGLGAGSLAQTVLAD >A7IIP4_XANP2_1 SIGNAL: Potential. MVGRSAVIRAAVLWTSIAAGTGAAHAQ >A7IL68_XANP2_1 SIGNAL: Potential. MPALYASSSRLVLSLFLVGMAAPALAA >A8G7T5_SERP5_1 SIGNAL: Potential. MPLKITRMPRPAVLAVAILCSMTTSALAY >A8GG96_SERP5_1 SIGNAL: Potential. MASFAPSFFSRGCLLALATTGGFSAVVNAA >A8GHU3_SERP5_1 SIGNAL: Potential. MKDKISHHLAVRKPLSKIYLALFSAPLLLMGSADMAR >A8GIR8 SERP5 1 SIGNAL: Potential. MKRLAIAIIAALPFCSAQAV >A8GL40_SERP5_1 SIGNAL: Potential. MKQATGKNKPALAPTWKLNALLCALLAAGGVQAA >B0KJB0_PSEPG_1 SIGNAL: Potential. MRKAPLLRFTLASLALACSQAFAA >B0KMU3_PSEPG_1 SIGNAL: Potential. MKSTSNPLRFDSIFYAVSTSLLLATPVETFAY >B0KU28_PSEPG_1 SIGNAL: Potential. MRCHRLLLPVALPLLTLPLLAHSQ >BOKUR3_PSEPG_1 SIGNAL: Potential. MIKRRNFTLSPLASAIGQLLLGASAVLFTGPAGAL >B0KUS7_PSEPG_1 SIGNAL: Potential. MHFTPNRLALCIALACAAFAPSAFAK >B0T124_CAUSK_1 SIGNAL: Potential. MQRKVLVATVATAPLLAMAFGAYAE >B0T4N4_CAUSK_1 SIGNAL: Potential. MMTRSSSKRTILAGSSLLVMAIAAAQPALAQ >B0U200_XYLFM_1 SIGNAL: Potential.

(Continued)



MKNMKTTFFPGSILVLTLVAFLSACGG >B0U225_XYLFM_1 SIGNAL: Potential. MKTIFAARTILASALAAALSACGD >B0U2J0_XYLFM_1 SIGNAL: Potential. MQKIKNKFIVRTILATTVTTVLSACGG >B0U5B1_XYLFM_1 SIGNAL: Potential. MTSNFTRSLLAFAITLTTTQGIAK >B0U5U7_XYLFM_1 SIGNAL: Potential. MERKNHKKTTLATLISVLLMGGAGATYAN >B0UED5_METS4_1 SIGNAL: Potential. MRFLSGVSLAAVITAIMGVGAARAQ >B1IVK2_ECOLC_1 SIGNAL: Potential. MNRTSPYYCRRSVLSLLISALIYAPPGMAA >B1IY80_ECOLC_1 SIGNAL: Potential. MHQSGSVSLCRSAISVLVATALYSPIALAS >B1J074 ECOLC 1 SIGNAL: Potential. MHSWKKKLVVSQLALACTLAITSQANAA >B1J4X2_PSEPW_1 SIGNAL: Potential. MPSFSPVTLRHVLHVASLAPLLLLTPQAMAQ >B1J6N8_PSEPW_1 SIGNAL: Potential. MLORLFCSLSLLTLAISAAHAA >B1J6R7_PSEPW_1 SIGNAL: Potential. MKSTSNPMRFDRIFYAVSTSMLLATPVETFAF >B1JE37_PSEPW_1 SIGNAL: Potential. MRKAPLLRFTLATLALACSQAFAA >B1JGR2_YERPY_1 SIGNAL: Potential. MDKTLLAGAISLSLVTLPVQVLAF >B1JHT7_YERPY_1 SIGNAL: Potential. MKKNRSTLSPCFRKTLIASLLVPLCSPLYSWAV >B1JLC5_YERPY_1 SIGNAL: Potential. MNNHKIWRLSAVAVALLISGNGYAD >B1JPQ0_YERPY_1 SIGNAL: Potential. MHNKFKANTLAISIAAILLSVSFNTLAV >B1JSJ9_YERPY_1 SIGNAL: Potential. MNTNSKKTYLSIAISSILYASTAMNANAD >B1LTI2_METRJ_1 SIGNAL: Potential. MTRGYGVSLAALAVALLGPGAQAQ >B1YPU8_BURA4_1 SIGNAL: Potential. MNRRFPFSQARSGRRLAHSALIAGAVIPWPSAAQ >B2FHR8_STRMK_1 SIGNAL: Potential. MRMMQFTPKFPSAKNQSDLARAIATALLIATSGAAGA >B2FI22_STRMK_1 SIGNAL: Potential. MNHPLHGRSSHSRSPLHSRLALAVSSSLLLAAAAPAMA >B2FPV5 STRMK 1 SIGNAL: Potential. MERTMMVRSVLATALAMALTACG >B2FSC8 STRMK 1 SIGNAL: Potential. MLLSKRPIRTLMAAAIALAALPAMA >B2FU91_STRMK_1 SIGNAL: Potential. MKHSKLSLALAGLIAVGAIA >B2FUQ4_STRMK_1 SIGNAL: Potential. MYRAVPRAFRPRRLAVSVLQALAVPSLLLTTASVAWA >B2I6G5_XYLF2_1 SIGNAL: Potential. MKLKFQKRKFLTVVIVFSMYGGSVVYAN >B2I753_XYLF2_1 SIGNAL: Potential. MKNMKTAFFPGSILVLTLVAFLSACGG

Table S2. (Continued)

>B2I7M9 XYLF2 1 SIGNAL: Potential. MKTIFAARTILASALAAALSACGD >B2I921_XYLF2_1 SIGNAL: Potential. MERKNHKKTTLATLISVLLMGSAGATYAN >B2I9C9_XYLF2_1 SIGNAL: Potential. MTSNFTRSLLAFAITLTTTQGIAK >B2JZU5_YERPB_1 SIGNAL: Potential. MHNKFKANTLAISIAAILLSVSFNTLAV >B2K394_YERPB_1 SIGNAL: Potential. MNNHKIWRLSAVAVALLISGNGYAD >B2K6L6_YERPB_1 SIGNAL: Potential. MKKNRSTLSPCFRKTLIASLLVPLCSPLYSWAV >B2K7Q9_YERPB_1 SIGNAL: Potential. MDKTLLAGAISLSLVTLPVQVLAF >B2K953_YERPB_1 SIGNAL: Potential. MNTNSKKTYLSIAISSILYASTAMNANAD >B2K9Q8 YERPB 1 SIGNAL: Potential. MKSRSNLNTRLLPLSILISSLIPGAVLAA >B3QEQ3_RHOPT_1 SIGNAL: Potential. MKKQLLLTTSLAPLFAVGVLGGSPAHAD >B3QI73_RHOPT_1 SIGNAL: Potential. MVAGVVAIGLVGATSVSAQAQ >B4EH84_BURCJ_1 SIGNAL: Potential. MKRKARNLGLGGMAILTGAVPVSAYA >B4EQ87 BURCJ 1 SIGNAL: Potential. MGSNKKRAEARPKLLVPTGMLVALVGAGIVPA >B4SIZ3_STRM5_1 SIGNAL: Potential. MSRAVPRAFRPRRLAVSVLQALAVPSLLLTTAGVAWAG >B4SJV7 STRM5 1 SIGNAL: Potential. MNHPLYGRSSHSRSPRSRLALAVASCLLLAAATPATAS >B4SLP2_STRM5_1 SIGNAL: Potential. MHPFPALPAHRKAILGSALLAALLGMAAAPAARAS >B4SLT8_STRM5_1 SIGNAL: Potential. MRMMQFTPKFPSTLVRSRLGLAVASSLMLAAMGTANAV >B4SP28_STRM5_1 SIGNAL: Potential. MKHSKLSLALAGLLGIGAIMAA >B4SRI9_STRM5_1 SIGNAL: Potential. MERTTMVRSVLATALAMALTACGG >B4STS4_STRM5_1 SIGNAL: Potential. MQLSKHPIRSLMAAAIALAALPAMAG >B5QTZ3_SALEP_1 SIGNAL: Potential. **MPTPONYSFIAIAVSAALASMVFPSOA** >B5QUX9 SALEP 1 SIGNAL: Potential. MTOKRTLLKYGILSLALAAPLSACA >B5R2C8 SALEP 1 SIGNAL: Potential. MNPMOKKKLISIAIALTLOSYYIPAIA >B5R4M8 SALEP 1 SIGNAL: Potential. MIVRKRRGRRTLCCLAGLMACSFFINTTYAWQ >B5R4X4_SALEP_1 SIGNAL: Potential. MHSWKKKLVVSQLALACTLAITSQANA >B5R5W4_SALG2_1 SIGNAL: Potential. MHSWKKKLVVSQLALACTLAITSQANA >B5R6Z9_SALG2_1 SIGNAL: Potential. MTQKRTLLKYGILSLALAAPLSACA >B5RG61_SALG2_1 SIGNAL: Potential.

(Continued)



MPTPONYSFIAIAVSAALASMVFPSOA >B5RGN6_SALG2_1 SIGNAL: Potential. MIVRKCRGRRTLCCLAGLMACSFFINTTYA >B8ENQ5_METSB_1 SIGNAL: Potential. MSDSNFNLKSLAAILLGGVALPLLLPASQALAA >B8ILY6_METNO_1 SIGNAL: Potential. MLGCESGRAHGLDRMRRLLLLGVSFAGLPVLATSALAQ >086135_BORPE_1 SIGNAL: Potential. MHIYGNMNRATPCRGAVRALALALLGAGMWTLSPPSAWA >Q07JX7_RHOP5_1 SIGNAL: Potential. MATNRSRRGRVTLLAVAIAAGPIAFTTADRAR >Q07P01_RHOP5_1 SIGNAL: Potential. MLVTHARRNGAPQRSTGWLASLTAIAVAILAYPELSLAQ >Q0AKS3 MARMM 1 SIGNAL: Potential. MPHTPTPLTLKSLLVASTAIVGIAAATPAFAQ >Q0AQ16_MARMM_1 SIGNAL: Potential. MRRRFLFASALVSLTVTAPAAFAD >Q0AT43_MARMM_1 SIGNAL: Potential. MAAVNRGVRTDAKAGAPRRVSMLLAGIALGFSAPVHAA >Q11GI1_MESSB_1 SIGNAL: Potential. MNVFETLSKGVSSAAICACIATLPGNLHAQ >Q11M33_MESSB_1 SIGNAL: Potential. MKTTAYLKLSGSLIALTAAAVSQAHAQ >Q12FI0_POLSJ_1 SIGNAL: Potential. MTLNYSFSYFMOKPIRYSLTAAACLLAFSAQAQ >Q132Z9_RHOPS_1 SIGNAL: Potential. MQRMTLAPSRRIAMLWSAALALSASASPALAQ >Q139I7_RHOPS_1 SIGNAL: Potential. MGAGRHFRNLSTLFLCTTFLVSAPVSAALYAA >Q1C198_YERPA_1 SIGNAL: Potential. MKTNRSTLSPCFRKTMIASLLVPLCSPLYSWAV >Q1C326_YERPA_1 SIGNAL: Potential. MNNHKIWRLSAVAVALLISGNSYAD >Q1C4W8_YERPA_1 SIGNAL: Potential. MDKTLLAGAISLSLVILPVQVLAF >Q1C5I0_YERPA_1 SIGNAL: Potential. MKSRHHLNTRLLPLSILISALIPAAVLAA >Q1C5I1_YERPA_1 SIGNAL: Potential. MKNSNTLNTRLLPLSILISSLVSGGAMAV >Q1CAV8_YERPA_1 SIGNAL: Potential. MHNKFKANTLAISIAAILLSVSFNTLAV >Q1CDG9 YERPN 1 SIGNAL: Potential. MKTNRSTLSPCFRKTMIASLLVPLCSPLYSWAV >O1CF77 YERPN 1 SIGNAL: Potential. MHNKFKANTLAISIAAILLSVSFNTLAV >O1CK98 YERPN 1 SIGNAL: Potential. MKNSNTLNTRLLPLSILISSLVSGGAMAV >Q1CK99 YERPN 1 SIGNAL: Potential. MKSRHHLNTRLLPLSILISALIPAAVLAA >Q1CKV3_YERPN_1 SIGNAL: Potential. MDKTLLAGAISLSLVILPVQVLAF >Q1CMH9_YERPN_1 SIGNAL: Potential. MNNHKIWRLSAVAVALLISGNSYAD >Q1GU46_SPHAL_1 SIGNAL: Potential. MRKTLLASTCLATLLSTAVHAE

Table S2. (Continued)

>Q1MHZ0_RHIL3_1 SIGNAL: Potential. MRIYRWLSASVGRHVGLATLFAGMALFLDAYG >Q1MK18_RHIL3_1 SIGNAL: Potential. MSPFCGSPTVLFSLLIPGTIMGGD >Q1QCE5_PSYCK_1 SIGNAL: Potential. MPRNISHAIVNPTTLKTLTKSMLAISLSMAGLAHAE >Q214Q0_RHOPB_1 SIGNAL: Potential. MGAGFFRDVSKLLLCTTFLVAAPVSAVLQAA >Q2GAY1_NOVAD_1 SIGNAL: Potential. MDRLRTTTILSTLAGTPVALALLVPQAANAA >Q2J1P8 RHOP2 1 SIGNAL: Potential. MOKARTRILAGFAFAMATSVSTGAVAAC >Q2KTY1_BORA1_1 SIGNAL: Potential. MTQYPARRPPSHALTAVVLALSSLA >Q2Y8Z0_NITMU_1 SIGNAL: Potential. MAKRKKSASLSLYAKFIIALLMAPVASLSSRAQ >Q3BM38_XANC5_1 SIGNAL: Potential. MSTNCTNMAAGVRVVLRWPLVFALLLLSTLYSGKAAA >Q3BVT0_XANC5_1 SIGNAL: Potential. MKKODVARTVLASALAVALTACG >Q3K4D0 PSEPF 1 SIGNAL: Potential. MPIOCKYKLOHLVLAVALAVGCVEFSLAE >Q3K4D1_PSEPF_1 SIGNAL: Potential. MPFPPQRLSFAIALLIATSAAHGK >Q3K5V0 PSEPF 1 SIGNAL: Potential. MIKOTLFVPLAGCLLAMACAQANAA >Q3KC67_PSEPF_1 SIGNAL: Potential. MORKISNVRLRDIRWGLVLSSFLAPFSQIAIGG >Q3KCT0_PSEPF_1 SIGNAL: Potential. MKNNNTPAQSGGGFRLKTLNVALLCAMATWGSAHAA >Q3KCT1_PSEPF_1 SIGNAL: Potential. MDVRIKPISVGTLLLVISATQAQAQ >Q3KD77_PSEPF_1 SIGNAL: Potential. MKTSLTSEEIKTTFCTVSSSILLCSSMEAQAG >Q3KDN6_PSEPF_1 SIGNAL: Potential. MFPRFLCSLSVLSLSIAAVHAA >Q3SNP6_NITWN_1 SIGNAL: Potential. MNVVVRASMPGNGALRRRVVAGGAFAFALSASSGAIAA >Q4ZMI6_PSEU2_1 SIGNAL: Potential. MTKTSRRWPFAACLLSLACGTAAAA >Q4ZQ07 PSEU2 1 SIGNAL: Potential. MQKSKCVGVVRYSFKPVATGALCALSFTFGCSAYAD >04ZUU5 PSEU2 1 SIGNAL: Potential. MNAPFVLRPLSWTLKTVIFLSPLLPGSHAFAO >04ZYT2 PSEU2 1 SIGNAL: Potential. MFRKTLLAMAMAATAVPACAE >Q664E5 YERPS 1 SIGNAL: Potential. MKQNRSTLSPCFRKTLIASLLVPLCSPLYSWAV >Q665R2_YERPS_1 SIGNAL: Potential. MNNHKIWRLSAVAVALLISGNGYAD >Q667C1_YERPS_1 SIGNAL: Potential. MHNKFKANTLAISIAAILLSVSFNTLAV >Q667Z0_YERPS_1 SIGNAL: Potential. MKSRHHLNTRLLPLSILISALIPAAVLAA >Q667Z1_YERPS_1 SIGNAL: Potential.

(Continued)



MKSRSNLNTRLLPLSILISSLIPGAVLAA >Q667Z2_YERPS_1 SIGNAL: Potential. MKNSNTLNTRLLPLSILISSLVSGGAMAV >Q66DI5_YERPS_1 SIGNAL: Potential. MDKTLLAGAISLSLVTLPVQVLAF >Q6N1B5_RHOPA_1 SIGNAL: Potential. MRRAASCQSACLPTVIPLSIAE >Q6N8G7_RHOPA_1 SIGNAL: Potential. MSTVGRFRHLSSLLLCTTFLVSAPMSAVLYAA

Publish with Libertas Academica and every scientist working in your field can read your article

"I would like to say that this is the most author-friendly editing process I have experienced in over 150 publications. Thank you most sincerely."

"The communication between your staff and me has been terrific. Whenever progress is made with the manuscript, I receive notice. Quite honestly, I've never had such complete communication with a journal."

"LA is different, and hopefully represents a kind of scientific publication machinery that removes the hurdles from free flow of scientific thought."

Your paper will be:

- Available to your entire community free of charge
- Fairly and quickly peer reviewed
- Yours! You retain copyright

http://www.la-press.com