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
Introduction
Bacterial autotransporters translocate via the Sec complex across the inner membrane of Gram-negative bacteria and translocate themselves across the outer membrane. This is accomplished by a translocator domain at the C-terminus of the autotransporter which adopts a β-barrel fold within the outer membrane resembling a porin-like domain. The trimeric autotransporter consist of an N-terminal signal sequence, a central “passenger domain”, and a β-barrel forming translocation unit. The β-barrel domain is necessary for the secretion of the passenger domain and connected via an α-helical linker region. 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. For translocation across the inner bacterial membrane autotransporters possess N-terminal signal sequences. In 2007 Dautin and Bernstein reported around 10% of the known 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. 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. “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. 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.

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. 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. Desvaux et al continued this approach and termed the N2 and H2 region the “extended signal peptide region” (ESPR). 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 al and 35 further long signal sequences of bacterial autotransporters taken from Henderson et al. Two-partner secreted proteins are known to possess an N-terminal conserved region important for their secretion. Additionally we performed a sequence database search in UniProtKB/SwissProt Release 14.7 using the sequence retrieval system (SRS, Release 7.1.3). 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 al and Szabady et al. From the sequences considered suitable by the work of Henderson et al Szabady et al and our own database search we assembled a dataset of 90 sequences. The signal peptidase cleavage sites were used as suggested in Henderson et al 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 retrieved from the UniProtKB/SwissProt Release 14.7 using the Sequence Retrieval System (SRS, Release 7.1.3). We searched for
“non-potential” bacterial signal sequences with evidence at protein level and a length of at least 40 residues. All retrieved sequences contain the twin-arginine (TAT) 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 14.7 using SRS (Release 7.1.3). 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, 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 referred 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 function (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 al. 14 are predicted to have a two-domain organization (Table 1). Of the 46 autotransporter signal sequences collected in Henderson et al. 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 al. 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 and for CAR56027 a signal anchor by SignalP. 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 al. and the conservation of the “N-terminal extension” reported by Szabady et al. within a margin of ±5 residues.

Szabady et al. further reported a conserved sequence pattern in the N-terminal extension of autotransporter and two-partner secretion systems signal peptides. This conservation is also present in 43 of 46 sequences compiled by Henderson et al. 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 al. although they are annotated to contain a domain similar to autotransporters. This might argue for the sequences of Szabady et al. 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 al. and Henderson et al. 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. The long signal peptide of subtype EspPα contains the conserved sequence pattern reported by Szabady et al. and for which experimental results were published by Peterson et al. 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

Bioinformatics and Biology Insights 2009:3
Table 1. NtraC analysis of 90 long bacterial signal peptides.

| Nr. | Accession number¹ | SP length² | NtraC³ | N-domain⁴ | C-domain | Predicted C-domain targeting | Organism          |
|-----|------------------|------------|--------|-----------|----------|-------------------------------|-------------------|
| 1   | NP_052685        | 55         | yes    | 1–26     | 27–55    | gram–      | E. coli²                     |
| 2   | AAC44731/ AAG37043 | 53         | yes    | 1–17(1–35)| 18–53(36–53)| gram–          | E. coli²                 |
| 3   | CAA11507         | 52         | yes    | 1–17     | 17–52    | gram–      | E. coli²                     |
| 4   | AAD23953         | 55         | yes    | 1–15(16) | 17–55    | SP/SA⁹     | E. coli¹⁰                  |
| 5   | CAA88252         | 56         | no     | –        | –        | –             | S. flexneri                |
| 6   | CAA46156         | 42         | yes    | 1–26     | 27(28, 29)–42 | gram–       | E. coli¹¹                 |
| 7   | CAC14227         | 75         | yes    | 1–40     | 41–75    | gram–      | Y. pestis                   |
| 8   | AAC43721         | 50         | no     | –        | –        | –             | H. influenzae              |
| 9   | CAC14202         | 62         | yes    | 1–27(–44)| 28–62    | gram–      | P. multocida                |
| 10  | AAK68872/ AAK09243 | 51         | yes    | 1–26     | 27–51    | gram–      | N. meningitidis             |
| 11  | 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  | AAA20254         | 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. meningitidis             |
| 20  | AAF67320         | 54         | yes    | 1–30     | 31–54    | gram–      | S. flexneri                 |
| 21  | AAG30168         | 49         | yes    | 1–29     | 30–49    | gram–      | E. coli¹²                  |
| 22  | AAK00474         | 52         | yes    | 1–33(1–36) (mTP,SP)| 37–52    | –          | 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. coli¹³                  |
| 25  | AAL78284         | 67         | yes    | 1–30(39) | 31(40)–67| gram–      | M. catarrhalis             |
| 26  | AAQ22366         | 56         | no     | –        | –        | –             | A. actinomycetemcomitans   |
| 27  | CAA88252         | 56         | no     | –        | –        | –             | 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. coli¹⁴                  |
| Accession | Length | Lipid | Start| End | Identity | Species               |
|-----------|--------|-------|------|-----|----------|-----------------------|
| CAC92482  | 69     | yes   | 1–37(41) | 38(42)–69 | gram–   | Y.pestis             |
| NP_274768 | 80     | yes   | 1–48(49)mTP | 49(50)–80 | gram–   | N.meningitis         |
| NP_308389 | 48     | yes   | 1–15(24)(38) | 16(25)(39)–48 | mTP15, SA, SP | E.coli16            |
| NP_519008 | 72     | yes   | 1–36(37)(39) | 33(40)–72 | gram–   | R.solanacearu        |
| NP_519896 | 60     | yes   | 1–30(–38)   | 31(–39)–60 | gram–   | R.solanacearu        |
| NP_522634 | 66     | yes   | 1–33(34)(41) | 34(35, 42)–66 | gram–   | R.solanacearu        |
| NP_636050 | 66     | yes   | 1–43(55,56,58) | 27(28)(31)(44)–66 | gram–   | X.campestris         |
| NP_807449 | 50     | yes   | 1–32(39)   | 29(30)–50 | SP, mTP  | S.enterica           |
| YP_553065 | 57     | yes   | 1–37(mTP)  | 30(–38)–57 | gram–   | B.xenovarans         |
| ZP_00033562 | 57   | yes   | 1–28(29)  | 29(30)–57 | gram–   | B.fungorum           |
| ZP_00041732 | 67  | yes   | 1–58(mTP) | 30(33, 35)–67 | SP      | X.campestris         |
| ZP_00088699 | 49  | yes   | 1–32(mTP) | 18(29, 32, 33)–49 | gram–   | A.vinelandii         |
| ZP_00122019 | 77  | yes   | 1–40      | 41(43, 48, 52, 53)–77 | gram–   | H.somnus             |
| YP_719000  | 69     | yes   | 1–37      | 38(39, 44, 49)–69 | SP      | H.somnus             |
| ZP_00132251 | 69  | yes   | 1–40      | 41(42, 43)–69 | gram–   | H.somnus             |
| AAC26634   | 52     | no    | –         | –            | –       | E.coli17             |
| AAC74583   | 61     | yes   | 1–29(40, mTP) | 30(41, 42)–61 | gram–   | E.coli18             |
| AAD41751   | 54     | yes   | 1–15(27, 28, 29) | 16(28, 29, 30)–54 | gram–   | E.coli19             |
| AAF43424   | 52     | yes   | 1–32      | 33–52       | gram–   | E.coli20             |
| AAG53941   | 65     | yes   | 1–38 (mTP) | 23(29, 30, 39)–65 | gram–   | B.bronchiseptica     |
| AAP33781   | 52     | yes   | 1–16      | 17–52       | gramp+, SP | E.coli21             |
| Q84GK0     | 56     | yes   | 1–16      | 17–56       | gram+    | E.coli22             |
| Q7BC4      | 52     | yes   | 1–16(17)  | 18–52       | gram+, SP | S.flexneri           |
| Q8CWC7     | 55     | yes   | 1–16      | 17–55       | SA       | E.coli23             |
| YP_840389  | 41     | yes   | 1–24(SA, mTP) | 25–41       | SP       | B.cenocepacia        |
| YP_916859  | 40     | yes   | 1–19(20)  | 21–40       | gram–, SP | P.denitrificans      |
| YP_00110606 | 47   | yes   | 1–18(19)  | 20–47       | gram–, SP, gram+ | B.vietnamiensis     |
| YP_001161762 | 50  | yes   | 1–15(28)  | 16(32)–50  | gram–, SP, gram+ | Y.pestis           |
| YP_001165485 | 48  | no    | –         | –           | –       | Enterobacter sp. 638 |
| YP_001281217 | 41  | no    | –         | –           | –       | Psychrobacter sp. PRwf–1 |
| YP_001371305 | 40  | no    | –         | –           | –       | O.anthropi           |
| YP_001371348 | 45  | yes   | 1–25(mTP) | 26–45       | SP, gram– | O.anthropi           |
| YP_001373319 | 45  | yes   | 1–20      | 21–45       | gram–, SP, gram+ | O.anthropi           |
| YP_001948780 | 41  | yes   | 1–18(24)  | 25–41       | gram–, SP, gram+ | B.multivorans       |
| Nr. | Accession number¹ | SP length² | NtraC³ | N-domain⁴ | 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     | –        | –        | –                           | B.ambifaria |
| 68  | YP_001811545      | 42         | no     | –        | –        | –                           | 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         | no     | –        | –        | –                           | 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(1–33mTP) | 18–47    | gram−, SP, gram+            | B.cepacia |
| 75  | QOPAN9            | 40         | yes    | 1–17    | 18–40    | gram−, SP, gram+            | C.jejuni |
| 76  | Q1MKD9            | 40         | yes    | 1–27(29)(mTP) | 30–40    | –                           | 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  | Q2JON4            | 44         | (yes)  | 1–37(mTP) | 38–44    | –                           | 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    | –                           | 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 negative SP            | B.avenum |
| 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  | Q6NSZ4            | 46         | yes    | 1–17(–21) | 24–46    | gram−, SP, gram+            | R.palustris |
| 89  | Q6N664            | 46         | yes    | 1–18(23) | 24–56    | gram−, SP, gram+            | R.palustris |
| 90  | P39180            | 52         | yes    | 1–35(u.–29 mTP) | 36–52    | SP, gram− (30–52)            | E.coli²⁴ |

¹Nr. 1–16 taken from Szabady et al.¹⁰; Nr. 17–51 taken from Henderson et al.¹⁰; Nr. 52–90 retrieved via SRS. ²gives the position of the last residue of the signal peptides as used in the publications or annotated in UniProKB v. 14.7. ³NtraC organization of the sequence. ⁴Length of the predicted N- or C-domain. Numbers in brackets refer to alternative possibilities for truncation of the domains. A targeting abbreviation in brackets means: only that length combination leads to the targeting function in brackets. Predicted C-domain targeting: predicted by SignalP or TargetP. ⁵/⁶: gram negative or gram positive signal peptide. ⁷strain = Sakai. ⁸strain = E2348/69. ⁹strain = EB1. ¹⁰SP: eukaryotic signal peptide, SA: signal anchor gram. ¹¹strain = 042. ¹²strain = K–12. ¹³strain = CFT073. ¹⁴strain = EH41. ¹⁵strain = 4797/97. ¹⁶mitochondrial targeting peptide. ¹⁷strain = Sakai. ¹⁸strain = 042. ¹⁹strain = K–12. ²⁰strain = H10407. ²¹strain = ML308–225. ²²strain = APEC13. ²³strain = O78:H11. ²⁴strain = O6:H1. ²⁵strain = K12.
sequences to mediate an interaction with an unknown cytosolic factor or to induce an unusual signal peptide conformation prior to protein translocation. 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 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. 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. 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 secretion a net negative charge or zero charge could result in considerably decreased rates of export. While Dierstein and Wickner reported that the N-terminal regions is not strictly required for processing by signal peptidase, Peterson et al demonstrated that the positive charges in the N-terminal part of the bacterial signal sequences may influence SRP recognition.

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.

|             | Histidine (His) | Lysine (Lys) | Arginine (Arg) | Aspartic Acid (Asp) | Glutamic Acid (Glu) |
|-------------|-----------------|--------------|----------------|---------------------|---------------------|
| N-domain    | 0.5 (±0.8)²⁶    | 1.9 (±1.5)   | 2.3 (±1.7)     | 0.4 (±0.6)          | 0.8 (±0.7)          |
| C-domain (+tra) | 0.3 (±0.5)     | 1.1 (±1.4)   | 1.5 (±1.4)     | 0.2 (±0.4)          | 0.3 (±0.6)          |
| N-domain (+tra) | 0.6 (±0.8)     | 2.3 (±1.8)   | 2.6 (±1.8)     | 0.4 (±0.7)          | 0.9 (±0.7)          |
| C-domain (−tra) | 0.2 (±0.5)     | 0.8 (±1.0)   | 1.1 (±1.2)     | 0.1 (±0.4)          | 0.2 (±0.5)          |

²⁶Domains of the NtraC algorithm. Excluding or including the transition area (domain border) predicted by the NtraC model. Standard deviation in brackets.
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 preliminary in silico results that 43 out of the 90 (48%) long signal peptide sequences of long autotransporters 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. non-polar 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 tendency of long signal 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.
Figure 2. Multiple sequence alignment of 79 NraC-organized autotransporter signal sequences (Matlab R2009a, Bioinformatics Toolbox Version 3.3). Red: negatively charged residues. Blue: positively 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
1. 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.
2. Yen YT, Stathopoulos C. Identification of autotransporter proteins secreted by type V secretion systems in gram-negative bacteria. Methods Mol Biol. 2009;370:33–46.
3. Dautin N, Bernstein HD. Protein secretion in gram-negative bacteria via the autotransporter pathway. Annu Rev Microbiol. 2007;61:89–112.
4. Jacob-Dubuisson F, Fernandez R, Coutte L. Protein secretion through autotransporter and two-partner pathways. Biochim Biophys Acta. 2004;1694:235–57.
5. Wells TJ, Tree JJ, Ulett GC, Schembri MA. Autotransporter proteins: novel targets at the bacterial cell surface. FEMS Microbiol Lett. 2007;274:163–72.
6. von Heijne G. Signal sequences. The limits of variation. J Mol Biol. 1985;184:99–105.
7. Clérico EM, Szmyńska 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.
8. 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.
9. 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.
10. 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.
11. Jacob-Dubuisson F, Locht C, Antoine R. Two-partner secretion in Gram-negative bacteria: a thrifty, specific pathway for large virulence proteins. Mol Microbiol. 2001;40:306–13.
12. Wu CH, Apweiler R, Barich A, et al. The Universal Protein Resource (UniProtKB): an expanding universe of protein information. Nucleic Acids Res. 2006;34:D187–91.
13. Harte N, Silventoinen V, Quevillon E, et al. Public web-based services from the European Bioinformatics Institute. Nucleic Acids Res. 2004;32:W3–9.
14. Bendtsen JD, Nielsen H, Widdick D, Palmer T, Brunak S. Prediction of twin-arginine signal peptides. BMC Bioinformatics. 2005;6:167–76.
15. 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.
16. Emanuelsson O, von Heijne G, Schneider G. Analysis and prediction of mitochondrial targeting peptides. Meth. Cell Biol. 2001;65:175–87.
17. Bendtsen JD, Nielsen H, von Heijne G, Brunak S. Improved prediction of signal peptides: SignalP 3.0. J Mol Biol. 2004;340:783–95.
18. Brander W, Schmidt H, Karch H. EspT, a novel extracellular serine protease of enterohaemorrhagic Escherichia coli O157:H7 cleaves human coagulation factor V. Mol Microbiol. 1997;24:767–78.
19. 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.
20. 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.
21. 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.
22. Hammen PK, Weiner H. Mitochondrial leader sequences: structural similarities and sequence differences. J Exp Zool. 1998;282:280–3.
23. 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.
25. 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.
26. Puziss JW, Fikes JD, Bassford PJ Jr. Analysis of mutational alterations in the hydrophilic segment of the maltose-binding protein signal peptide. J Bacterial. 1989;171:235–57.
27. Dierstein R, Wickner W. Requirements for substrate recognition by bacterial leader peptidase. EMBO J. 1986;5:427–31.
28. 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.

| Signal Sequence | Description |
|-----------------|-------------|
| AAUA_ALCFA_1    | Tat-type signal. |
| MRWLDKFGEMLSRVSVAHKTSRSRVSRLRSVGLMVGSAPVPLPVPLVVARA | |
| AFB2_STRCX_1    | Tat-type signal. |
| MCTREAVRMSREHDPEIPSRRLKGGAAAGALTAVPGEAA | |
| CHOD_BREST_1    | Tat-type signal. |
| MTDSDANRADAATRGAVS5RVRFLAGPLTAGAIALSMSTASASA | |
| CHOD_STRS0_1    | Tat-type signal. |
| MTQQHLSRRRLMGMAAFGAAALAGTTTAAPRAAAAAAAAKSA | |
| CPRA_DESHA_1    | Tat-type signal. |
| MENNQKRQQGRSRLKFVGAATTMGVIAGAKPAKVA | |
| DHAQ_ACEPO_1    | Tat-type signal. |
| MRGLNRFLGKDGRREQASLSRGRFLVTSLGGAGVMGFARPSA | |
| DHLH_METEX_1    | Tat-type signal. |
| MLGKSDFFDKEMRSMRVAHGTSRSRGFIGRGTAVAGVLPPLPDRGRGSRASANA | |
| DHLH_PARDE_1    | Tat-type signal. |
| MLGNRFDVMEKLSRRAVGQTSRSRSVGLGTMALGIVPLPDRGRGSRASANA | |
| DHLH_PARVE_1    | Tat-type signal. |
| MLGNRFDVMEKLSRRAVGQTRGIGRGTATGIVPLPDRGRGSRASANA | |
| DMSA_ECOLI_1    | Tat-type signal. |
| MKTKIDAVLAASEVSRGVLTTAIGLAMASSAALTLPFSRIAHA | |
| DMSA_RHOCA_1    | Tat-type signal. |
| MTKFSQGLELSRLSFALSYVATLGGALGLGMAPKIAWA | |
| DMSA_RHOSH_1    | Tat-type signal. |
| MKLSEQHELHAELRSRLSFALSTYAAAGALGLGMAPKIAWA | |
| GADH3_PECGY_1   | Tat-type signal. |
| MSKHGNGTREDFTLRTITLAPAMAVSTGAMALVAPAAGA | |
| GFO_ZYMO_1      | Tat-type signal. |
| MTKISSSDNLNASMATPRTLALRAVGGVGLAAGALASGQLQ | |
| MBHS_AZOVI_1    | Tat-type signal. |
| MNQETFQYVMRQGTVRSLKYSQMTSLAAGLAPAFAPFRAHA | |
| MBHS_ECOLI_1    | Tat-type signal. |
| MNNEETFQYVMRQGTVRSLKYSQMTSLAAGLAPAFAPFRAHA | |
| MBHS_Olico_1    | Tat-type signal. |
| MPFETQYVMRQGTVRSLKYSQMTSLAAGLAPAFAPFRAHA | |
| MBHS_RALEH_1    | Tat-type signal. |
| MVEFETQYVMRQGTVRSLKYSQMTSLAAGLAPAFAPFRAHA | |
| NOSZ_AHCY_1     | Tat-type signal. |
| MSKJHKLRASLFTATAGSAILGTVGAALSLGAAGLATPARA | |
| NOSZ_PARDE_1    | Tat-type signal. |
| MSKJEKLRASLFTATAGSAILGTVGAALSLGAAGLATPARA | |
| NOSZ_PSEST_1    | Tat-type signal. |
| MSKJDSKNTPQVPEKGLGSRGLGASVTGAAATTLGAGAVMTRESWAQA | |
| PHNS1_DESCV_1   | Tat-type signal. |
| MRFSVGLGKGEAEERLARGRVSRRDFLKFCTAIAVTMGMGAPAEVAR | |
| PHNS1_DESCF_1   | Tat-type signal. |
| MNSVGLGKREAEKRLVQNGVSRDFLKFCTAIAVTMGMGAPAEVAR | |
| PHNS1_DESCM_1   | Tat-type signal. |
| MKISGLKGEVIERLAIERSGVRDFLKFCTAIAVTMGMGAPAEVAR | |
| PHNS1_STR_V_1   | Tat-type signal. |
| MQQGDFQDRIKQSPAPLNGMRSRGRGLLGAGTLTALATASAGLPLLPGTAHA | |
| PHNS2_ECOLI_1   | Tat-type signal. |
| MSNQQEYEPDNYKHEPDLLSLTRDILKSVAAATATAVVPHTLPLASVPA | |
| PHNS2_ECOLI_1   | Tat-type signal. |
| MKKQFNFKESDKTAEVSFPMRQVBLKALGISATALSPLPAAHA | |
| PHNS2_ECOLI_1   | Tat-type signal. |
| MSKNERMGISRRTLVSTAILSALAAAGGFLSPFLTRNAAAAA | |
Table S2. 228 short bacterial signal peptides associated with autransporters.

| Signal Peptide | Potentiality |
|----------------|--------------|
| >BIGA_SALTY_1 SIGNAL: Potential. MNMPQKKKLISIAALTLQSYIIPAIA | |
| >ESTA_PSEAR_1 SIGNAL: Potential. MRRMLKPLVAACLASSLSTAPQA | |
| >ESTA_PSEP0_1 SIGNAL: Potential. MRRMLKPLVAACLASSLSTAPQA | |
| >HAP1_HAEIN_1 SIGNAL: Potential. MKKTVFRLNFTACVSLGISASQANA | |
| >HAP2_HAEIN_1 SIGNAL: Potential. MKKTVFRLNFTACVSLGISASQANA | |
| >IGA0_HAEIN_1 SIGNAL: Potential. MLNKKFLNIALTAVAYLPTYYEA | |
| >IGA1_HAEIN_1 SIGNAL: Potential. MLNKKFLNIALTAVAYLPTYYEA | |
| >IGA3_HAEIN_1 SIGNAL: Potential. MLNKKFLNIALTAVAYLPTYYEA | |
| >IGA4_HAEIN_1 SIGNAL: Potential. MLNKKFLNIALTAVAYLPTYYEA | |
| >IGA_NEIGO_1 SIGNAL: MKAKRMKINAKSLSIFAYALTPYSEA | |
| >LIP1_PHOLU_1 SIGNAL: MKRSFIFAPGMLALSISAISNAHA | |
| >OMPA_RICCN_1 SIGNAL: Potential. MANISPKLFQAIQQGKFGYGLSLHSSLA | |
| >OMPA_RICFI_1 SIGNAL: Potential. MANISPKLFQAIQQGKFGYGLSLHSSLA | |
| >PERT_BOR1_1 SIGNAL: Potential. MNMMLRIVKAPPRGRTTLAMALGAAPAAHA | |
| >PERT_BOR2_1 SIGNAL: Potential. MNMMLRIVKAPPRGRTTLAMALGAAPAAYA | |
| >PERT_BORP_1 SIGNAL: Potential. MNMMLRIVKAPPRGRTTLAMALGAAPAAHA | |
| >PRTS_SERMA_1 SIGNAL: Potential. MILNKKKLYCFLYGSLISHSSLA | |
| >PRTT_SERMA_1 SIGNAL: By similarity. MILNKKKLYCFLYGSLISHSSLA | |
| >SCA1_RICCN_1 SIGNAL: Potential. MNKLTQHNLKSKRFLIESLALISISVGA | |
| >SCA1_RICFE_1 SIGNAL: Potential. MNKLTQHNLKSKRFLIESLALISISVGA | |
| >SCA2_RICCN_1 SIGNAL: Potential. MNLQNSHHKVKTLVTBMSTCLTLSSSLSTSA | |
| >SCA2_RICFE_1 SIGNAL: Potential. MLSQNSHHKVKTLVTBMSTCLTLSSSLSTSA | |
| >SCA2_RICS1_1 SIGNAL: Potential. MSTCLTLSSSLSTSA | |
| >SSA1_PASHA_1 SIGNAL: Potential. MYKIKHSFKNKLIAISISSSSLIA | |
| >VACA1_HELPP_1 SIGNAL: Potential. MEIQHTHRKINRPLVSLALVLGALVSITPPQOSHA | |
| >VACA2_HELPP_1 SIGNAL: Potential. MEIQHTHRKINRPLVSLALVLGALVSITPPQOSHA | |
| >VACA4_HELPP_1 SIGNAL: Potential. MEIQHTHRKINRPLVSLALVLGALVSITPPQOSHA | |
| >VACA7_HELPP_1 SIGNAL: Potential. MEIQHTHRKINRPLVSLALVLGALVSITPPQOSHA | |
| >YADA1_YEREN_1 SIGNAL: Potential. MTDFKISVSAALSLAFSSPYAFA | |
| >YADA2_YEREN_1 SIGNAL: Potential. MTDFKISVSAALSLAFSSPYAFA | |
| >YADA3_YEREN_1 SIGNAL: By similarity. MTDFKISVSAALSLAFSSPYAFA | |
| >YADA4_YEREN_1 SIGNAL: By similarity. MTDFKISVSAALSLAFSSPYAFA | |

(Table S2. (Continued))
Table S2. (Continued)

| Sequence | Signal | Potential |
|----------|--------|-----------|
| MTSNFRSLLAFAILTLTQTQIAK | Potential |
| Q3RAY9_XYLF A_1 | Potential |
| MKTIFAPARTIALAALASACGD | Potential |
| Q3R92_XYLF A_1 | Potential |
| MKNKRTTFPGPSLTVLTVFLASCGG | Potential |
| Q3RE25_XYLF A_1 | Potential |
| MKIKKNIKFIVRTILATTVTTLASCGG | Potential |
| Q3RE26_XYLF A_1 | Potential |
| MKTIAPARTIALAALASACGD | Potential |
| Q3RGS6_XYLF A_1 | Potential |
| MTSKFRSLLAFAILTLTQTQIAK | Potential |
| Q3RGT2_XYLF A_1 | Potential |
| MKTIFAPARTIALAALASACGD | Potential |
| Q3RGG2_XYLF A_1 | Potential |
| MKNKRTTFPGPSLTVLTVFLASCGG | Potential |
| Q3RGT6_XYLF A_1 | Potential |
| MKIKKNIKFIVRTILATTVTTLASCGG | Potential |
| Q3RY73_SALB N_1 | Potential |
| MNKIYALKYSVRQGALVPV | Potential |
| Q83YP8_ACTAC G_1 | Potential |
| MKIKKLFPRSVTTALTVFNPQAPA | Potential |
| Q93S98_BORPE _1 | Potential |
| MKNNSRSRKPAPRRLRTMLMALGAGAAPAHA | Potential |
| Q93S99_BORPE _1 | Potential |
| MKNNSRSRKPAPRRLRTMLMALGAGAAPAHA | Potential |
| A0B1Z1_BURCH_1 | Potential |
| MFLALSAGIVPAHATSCSTAG | Potential |
| A0LDC0_MAGSM_1 | Potential |
| MKRNSLTLTSLTTGLMGMTQPAV | Potential |
| A1JM6W_YERE8 G_1 | Potential |
| MKIYRIFKTMPVSCPVLGFPSVDA | Potential |
| Q1J5Q7_YERE8 G_1 | Potential |
| MNNINIALPCFRKTTLASLLVLPLAPL | Potential |
| A1JJS0_BURV G_1 | Potential |
| MNHRSPFSLRTGGRALHAILAGAVMPASS AQ | Potential |
| A4THP4_YERP P_1 | Potential |
| MNHMKVRLSAAVALLLSIGNSYAD | Potential |
| A4TNM13_YERP P_1 | Potential |
| MNKTNLTLRLLPLSILSLVSgggAV | Potential |
| A4TNM4_YERP P_1 | Potential |
| MKSRHHLNTRLLPLSILSLIPAVAALAA | Potential |
| A4TP40_YERP P_1 | Potential |
| MDKTLGAGAALLSLGVLPVQVLAF | Potential |
| A4T7Q3_YERP P_1 | Potential |
| MKSTNKLSPCFCRKMIALSLLVLPCSLP | Potential |
| A4W648_Ent39_1 | Potential |
| MKTTFRLOQVATSLISSSSVSVIPOTAL | Potential |
| A4W9S4_Ent39_1 | Potential |
| MIVCVVSCFGLASPAALAA | Potential |
| A4WEB5_Ent39_1 | Potential |
| MKKVLCSQILSLVASTAAGGLTTHALAV | Potential |
| A4WED8_Ent39_1 | Potential |
| MKKLKRSLCSQLCVLMSLTSLSAFAA | Potential |
| A4WQQ4_Ent39_1 | Potential |
| MSKTNTPPATTDRKVLLGSGSGAALSSAE | Potential |
| A4XFF0_NOVAD_1 | Potential |
| MKTNKSLGALAGAASAVAVGLAQAQQVAA | Potential |
| A4XXQ8_PSEMY G_1 | Potential |
| MALRVALIGLQLLQVATGVGAV | Potential |

Table S2. (Continued)

| Sequence | Signal | Potential |
|----------|--------|-----------|
| >A5FVP6_ACICJ_1 | Potential |
| MKSRRFSLLAATALSLSACSG | Potential |
| >A5FX02_ACICJ_1 | Potential |
| MGSEMTTFTPCCVRLLAIGALAAGFGLAPAR | Potential |
| >A5V6F0_SPHW_1 | Potential |
| MRYLLASTCLAIAAIAVFPVAE | Potential |
| >A5VXL2_PSEP1_1 | Potential |
| MKRAPLRRFTLIALALACSAQAL | Potential |
| >A5W1C9_PSEP1_1 | Potential |
| MKSTSNFLRFDSIFYAVSTSILLLATPVETIAY | Potential |
| >A5W39_PSEP1_1 | Potential |
| MRLRMFLGLSLPLGLMVTFAQAN | Potential |
| >A5W741_PSEP1_1 | Potential |
| MGIVKQORRGLVRAKVMSALMLSPIAQL | Potential |
| >A5UC19_SINMRV_1 | Potential |
| MvA4VRKCFSTVLSGALLGLPFCVQGSEASAA | Potential |
| A6X2Y0_OCHAA_1 | Potential |
| MTGVLRHKSMLMTTAALGFPYTATG | Potential |
| A6X4K3_OCHAA_1 | Potential |
| MCRRNIGSARLRFSLTALSSLGTQLQSTPGMAA | Potential |
| A6X7J0_OCHAA_1 | Potential |
| MKKLWLASTAISLPSFSAWASA | Potential |
| A6X8H2_OCHAA_1 | Potential |
| MLKRLNGKVLFRFLPLFSLGATLAMPTVLQA | Potential |
| A7H7Z2_PARLI_1 | Potential |
| MTARNRTAAARRHHIALMLGTLALAPHSQGASD | Potential |
| A7IL68_XANP2_1 | Potential |
| MPALYASSRSLVLSLFLVMAAAPALAA | Potential |
| A8G7T5_SERP5_1 | Potential |
| MPLKTRMPRPRLAVAILCMMNTSMTSLAY | Potential |
| A8GG96_SERP5_1 | Potential |
| MASFPASFSRGCCALLATGTGSFPAVVNA | Potential |
| A8GHU3_SERP5_1 | Potential |
| MKDKSISHLAVRPKLSIYALAPSFAPLMLMDSAMAR | Potential |
| A8GIR8_SERP5_1 | Potential |
| MKRLALAI11AALPFCSAQAV | Potential |
| A8GL40_SERP5_1 | Potential |
| MKQATQKINLAPWTKINALLCALLAAGGQVA | Potential |
| >B0KJ8B_PSEP1_1 | Potential |
| MKRAPLRRFTLIALALACSAQAA | Potential |
| >B0KM3U_PSEP1_1 | Potential |
| MKSTSNFLRFDSIFYAVSTSILLLATPVETFAY | Potential |
| >B0KI28_PSEP1_1 | Potential |
| MRCRLLLLPLVPLLLTPLLHAHQ | Potential |
| >B0KUR2_PSEP1_1 | Potential |
| MKRNNFSTLSPLASAOQGLLGSASVLFPGAGAL | Potential |
| >B0KUS7_PSEP1_1 | Potential |
| MQVRKLVATVAPLMAFAQY | Potential |
| >B0T124_CAUSK_1 | Potential |
| MMRTRSKRTILAGSSSLVMAIAAAAQPALAQ | Potential |
| >B0U200_XYLFM_1 | Potential |
| MKNMKTTPFPFGISILVTLTLVAPLSACGG |
|----------------------------------|
| >B0U225_XYLFM_1 SIGNAL: Potential. |
| >B0U2J0_XYLFM_1 SIGNAL: Potential. |
| >B0U5B1_XYLFM_1 SIGNAL: Potential. |
| >B0U5U7_XYLFM_1 SIGNAL: Potential. |
| MERKKHKTTATLISVVLMGGAGATYAN |
| >B0UED5_METS4_1 SIGNAL: Potential. |
| MRFLSGVSLAIAVITAIMGVGAARAA |
| >B1IV2_XECOL_1 SIGNAL: Potential. |
| MRTPSPYCRSSVLSSLISLIAYPFGMAA |
| >B1IY0_ECOL_1 SIGNAL: Potential. |
| MHSWKKKLVSQLQQALTACTTSQANAA |
| >B1J4X2_PSEPW_1 SIGNAL: Potential. |
| MPSFSVPVRHLVHLSVAPK PLLTFPCA |
| >B1J6X8_PSEPW_1 SIGNAL: Potential. |
| MLQRFCSLSSLTLAAIASHHAA |
| >B1J6R7_PSEPW_1 SIGNAL: Potential. |
| MKSTSNMRFDRIFVYAVSSTMLATTPVET |
| >B1JE37_PSEPW_1 SIGNAL: Potential. |
| MKAPLRLFTTLATLACRQAA |
| >B1JGR2_YERPY_1 SIGNAL: Potential. |
| MDTKLAIASLSLTVLPQVVLAF |
| >B1JHT7_YERPY_1 SIGNAL: Potential. |
| MKRRNSTLSPCFRTKTLIASLVLPSLYS |
| >B1JLC5_YERPY_1 SIGNAL: Potential. |
| MNNHKWRLSVAVALLISGGYAD |
| >B1JPQ0_YERPY_1 SIGNAL: Potential. |
| MNNHKWRLSVAVALLISGGYAD |
| >B1JS9_YERPY_1 SIGNAL: Potential. |
| MNNSKKTYLSSILSILSILSILSIL |
| >B1LT2_METRJ_1 SIGNAL: Potential. |
| MTRGYYVSLAALAVGLGPGQAQ |
| >B1YP8_BURCJ_1 SIGNAL: Potential. |
| MNRRFPFSOARSSRGRLHSALIAGAVPVS |
| >B2FHR8_STRM5_1 SIGNAL: Potential. |
| MRMQFTPFPFSAKQDXDLARRIAATALLATSGAGA |
| >B2F2I2_STRM5_1 SIGNAL: Potential. |
| MNHPHGRSSHSPHRLSRLAVSSSLLLLAAAPA |
| >B2F5VS_STRM5_1 SIGNAL: Potential. |
| MERTMVRSVLATMALALAC |
| >B2FSC8_STRM5_1 SIGNAL: Potential. |
| MLLSRKPIRMLAAAILAALAPA |
| >B2FU91_STRM5_1 SIGNAL: Potential. |
| MKHSKLALAGLIAVAGAA |
| >B2FU94_STRM5_1 SIGNAL: Potential. |
| MYARVPRAFRRPLASVQLALAVPSLLTTGSCVA |
| >B2I6G5_XYLF2_1 SIGNAL: Potential. |
| MLKLFQKRFKFLTVVIFPSMYGGSVYVA |
| >B2I753_XYLF2_1 SIGNAL: Potential. |
| MKNMKTAPPFGISILVTLTLVAPLSACGG |

**Table S2.** (Continued)
Table S2. (Continued)

| Signal Sequence | Signal: Potential |
|-----------------|-------------------|
| MPTPQNSPIAIAVSAALASMVFPQQA | |
| >B5RG96_SALG2_1 | Potential. |
| MIVRKCRGRTLCCAGLMACSFFINTTYA | |
| >B8ENQ5_METBO_1 | Potential. |
| MDSFNLKSLAAILLGVPLLLPASQLAA | |
| >B9ILY6_METBO_1 | Potential. |
| MLGCESGRAHGLDRMLRLLLVFGAPLVLATSALQ | |
| >O86163_MORDE_1 | Potential. |
| NHYNQMMRRATPGRGSHGLALLGAGMWTLSPPSAWA | |
| >Q07JX7_ROHOP_1 | Potential. |
| MATNRKRGRTVLAAVAIGAPIAAPTFDAR | |
| MLTHARRNGAPRSTGWLLASLTLAIAVAYAPLSLAQ | |
| >Q0AKS3_MORM1_1 | Potential. |
| MRRFLPASVLTVPTTAPAAD | |
| >Q0AT43_MORM1_1 | Potential. |
| MAVNVGVRADAKAGPVRVSLMLAGALGFSAPVAA | |
| >Q11GI1_MESSB_1 | Potential. |
| MKTAYKLGLSIALAATAVSAQHAQ | |
| >Q12FI0_POLSJ_1 | Potential. |
| MTLNYSFYSFMQKPYRLSTAAACLAFLASFAQQA | |
| >Q132Z9_ROHOP_1 | Potential. |
| MQRMTLPARIWLSRAALASASSAPQ | |
| >Q1397_ROHOP_1 | Potential. |
| MGAHFRNLSTFLCCTTFLVSAPVSAALYAA | |
| >Q1C198_YERP_1 | Potential. |
| MKTNRTLSPCRFTKMASSLVLVPCPLSPLSWAV | |
| >Q1C326_YERP_1 | Potential. |
| MNHKIWRSAVALLISGNASY | |
| >Q1C4W8_YERP_1 | Potential. |
| MTDKTLAGAILSSLVSILVPQVGLAF | |
| >Q1C510_YERP_1 | Potential. |
| MMRHHLTRLPISLILASPAVLLAA | |
| >Q1C511_YERP_1 | Potential. |
| MKNSNLTRLPSLILSLSVSGGAMAV | |
| >Q1CAV8_YERP_1 | Potential. |
| MNHKFKNLTAISAIALLLSVSFSNTLAV | |
| >Q1CDG9_YERP_1 | Potential. |
| MKDNRTLSPCRFTKMASSLVLVPCPLSPLSWAV | |
| >Q1CF77_YERP_1 | Potential. |
| MNHKFKNLTAISAIALLLSVSFSNTLAV | |
| >Q1CK98_YERP_1 | Potential. |
| MKNSNLTRLPSLILSLSVSGGAMAV | |
| >Q1CK99_YERP_1 | Potential. |
| MMRHHLTRLPISLILASPAVLLAA | |
| >Q1CKV3_YERP_1 | Potential. |
| MTDKTLAGAILSSLVSILVPQVGLAF | |
| >Q1CMH9_YERP_1 | Potential. |
| MNHKIWRSAVALLISGNASY | |
| >Q1GU46_SPHAL_1 | Potential. |
| MKTLLACATLLSTAVHAE | |

(Continued)

Table S2. (Continued)

| Signal Sequence | Signal: Potential |
|-----------------|-------------------|
| >Q1MHZ0_ROHIL3_1 | Potential. |
| MRYRKLASAVGHVGLATIFALPAALMFLDAYG | |
| >Q1MK18_ROHIL3_1 | Potential. |
| MSPFCGSPRTLVSLLIPTTGMD | |
| >Q1QCE5_PSYCK_1 | Potential. |
| MRPNISAVTPTTTLKLTCLASLMLASMAQGAA | |
| >Q214Q0_ROHBP_1 | Potential. |
| MGAFGRDVSKLLCTTFVLAVAPVSQALAA | |
| >Q2GAY1_NOVAD_1 | Potential. |
| MDRRTTTILSTLGTAFVLALLVPQANAA | |
| >Q2J1P8_ROHBP_1 | Potential. |
| MQRKARTLIAGFAMATSVSTGAAC | |
| >Q2KTY1_BORA1_1 | Potential. |
| MTOPRSRPSHALVLSLSSLA | |
| >Q2Y820_NITMU_1 | Potential. |
| MAIRKKSALSLYLHFIALLMVPASLSRSAQ | |
| >Q3BVT0_YANC5_1 | Potential. |
| MKKOaVTVLASALVAATTC | |
| >Q3K4D0_PSEPF_1 | Potential. |
| MPIQCKYKLQHLVLAVALVCVESLSL | |
| >Q3K4D1_PSEPF_1 | Potential. |
| MPPFQRLSPFAIALLISAAHKG | |
| >Q3K5V0_PSEPF_1 | Potential. |
| MKQTFLPVLPCALLMACAQANAA | |
| >Q3KC67_PSEPF_1 | Potential. |
| MQKSKCQYRLDRWGLVLSFLAPFSQIAIG | |
| >Q3KCT0_PSEPF_1 | Potential. |
| MKNNNTPAQSGGGFKLKTVLAVCAMATWSSAHA | |
| >Q3KCT1_PSEPF_1 | Potential. |
| MDVRKIPISVTLTVLSIAATQAQAA | |
| >Q3KD77_PSEPF_1 | Potential. |
| MKTSQSEEEKIPTFCVSSILSCSSMEEAQ | |
| >Q3KD66_PSEPF_1 | Potential. |
| MFPRFCSLSLVSLSIAAVHAA | |
| >Q3SN6_PITW_1 | Potential. |
| MNNVVRAPMNGALRRVRVAGAFALSSAGAA | |
| >Q4ZM16_PSEU2_1 | Potential. |
| MTKTRNWPAECALLSSSCLAGTAAA | |
| >Q4ZQ07_PSEU2_1 | Potential. |
| MQKSKCGVVRYSFKPVTGALCSFTFCGSAAYD | |
| >Q4UU5_PSEU2_1 | Potential. |
| MNAFFVLPLKTLTVTPSLPLPSGAFA | |
| >Q4YT2_PSEU2_1 | Potential. |
| MFPRFTLAMAATAVPCACE | |
| >Q6645_YERP_1 | Potential. |
| MKQRNTLSPCRFTKMASSLVLVPLSPLSWAV | |
| >Q665R2_YERP_1 | Potential. |
| MNNKIKWRSVALLISRSN | |
| >Q667C1_YERP_1 | Potential. |
| MKKKQDVARTVLASALVAMATWSSAHA | |
| >Q667Z0_YERP_1 | Potential. |
| MNNKIKWRSVALLISRSN | |
| >Q667Z1_YERP_1 | Potential. |
| MKKKQDVARTVLASALVAMATWSSAHA | |

(Continued)
Table S2. (Continued)

MKRSNLNLPLLISLISISLPLGAVLAA
Q667Z2_YERPS_1 SIGNAL: Potential.
MKNSNMLNLPLLISLISLSVSGAMAV
Q66DI5_YERPS_1 SIGNAL: Potential.
MDKTLGKAIISLSLVTLPQVLAF
Q6N1B5_RHOPA_1 SIGNAL: Potential.
MRRAASCQSACLPTVIPLSIAE
Q6N8G7_RHOPA_1 SIGNAL: Potential.
MSTVGRFRHLSSLCTTFLVSAPMSAVLYAA

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