

# Bioinformatic analysis of the TonB protein family

Byron C. H. Chu · R. Sean Peacock ·  
Hans J. Vogel

Received: 22 September 2006 / Accepted: 28 November 2006 / Published online: 16 January 2007  
© Springer Science+Business Media B.V. 2007

**Abstract** TonB is a protein prevalent in a large number of Gram-negative bacteria that is believed to be responsible for the energy transduction component in the import of ferric iron complexes and vitamin B<sub>12</sub> across the outer membrane. We have analyzed all the TonB proteins that are currently contained in the Entrez database and have identified nine different clusters based on its conserved 90-residue C-terminal domain amino acid sequence. The vast majority of the proteins contained a single predicted cytoplasmic transmembrane domain; however, nine of the TonB proteins encompass a ~290 amino acid N-terminal extension homologous to the MecR1 protein, which is composed of three additional predicted transmembrane helices. The periplasmic linker region, which is located between the N-terminal domain and the C-terminal domain, is extremely variable both in length (22–283 amino acids) and in proline content, indicating that a Pro-rich domain is not a required feature for all TonB proteins. The secondary structure of the C-terminal domain is found to be well preserved across all families, with

the most variable region being between the second  $\alpha$ -helix and the third  $\beta$ -strand of the antiparallel  $\beta$ -sheet. The fourth  $\beta$ -strand found in the solution structure of the *Escherichia coli* TonB C-terminal domain is not a well conserved feature in TonB proteins in most of the clusters. Interestingly, several of the TonB proteins contained two C-terminal domains in series. This analysis provides a framework for future structure-function studies of TonB, and it draws attention to the unusual features of several TonB proteins.

**Keywords** Gram-negative bacteria · Iron transport · TonB · Multiple sequence alignment

## Abbreviations

CTD	carboxy-terminal domain
DSS	2,2-dimethyl-2-silapentane-5-sulfonate
HSQC	Heteronuclear single quantum coherence
MSA	multiple sequence alignment
NTD	amino-terminal domain
NOE	nuclear overhauser effect
OMT	outer membrane transporter

Byron C. H. Chu and R. Sean Peacock contributed equally to this work.

B. C. H. Chu · R. S. Peacock · H. J. Vogel (✉)  
Structural Biology Research Group, Department of  
Biological Sciences, University of Calgary, Calgary,  
Alberta, Canada, T2N 1N4  
e-mail: vogel@ucalgary.ca

## Introduction

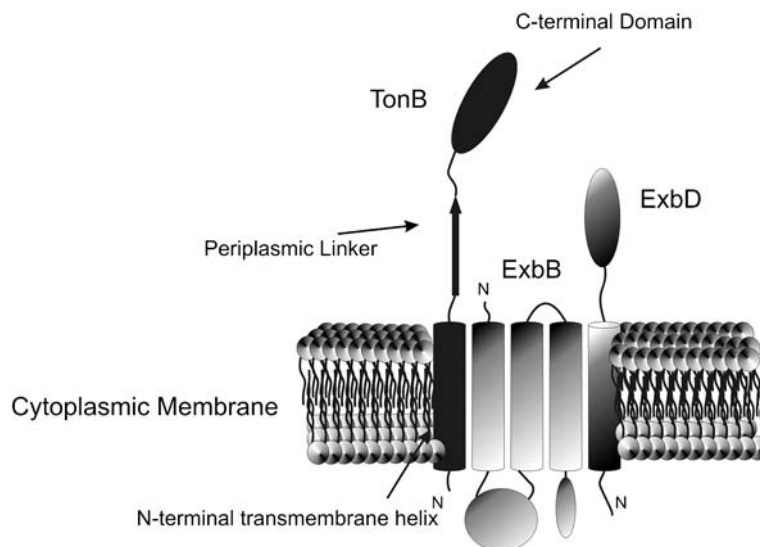
Iron is essential for the survival of almost all living organisms (Ratledge and Dover 2000). The lynch-

pin of the iron uptake system in Gram-negative bacteria is the TonB protein which, with the help of the ExbB/ExbD complex, transduces the energy of the cytoplasmic membrane proton motive force across the periplasmic space to the Outer Membrane Transporters (OMTs) of a variety of iron-ligand complexes. This allows for passage of the iron loaded ligand into the periplasmic space (recently reviewed in (Braun and Braun 2002; Postle and Kadner 2003; Wiener 2005)). Iron withdrawal is a well established method of inhibiting bacterial growth (Jurado 1997; Weinberg 1984), and disruption of the TonB gene is known to decrease virulence in several organisms (Stork et al. 2004; Beddek et al. 2004; Torres et al. 2001).

The TonB protein from *Escherichia coli* is thought to have three domains (Fig. 1). Topology studies have shown that *E. coli* TonB has a short cytoplasmic domain consisting of 12 residues, followed by a single  $\alpha$ -helical transmembrane domain ending approximately at residue 32 (Roof et al. 1991). Residues 33–150 form a linker domain that includes a Pro-rich region (Evans et al. 1986) and a flexible region (Peacock et al. 2005). The last 90 residues form a structurally well

defined C-terminal domain (CTD) (Peacock et al. 2005; Koedding et al. 2005)

Although many studies have been published about the TonB protein, the mechanism by which it promotes ligand transport through OMTs is still very much unknown. There are some indications that TonB does not remain anchored to the cytoplasmic membrane during the import of the ferric ligand (Letain and Postle 1997; Larsen et al. 2003). It is also not known if the TonB protein dislodges the cork domain from the OMT completely, partially or whether it only causes smaller scale structural rearrangements (Klebba 2003; Eisenhauer et al. 2005; Chimento et al. 2005). Even the oligomeric state of the active form of the protein is a matter of considerable debate (Chang et al. 2001; Koedding et al. 2004; Khursigara et al. 2004; Ghosh and Postle 2004; Sauter et al. 2003). Our studies have shown that a monomeric form of *E. coli* TonB-CTD can bind to give a 1:1 complex with several TonB box regions of various OMTs (Peacock et al. 2005). This was later confirmed by two crystal structures of two different constructs of the periplasmic domain of TonB in complex with the OMT



**Fig. 1** Topology diagram of the TonB/ExbB/ExbD system. The three main functional domains of the TonB protein are indicated. The proline-rich domain is represented by a filled arrow. The N-terminus of the three proteins is shown for clarity. The topology was determined in all cases for the *E. coli* proteins: TonB (Roof et al.

1991), ExbD (Kampfenkel and Braun 1992) and ExbB (Kampfenkel and Braun 1993). The stoichiometry of the proteins in this complex has been estimated to be 1:2:7 TonB:ExbD:ExbB, but is not shown here for clarity (Higgs et al. 2002)

proteins FhuA and BtuB (Pawelek et al. 2006; Shultis et al. 2006). Although physiologically relevant dimeric forms may exist in vivo (Ghosh and Postle 2004), a truncated TonB-CTD that forms dimers in solution does not interact with FhuA at all (Koedding et al. 2004). This suggests that if TonB dimerization is important to the transport mechanism then it is probably after OMT recognition and binding.

In this report we present an analysis of the TonB sequences we have found in the NCBI Entrez database. Our analysis shows that the only well conserved elements of TonB are the single predicted transmembrane helix which is usually located at the N-terminus, and a well defined 90 amino acid C-terminal region.

## Materials and methods

### Sequence retrieval

We examined all the completed and annotated microbial genomes from the Entrez Genome Project (<http://www.ncbi.nlm.nih.gov/genomes/lproks.cgi>). From these we sorted for Gram-negative bacteria and searched for annotated TonB proteins in all the completed genomes (as of August 10, 2006). We then verified the completeness of the list by performing a PSI-BLAST search with three iterations (Altschul et al. 1997), adding in any sequences of completely sequenced genomes that the annotation search had missed.

### Sequence analysis

All the retrieved sequences were entered into a VectorNTI database (Invitrogen). The proteins were subjected to multiple sequence alignment (MSA) using default Clustal X (Thompson et al. 1997) settings and then subdivided into N-terminal, periplasmic linker and C-terminal domains using the amino acid sequence of the *E. coli* TonB as a guide. The N-terminal domains were then aligned in the Clustal X software using the default settings. The C-terminal domain was aligned using a gap opening penalty of 8 to accommodate some of the variable loops that were found in the TonB sequences using the Clustal W algorithm. The phylogenetic trees were created using a

Neighbor-Joining Bootstrap tree with 1,000 bootstrap trials in Clustal X based on the MSA. The phylogenetic trees were then visualized in the Tree View program (<http://taxonomy.zooloogy.gla.ac.uk/rod/treeview.html>). N and C-terminal domain clusters were assigned based on the phylogenetic trees.

The N-terminal domains were analyzed for predicted transmembrane helices using the Ave-Has program (Zhai and Saier 2001) with the multiple sequence alignments of the N-terminal domains subdivided into three clusters. The average hydropathy and similarity values at each position of the MSA were obtained and plotted in Microsoft Excel. For the MecR1 containing TonB proteins, a profile based alignment procedure was performed with the BlaR1 protein from *Bacillus licheniformis* used as the profile. Secondary structural elements were obtained from the numbering from the BlaR1-sensor domain crystal structure (Birck et al. 2004).

The C-terminal domains were subdivided into nine total clusters and then analyzed for secondary structure using the Jpred webserver with the Jalign algorithm (Cuff and Barton 2000). In addition, individual representative sequences were subjected to secondary structure prediction using the Jnet algorithm.

### Expression and purification of *E. coli* TonB-CTD and NMR spectroscopy

<sup>15</sup>N labeled *E. coli* TonB-CTD (residues 103–239) was expressed and purified as described previously (Peacock et al. 2005). EP and KP peptides consisting of TonB residues 70–83 (EP peptide: EPEPEPEPIPEPPK) and residues 84–102 (KP peptide: EAPVVIEKPKPKPKPKPKP) were synthesized by Anaspec (San Jose, CA). The peptides were over 98% pure as determined by analytical HPLC and mass spectrometry.

<sup>15</sup>N labeled *E. coli* TonB-CTD was concentrated to 0.5 mM (8 mg/ml). A 500 µl 10% D<sub>2</sub>O sample in 10 mM sodium phosphate pH 7.0 was added to an NMR tube. Experiments were performed at 25°C on a Bruker Avance 500 MHz Spectrometer with a triple resonance inverse TXI cryoprobe. All data was processed using NMRPipe (Delaglio et al. 1995) and

analyzed with NMRVIEW (Johnson and Blevins 1994).  $^1\text{H}$ ,  $^{15}\text{N}$  HSQC spectra were acquired with a flip back pulse for water suppression, and steady state NOE spectra were acquired as described previously (Kay et al. 1989). Proton chemical shifts were referenced to 2,2-dimethyl-2-silapentane-5-sulfonate (DSS) as 0 ppm and  $^{15}\text{N}$  chemical shifts were referenced indirectly to DSS. Interactions of TonB-CTD with the Pro-rich region peptides were studied by adding a stoichiometric amount of peptide to the TonB sample.

## Results

Initial BLAST searches with both the full-length and the C-terminal domain of the *E. coli* TonB protein (AAC74334) showed that TonB sequences are only found in Gram-negative bacteria. Using the Entrez Genome database we searched 217 completely sequenced Gram-negative bacterial genomes for annotated TonB proteins (August 10, 2006). Manual curation of these genomes yielded 219 TonB sequences from 124 organisms with unique taxonomic IDs. Additional PSI-BLAST searching added 26 TonB sequences and six organisms. A further 18 sequences from 14 organisms were added to the TonB database based on manual examination of the TonB literature. These additional organisms do not have completely sequenced genomes available but their TonB proteins have been examined experimentally and are important to our current understanding of TonB structure and function. In total, 263 TonB sequences from 144 gram-negative bacteria were examined in this study. Members of the *Proteobacteria*, *Cyanobacteria*, *Bacteroidetes/Chlorobi*, *Chlamydiae/Verrucomicrobia* and *Fusobacteria* phyla are all represented. These TonB proteins are extremely variable in length, ranging from 170 residues for one of the four TonBs in *Bacteroides fragilis*, to 700 residues in one of the four TonBs in *Colwellia psycherythraea*. However, the majority of the TonB proteins are between 225 and 300 residues long (Fig. 2A). Approximately 42% of these organisms have more than one TonB protein encoded in the genome, with numbers ranging between two and nine unique copies of

different TonB proteins within the same organism (Fig. 3). The organism that holds the record to date with nine TonBs is *Pseudomonas syringae* pv. *syringae* B728a. The majority of organisms with multiple TonBs have either two or three (67%) annotated TonBs.

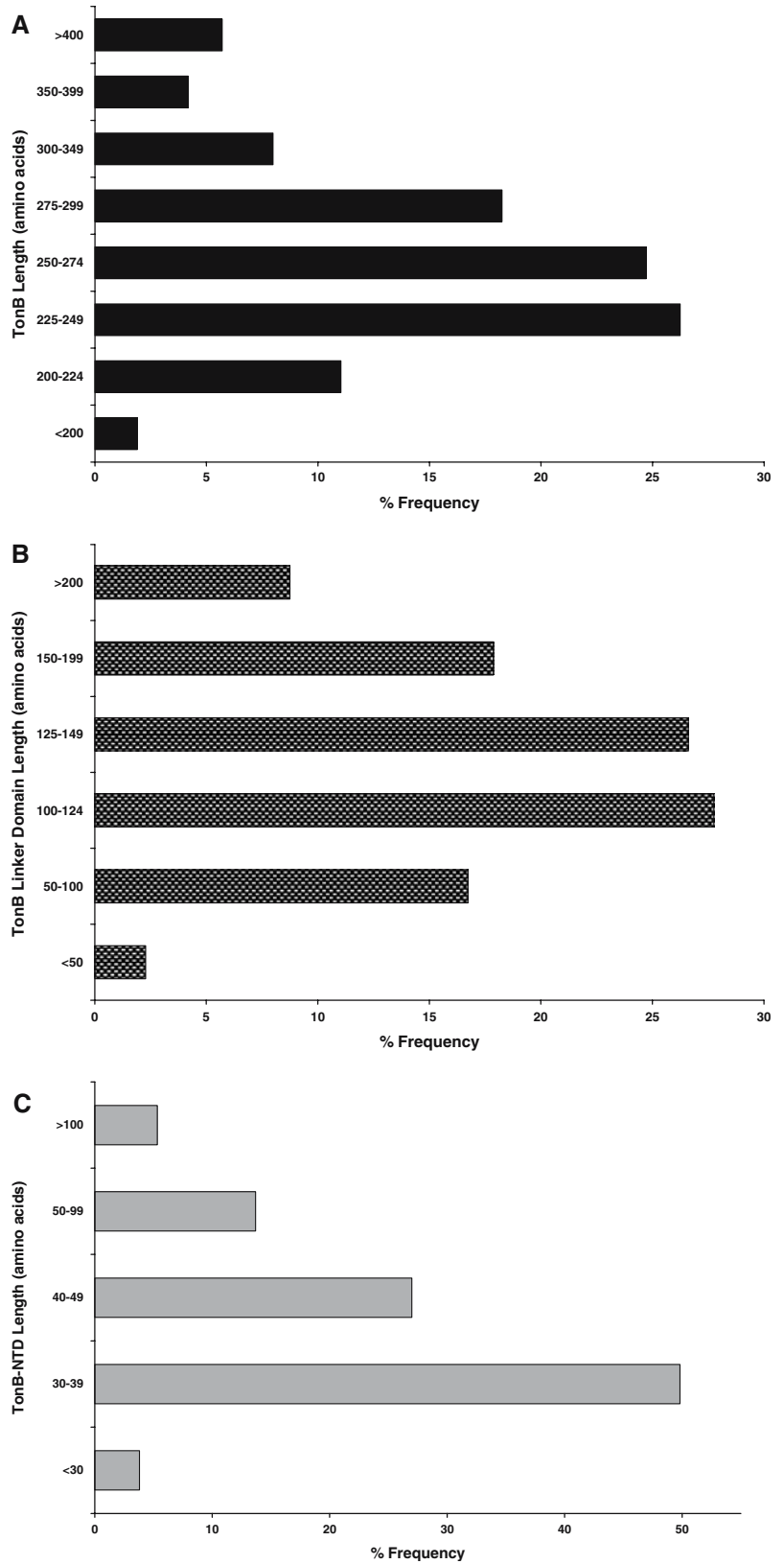
Based on the currently available genome sequence data and annotation, approximately a third of Gram-negative bacteria appear to not have a TonB. Members of the *Chlamydia*, *Chlamydophila*, *Chromohalobacter*, *Coxiella*, *Ehrlichia*, *Francisella*, *Geobacter*, *Halobacter*, *Janaschia*, *Lawsonia*, *Leifsonia*, *Legionella*, *Rickettsia*, *Wigglesworthia* or *Wolbachia* genus were not found to have any known TonB proteins. *Wollicella succinogenes* had one protein annotated as TonB that was 105 amino acids long and contained no apparent N-terminal or periplasmic linker domain, which we excluded from our analyses. Additionally, some species have annotated OMT receptors in the Entrez database, but no obvious TonB orthologues. It is unclear whether any of these receptors are functional, or if and how they might be activated. At least one organism, *Aquifex aeolicus*, has annotated ExbB and ExbD genes, but no TonB protein. Perhaps this organism, or the organisms with OMT proteins lacking TonB, encodes a protein that is too divergent from other known TonB sequences to be recognized by BLAST, that fulfils the same function.

TonB is generally acknowledged to contain three domains—an N-terminal domain containing a single transmembrane helix, a periplasmic linker domain containing a Pro-rich domain, and a C-terminal domain for which a large amount of biophysical and structural information is available. Our attempts to do phylogenetic analysis on TonB using full-length sequences were unsuccessful due to the hypervariable linker region. Therefore, we subdivided each of the proteins into three domains using the *E. coli* TonB protein as a guide and all further analysis was done based on these domain divisions.

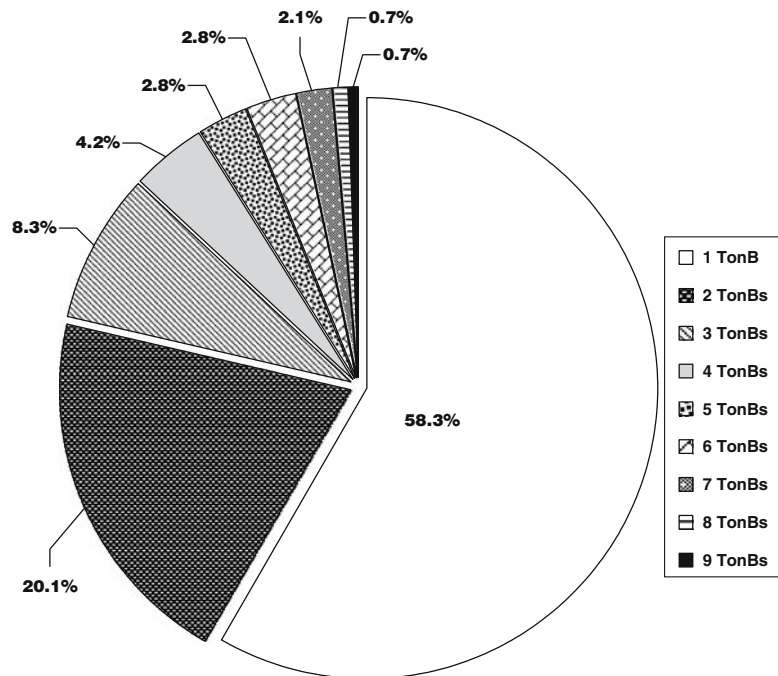
### C-terminal domain

The C-terminal domain is the most well studied and best conserved portion of the TonB protein.

**Fig. 2** Histograms of overall TonB length (A), and TonB periplasmic linker (B) and N-terminal (C) domain lengths (amino acid residues) from the 263 analyzed TonB sequences



**Fig. 3** A total of 144 Gram-negative bacteria were found to possess an annotated TonB. Of these, 60 bacteria had more than one annotated TonB. The distribution frequency of TonBs in these Gram-negative bacteria ranges between 1 and 9 as indicated above



In *E. coli* the structured C-terminus appears to span approximately the last 90 amino acids of the protein (Peacock et al. 2005; Koedding et al. 2005). Phylogenetic analysis of the C-terminal domains yielded nine clusters (Fig. 4). Organization of these clusters does not appear to be dependent on the taxonomy of the organism. Bacteria from different taxonomy group together within certain clusters, and TonBs from the same species can be found throughout several different CTD-domain clusters. For example, *Pseudomonas syringae* pv. *syringae* B728ABh has nine different TonB proteins that classify to clusters 1C, 2B, 2C and 4. Representative protein sequences for each CTD cluster are listed in Table 1, and an exhaustive list can be found on the web at <http://groningen.bio.ucalgary.ca/vogel/tonb>.

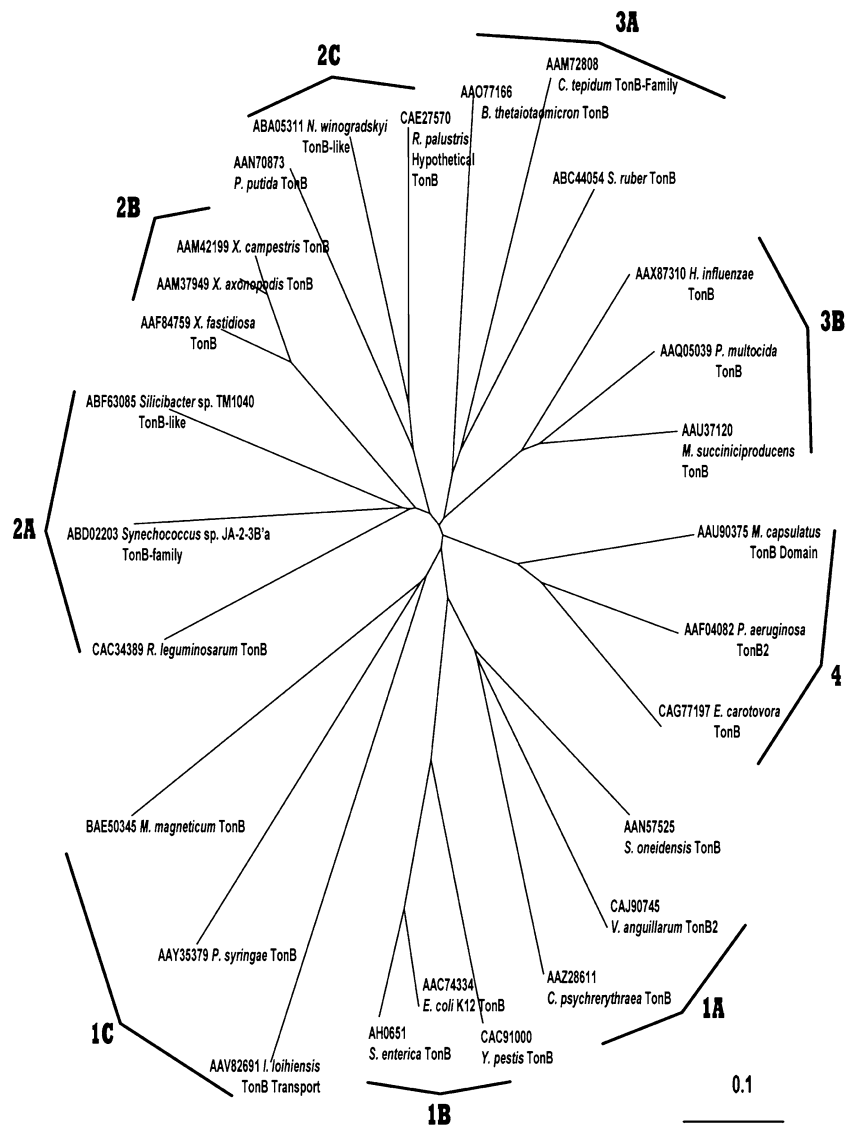
All clusters except one have a predicted secondary structure consisting of 3  $\beta$ -strands and 2  $\alpha$ -helices as determined by Jpred (Fig. 5). Cluster 1C is the only exception as only the extreme C-terminal  $\beta$ -strand and the amphipathic  $\alpha$ -helix are predicted. Several of the clusters (2A, 2C and 3B) often predict to have a longer  $\alpha$ -helix at the N-terminus of this domain as compared to the *E. coli* TonB. This is not dissimilar to the

related TolA protein (Lubkowski et al. 1999; Witty et al. 2002).

The most notable differences across the rest of the families appear to be a variable loop size between  $\alpha$ -helix 2 and  $\beta$ -strand 3 (Fig. 6). The exact alignment of these loops is not always inherently obvious, and is highly affected by the Gap opening penalty applied in Clustal X. However, we know that this is not an artifact of the alignments, as the solution structure of TonB2 from *Vibrio anguillarum* has confirmed the existence of this extra long loop (Peacock et al. submitted). In fact, the *E. coli* TonB has one of the shortest loops of all the known TonBs, and the loop extension appears to be a general feature of many TonB proteins.

Several TonBs have relatively lengthy C-terminal extensions of approximately 30–190 residues, for example the TonB from *Salinbacter ruber* (ABC44054) displays this feature. Generally the sequence in this extension is unremarkable, however three TonB proteins unexpectedly contain a second complete TonB C-terminal domain (Fig. 7). The examples of this are from *Bacteroides fragilis* (CAH09447 and BAD50734) and *Bacteroides thetaictamicron* (AAO79003). Although the sequences of the two fused TonB

**Fig. 4** Phylogenetic tree diagram of TonB C-terminal domains (CTDs). Three seed sequences from each cluster were used to improve visual clarity. The clusters (1A–C, 2A–C, 3A–B and 4) were analyzed for sequence conservation and are indicated on the diagram. The trees are based on CLUSTAL X-derived multiple sequence alignments (MSA). The tree was drawn with the TreeView program



CTDs are not identical, they do belong to the same clusters in our phylogenetic analysis.

The most highly conserved features across all TonB families is the “YP motif”, found along the extension from the beginning of the known structure of the C-terminal domain of TonB and  $\alpha$ -helix 1, as well as the amphipathic  $\alpha$ -helix 2 (Fig. 8A). The tyrosine from this motif is in close contact with the TonB box of the OMTs as seen in the recently published crystal structures of the FhuA-TonB periplasmic domain (Pawelek et al. 2006) and BtuB-TonB-CTD complexes (Shultis

et al. 2006), as well as the solution structure of TonB-CTD bound to a 10mer peptide of the FhuA TonB-box (Peacock et al. in preparation). The alignments of the TonB-CTDs are more variable further towards the N-terminus of the protein. This could be in part due to the differences in predicted secondary structure in this region, and could accommodate for longer  $\alpha$ -helices in this area of the protein.

Many TonBs (not including the *E. coli* version) have a conserved SSG motif in the loop joining  $\beta$ -strand 2 to  $\alpha$ -helix 2. Although this

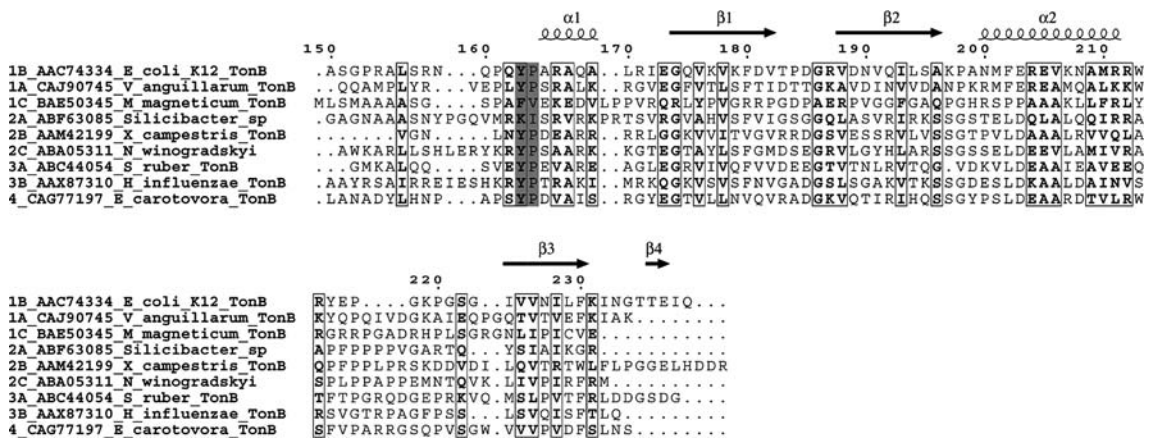
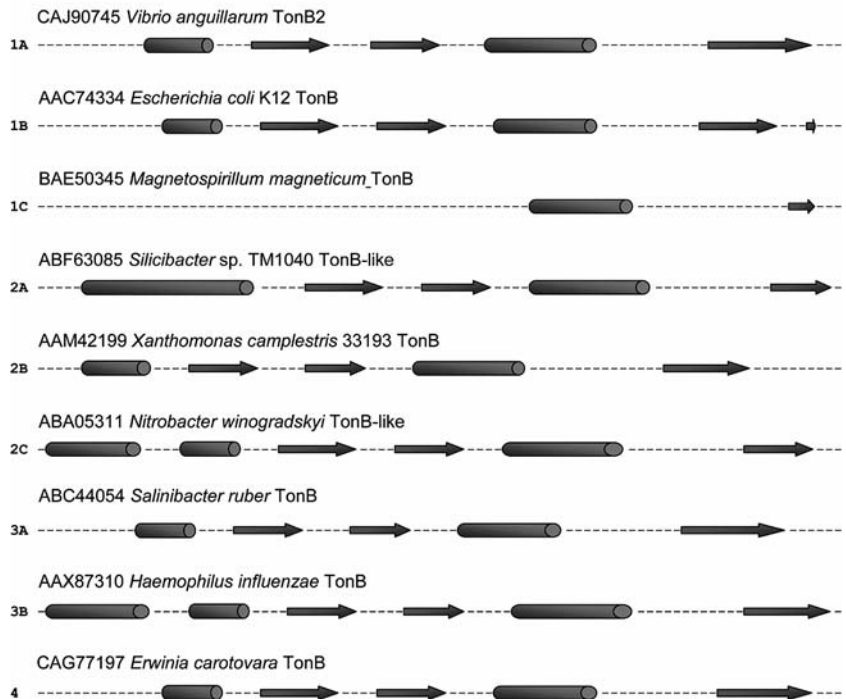
**Table 1** A partial list of annotated TonB's organized by C-terminal domain clustering. Additional information regarding the N-terminal domain and linker region connecting the N- and C-terminal domains is listed. A complete list of annotated TonB's will be available online at <http://growningen.bio.ucalgary.ca/vogel/tonb>

CTD cluster	Bacterial name	Taxonomic classification	Protein name	Genbank PID	Protein length	NTD length	NTD cluster	Length of linker domain	% Proline content (Linker domain)
1A	<i>Shewanella oneidensis</i> MR-1	Gammaproteobacteria	TonB2	AAN57525	203	32	1	81	18.52
1A	<i>Colwellia psychrerythraea</i> 34H	Gammaproteobacteria	TonB	AAZ28611	367	50	1	227	1.76
1A	<i>Vibrio anguillarum</i>	Gammaproteobacteria	TonB2	CAJ90745	206	32	1	84	10.71
1B	<i>Escherichia coli</i> K12	Gammaproteobacteria	TonB	AAC74334	239	33	2	116	25.86
1B	<i>Salmonella enterica</i> subsp. <i>Enterica</i> serovar <i>typhi</i> str. CT18	Gammaproteobacteria	TonB	AH0651	242	33	2	119	28.57
1B	<i>Yersinia pestis</i> CO92	Gammaproteobacteria	TonB	CAC91000	252	36	2	126	23.02
1C	<i>Idiomarina loihiensis</i> L2TR	Gammaproteobacteria	biopolymer system	AAV82691	280	46	1	144	1.39
1C	<i>Pseudomonas syringae</i> pv. <i>syringae</i> B728a	Gammaproteobacteria	TonB	AAV35379	221	30	1	101	7.92
1C	<i>Magnetospirillum magneticum</i> AMB-1	Alphaproteobacteria	TonB	BAE50345	307	30	1	187	28.34
2A	<i>Synechococcus</i> sp. JA-2-3B'a(2-13)	Cyanobacteria	TonB family	ABD02203	464	103	2	271	15.13
2A	<i>Silicibacter</i> sp. TM1040	Alphaproteobacteria	TonB-like	ABF63085	304	33	1	181	15.47
2A	<i>Rhizobium leguminosarum</i>	Alphaproteobacteria	TonB	CAC34389	454	82	1	282	12.06
2B	<i>Xylella fastidiosa</i> 9a5c	Gammaproteobacteria	TonB	AAF84759	221	46	2	163	9.82
2B	<i>Xanthomonas axonopodis</i> pv. <i>citri</i> str. 306	Gammaproteobacteria	TonB	AAM37949	291	40	2	161	10.56
2B	<i>Xanthomonas campestris</i> pv. <i>campestris</i> str. ATCC 33913	Gammaproteobacteria	TonB	AAM42199	291	40	2	161	10.56
2C	<i>Pseudomonas putida</i> KT2440	Gammaproteobacteria	TonB	AAN70873	243	35	1	118	33.05
2C	<i>Nitrobacter winogradskyi</i> Nb-255	Alphaproteobacteria	TonB-like	ABA05311	256	36	2	130	15.38
2C	<i>Rhodopseudomonas palustris</i> CGA009	Alphaproteobacteria	hypothetical	CAE27570	285	42	1	153	24.18
3A	<i>Chlorobium tepidum</i> TLS	Bacteroidetes/Chlorobi	TonB family	AAM72808	267	54	1	123	9.76
3A	<i>Bacteroides thetaiotaomicron</i> VPI-5482	Bacteroidetes/Chlorobi	TonB	AAO77166	227	41	1	96	12.5
3A	<i>Salinibacter ruber</i> DSM 13855	Bacteroidetes/Chlorobi	TonB, C-terminal	ABC44054	564	331	MecR1	38	13.16
3B	<i>Pasteurella multocida</i>	Gammaproteobacteria	TonB	AAQ05039	256	31	2	135	16.3
3B	<i>Mannheimia succiniciproducens</i> MBEL55E	Gammaproteobacteria	TonB	AAU37120	249	29	2	130	10
3B	<i>Haemophilus influenzae</i> 86-028NP	Gammaproteobacteria	TonB	AAV87310	264	31	2	143	13.99
4	<i>Pseudomonas aeruginosa</i>	Gammaproteobacteria	TonB2	AAF04082	270	66	1	114	42.11
4	<i>Methylococcus capsulatus</i> str. Bath	Gammaproteobacteria	TonB domain	AAU90375	248	38	2	120	25.83
4	<i>Erwinia carotovora</i> subsp. <i>atroseptica</i> SCRI1043	Gammaproteobacteria	TonB	CAG77197	287	41	1	156	23.08



**Fig. 5** Secondary structure predictions of TonB C-terminal domain clusters were performed using Jpred (Cuff and Barton 2000). Individual sequences from each TonB C-terminal domain cluster were chosen as representatives and were subjected to secondary structure prediction using the Jnet algorithm. The overall structural similarities and differences between C-terminal domain clusters are highlighted using this technique. The molecular basis for these secondary structure differences are examined further in Fig. 8.  $\alpha$ -helices and  $\beta$ -strands are depicted as barrels and arrows, respectively

**TonB-CTD Clusters (Representative Sequences):**



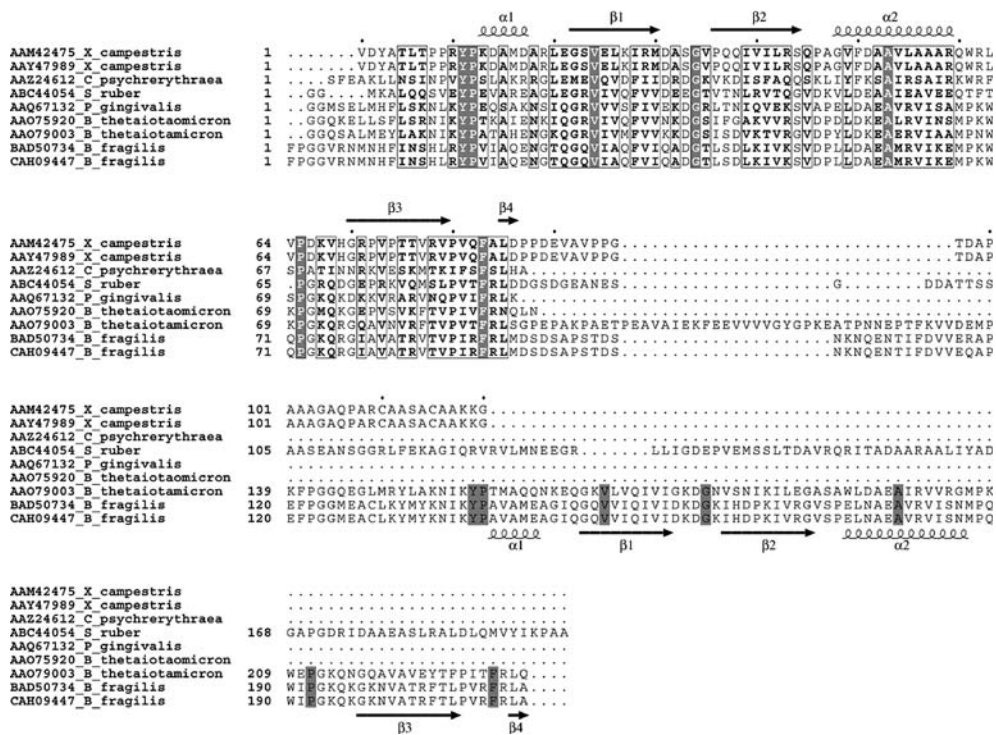
**Fig. 6** Multiple sequence alignment of representative sequences from TonB C-terminal domain clusters. Regions of high similarity (0.5) are boxed and slightly shaded. The highly conserved YP motif is boxed and highlighted in

grey. In addition, previously identified secondary structure elements from the *E. coli* TonB C-terminal domain are drawn and numbered above the alignment

region does not make any direct contact to the TonB box in any TonB-complex structures, it is in close proximity to the outer membrane receptor and therefore may play a role in receptor recognition.

The periplasmic linker domain

The linker domain of the *E. coli* TonB protein is comprised of 38 amino acids of unknown function (33–70), a 33 amino acid Pro-rich domain (70–



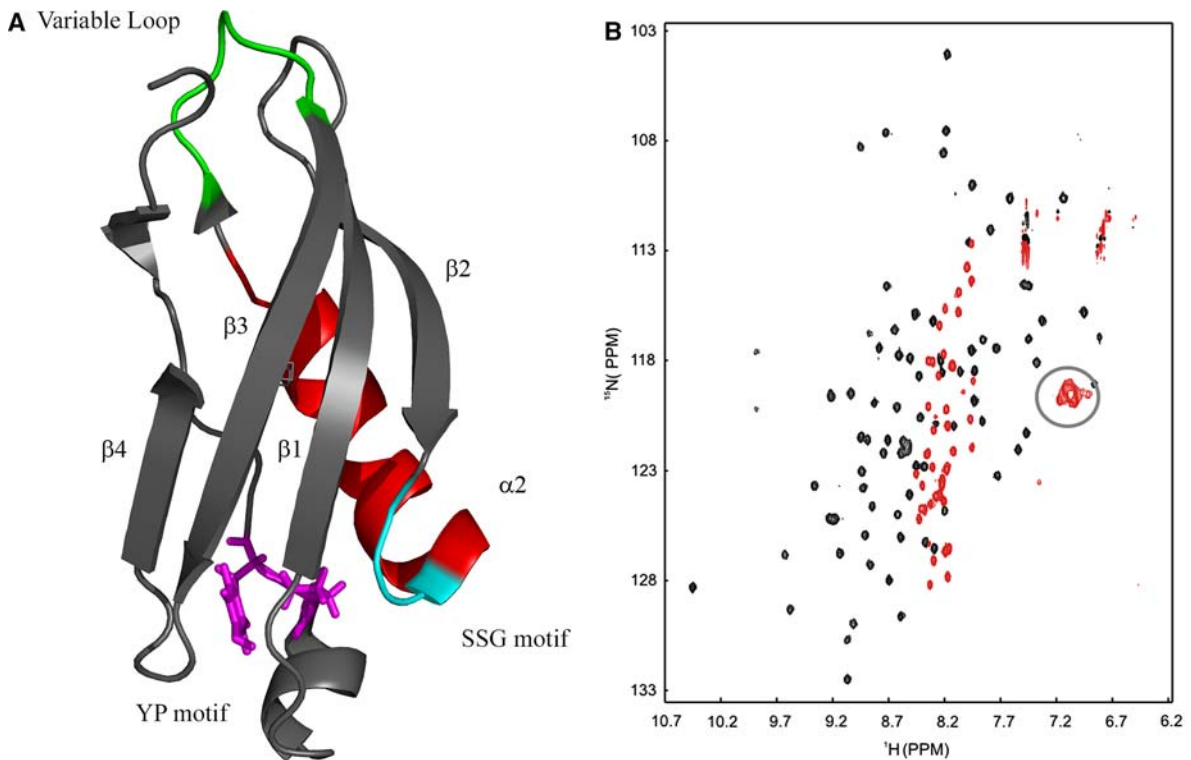
**Fig. 7** Multiple sequence alignment of C-terminal domains from annotated TonBs with long N-terminal domains (>291 residues). Residues that are boxed and slightly shaded are highly similar and those highlighted in grey are perfectly conserved. Secondary structural elements from the *E. coli* TonB CTD structure are shown

102) and an additional unfolded region from residues 103–150. The Pro-rich domain has previously been found not to be essential for the function of *E. coli* TonB (Larsen et al. 1993) and *Vibrio cholerae* TonB1 (Seliger et al. 2001). Nevertheless, the function of such shortened TonBs is slightly impaired in media that have an increased osmolarity, implying that this domain's main function is purely to provide a spacer allowing TonB to span the periplasmic space. It has been speculated that the Pro-rich region adopts a relatively rigid conformation, which may help in the transduction of mechanical energy from the cytoplasmic membrane to the outer membrane (Evans et al. 1986). Our analysis has found that the linker region between the N-terminal domain and the C-terminal domain is not necessarily Pro-rich at all, with the Pro contents of this domain ranging between 1.39% and 43.43%. The length of

above the alignment. In addition, a second potential TonB C-terminal domain can be found in TonB's from *B. thetaiotaomicron* (AAO79003) and *B. fragilis* (BAD50734 and CAH09447). The position of *E. coli* TonB CTD structural elements are shown below the alignment for the sequences with additional C-terminal domains

this region is also highly variable, ranging from 22 to 283 residues (Fig. 2B).

Previously we have discovered that in a truncated *E. coli* TonB (103–239), residues 103–150 have essentially no defined structure in solution (Peacock et al. 2005). To further check whether the Pro-rich domain has an influence on the structure of these residues, we titrated a  $^{15}\text{N}$ -labeled sample of 500  $\mu\text{M}$  TonB-CTD with peptides representing the EP repeat and KP repeats of the proline rich domain, as well as with both peptides together. In both cases, identical new amide chemical shifts were observed in the 6.5–7.5 ppm region. However, steady state NOE measurements of these samples showed that these new peaks were not indicative of any stable structure, as all of the new amides had negative steady state NOEs (Fig. 8B). This indicates that the Pro-rich domain in *E. coli* has no real influence on the flexible structure near the



**Fig. 8** (A) Structurally conserved and variable elements of TonB-CTD. The structure of TonB-CTD is illustrated with key aspects highlighted. The YP motif is colored in magenta, the amphipathic helix is colored in red and the variable length loop region between alpha helix 2 and beta sheet 3 is in green. The SSG motif loop is shown in cyan.

(B) Steady state NOE spectra of KP and EP peptide in complex with *E. coli* TonB-CTD. Black contours are positive and red contours are negative. Negative contours are an indicating of flexibility and weak binding. The new peaks which appear when the peptides are added are circled in gray

C-terminal domain, and that this region is likely flexible. This is unlike the TolA protein which is thought to have a helical bundle in its periplasmic domain joining the N and C-terminal domains (Witty et al. 2002).

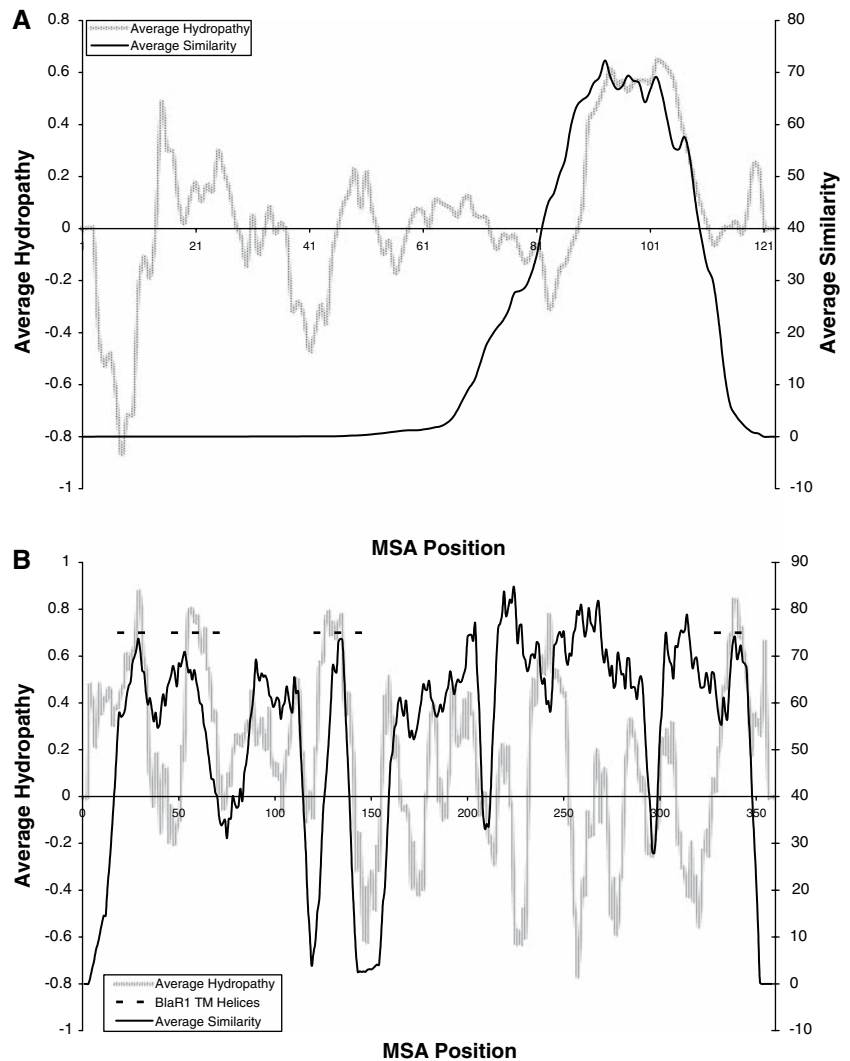
#### N-terminal domain

The TonB N-terminal domain functions in at least three different capacities in which it acts as: (1) a signal sequence to facilitate translocation of TonB to the periplasm; (2) an anchor to tether TonB to the cytoplasmic membrane; and (3) a partner to interact with ExbB /ExbD in an undefined manner to transduce energy to OMTs (Karlsson et al. 1993). TonB N-terminal domains range in length from 25–128 residues, but the majority are between 30 and 50 residues (Fig 2C). All of the TonB proteins have at least one predicted N-

terminal transmembrane  $\alpha$ -helix as determined by AveHas analysis of the N-terminal domain of TonB (Fig. 9A). Broadly we have categorized the N-terminal domains into three clusters where the majority of TonBs are found in clusters 1 and 2, and only six TonBs are found in cluster 3. The overall sequence similarity among the clusters was poor except in the region of the transmembrane  $\alpha$ -helix where a large number of hydrophobic residues are found. Previous studies have shown that a conserved SXXXH motif located in the transmembrane  $\alpha$ -helix is important for coupling TonB to the cytoplasmic membrane electrochemical gradient (Larsen and Postle 2001). This motif is conserved in N-terminal domain clusters 2 and 3 (e.g. *E. coli* K12 TonB), but not cluster 1 (e.g. *V. anguillarum* TonB2).

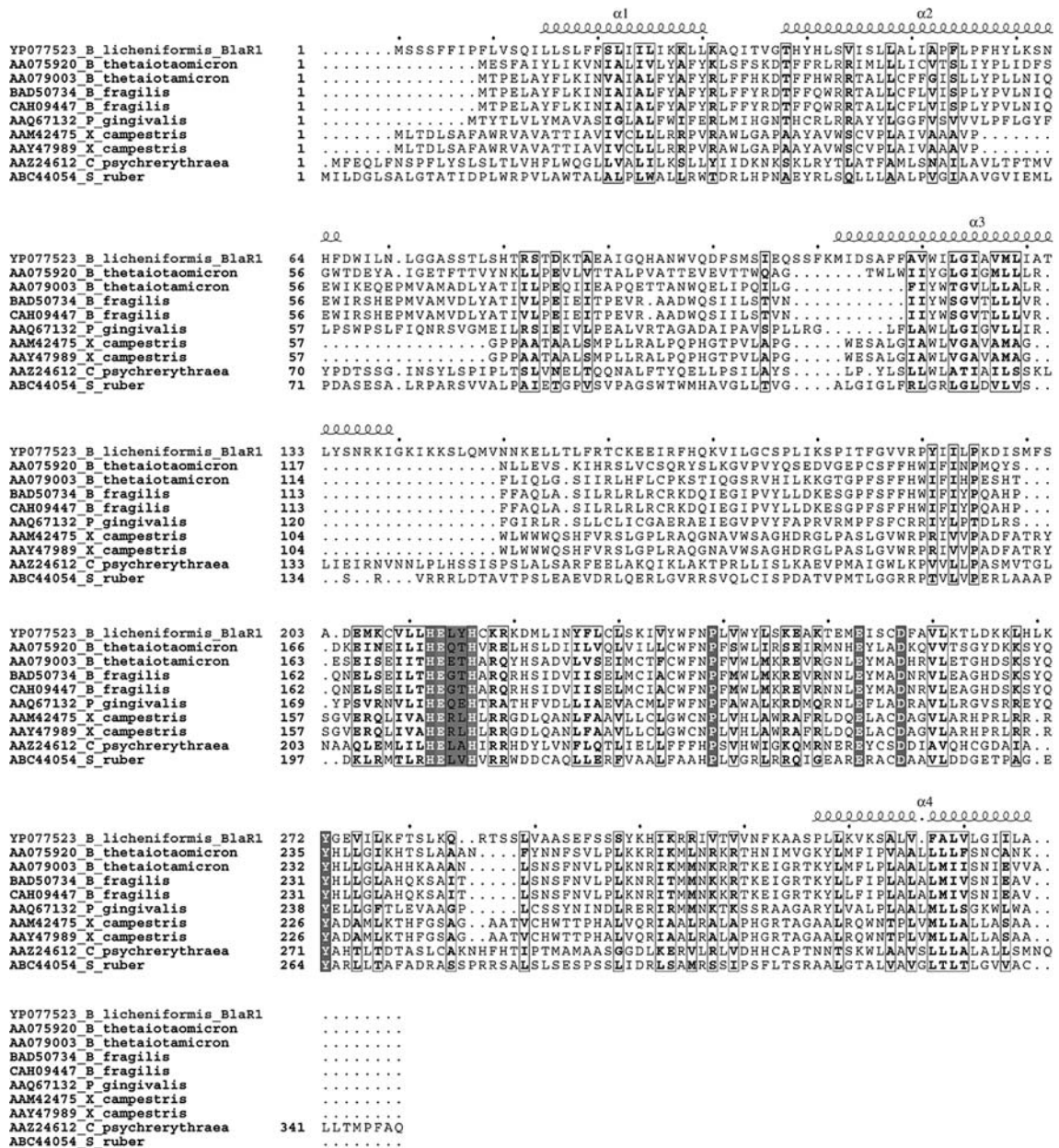
A number of TonB proteins have medium sized N-terminal extensions between 50 and 100

**Fig. 9** (A) TonB N-terminal domains were grouped into three major clusters based on Clustal X-derived multiple sequence alignments and phylogenetic analysis similar to the C-terminal domains. Average hydropathy and similarity plots of all three clusters were obtained using the AveHas program (Zhai and Saier 2001). The results from analysis of the N-terminal domain cluster 1 are displayed as a representative plot. (B) Average hydropathy and similarity values for Clustal X-derived multiple sequence alignments of TonB N-terminal domains (>291 residues) with annotated MecR1 regions. The location of the four predicted transmembrane helices from the BlaR1 transmembrane domain are shown (dashed lines). High average hydropathy and similarity values are observed for the four transmembrane helices and a high degree of similarity is observed in the loop 3 region where the putative zinc-binding motif, HEXXH, is located



amino acids, including *P. aeruginosa*, where the N-terminal extension has been shown to be essential for TonB1 stability (Zhao and Poole 2002), although the function of these extensions are unknown. A number of other previously undescribed TonB proteins however, have much longer N-terminal domains (291–348 residues). Examples of organisms containing these TonB proteins include *B. fragilis* and *X. camprestris*. These longer N-terminal domains are predicted by AveHas to have four transmembrane regions (Fig. 9B) as well as a highly conserved loop region that codes for an M56-Zn<sup>2+</sup> peptidase. The predicted secondary structure of these longer N-terminal domains is homologous to

the N-terminal domains of the highly similar BlaR1 and MecR1 proteins in *Staphylococcus aureus* and *S. sciuri*, that are involved in  $\beta$ -lactam antibiotic resistance (Hardt et al. 1997; Berger-Bachi and Rohrer 2002) (Fig. 9B). The profile alignment of the longer TonB N-terminal domains with the BlaR1 from *S. aureus* also shows perfect conservation of a putative zinc-binding motif (HEXXH) in the correct loop region and a number of conserved residues in an additional loop region (Fig. 10). The loss of interaction between the latter residues and the extracellular C-terminal sensor domain is important for activation of MecR1/BlaR1 (Hanique et al. 2004).



**Fig. 10** A profile alignment of the *B. licheniformis* BlaR1 transmembrane domain with the N-terminal domains of TonBs with annotated MecR1 regions. Areas with high degrees (0.5) of similarity are boxed and slightly shaded and perfectly conserved residues are highlighted in gray.

**Discussion**

Although the mechanism of TonB-energized transport is currently not well understood, two models have been proposed – the propeller model

The putative metalloprotease (zinc) binding motif HEX-XH located in loop 3 of the BlaR1 domain is conserved in all TonB MecR1 regions and is also highlighted in gray. The location of predicted BlaR1 transmembrane helices are shown above the sequence alignment

(Chang et al. 2001) and the shuttle model. However the discovery that the monomeric TonB-CTD binds to the TonB box region of the OMTs raises serious questions about the validity of the propeller model as well as the shuttle model. The shuttle model

involves complete dissociation of TonB from the inner membrane ExbB/ExbD complex (Larsen et al. 2003). The existence of the TonB proteins with multiple transmembrane helices, as seen here for *B. fragilis*, would tend to argue against such a mode of operation, as the energetic cost of removing this entire protein from the membrane would undoubtedly be much larger than a single transmembrane helix from a protein sleeve. No biological data exists however to show that these uncommon TonB proteins are active, and all of the organisms that we have identified with these proteins have multiple copies of TonB, so whether these long N-terminal extension TonBs are active in vivo is as of yet unknown and needs to be studied. Alternative models would include a pulling force exerted by the TonB/ExbB/ExbD proteins on the cork domain of the OMT, but there is currently little experimental evidence to support this idea.

Both MecR1/BlaR1 and TonB proteins are involved in signal transduction and possess similar structural organizations: a membrane anchored N-terminal domain, followed by a periplasmic linker and a highly conserved C-terminal domain. MecR1/BlaR1 is a receptor that detects and is activated by  $\beta$ -lactam antibiotics. Interaction with their sensor domain, initiates signal transduction through the transmembrane domain and activation of the M56-Zn<sup>2+</sup> peptidase that regulates expression of genes that encode for  $\beta$ -lactamase and penicillin-binding protein 2A through cleavage of the MecI/BlaI repressor (Hanique et al. 2004).

As mentioned previously, the vast majority of TonBs possess N-terminal domains in the range of 30–50 residues in length. TonB N-terminal domains are required for energy transduction and it has been suggested that the spatial relationship between the Ser and His residues of the SXXXH motif, located within the predicted transmembrane  $\alpha$ -helix, defines the minimum required transduction element (Larsen and Postle 2001). However, this applies only to TonBs with cluster 2 and 3 NTDs since a large number of cluster 1 TonB NTDs lack either the conserved Ser (e.g. *P. aeruginosa* TonB1) or His (e.g. *Vibrio vulnificus* TonB) residues individually or the motif in its entirety (e.g. *P. syringae phaseolicola* TonB3). His98 from TonB1 (cluster 1) of

*P. aeruginosa* aligns with the His20 from the SXXXH motif of the *E. coli* TonB (cluster 2) but has been shown by site-directed mutagenesis to be non-essential to TonB1 function (Zhao and Poole 2002).

There is also some evidence to suggest that TonBs have host-organism specific requirements that are related to their ability to interact with a variety of OMTs and/or ExbB/ExbD. The specific elements that define these requirements have not yet been identified and could be reflected in the N-terminal and C-terminal domain clustering we observed in this study. Some evidence that supports these claims can be found in TonB-swapping experiments. For example, TonB1 from *P. aeruginosa* is functional in *E. coli* but the TonB from *E. coli* is non-functional in *P. aeruginosa* and only partial functionality is restored when a chimeric TonB containing the TonB1 N-terminal extension of *P. aeruginosa* and the C-terminal domain of *E. coli* is expressed (Zhao and Poole 2002). TonB proteins can serve many different OMTs within a single organism, but in organisms with multiple TonB genes, each protein often only functions with a subset of the OMTs. For example, *Caulobacter crescentus* has 65 OMTs and only one TonB (Nierman et al. 2001), but *Vibrio anguillarum*, which has far fewer OMTs, has two TonBs, only one of which will import the native siderophore anguibactin (Stork et al. 2004). Discovering how and why the TonB protein can be so promiscuous and yet sometimes selective about which OMTs they can energize will be a large step in our understanding of the TonB mechanism.

The TonB proteins containing two independent TonB C-terminal domains are also of interest. A number of publications have suggested that TonB dimerization may be involved in the mechanism of transport so this could be a way in which a single TonB could dimerize (Chang et al. 2001; Ghosh and Postle 2004; Khursigara et al. 2005; Sauter et al. 2003). No biological data exists on these TonB proteins either though, so the effect of having a tandem TonB C-terminal domain is also unknown. It is tempting to speculate that the two TonB-CTD domains could confer specificity to different subsets of receptors using only one protein. An argument against this is that in all cases the individual CTDs for

sequences with tandem CTDs cluster to the same family. It should be emphasized that we have no evidence that the cluster the TonB-CTD belongs to in any way confers receptor specificity.

Overall though, the C-terminal domain of TonB is well conserved, both in terms of sequence and in terms of predicted secondary structure elements. It appears as though the portion of TonB that is the least well conserved is the loop between  $\alpha$ -helix 2 and  $\beta$ -strand 3. This area of TonB is distal from the interaction site between the outer membrane receptors as reported in the recent crystal structure of the complexes of TonB with the BtuB and FhuA OMTs (Shultis et al. 2006; Pawelek et al. 2006).

The function of the periplasmic linker region of *E. coli* TonB-CTD is uncertain. A recent report indicated that the Pro-rich domain was essential in forming a 2:1 TonB:FhuA complex (Khursigara et al. 2005). This is supported by early NMR studies on a peptide derived from the Pro-rich region which found an interaction between this region and FhuA (Brewer et al. 1990). It is also supported by phage display results that indicate that TonB has more than one interaction site with outer membrane receptors (Carter et al. 2006a), possibly due to an interaction with the Pro-rich domain. However it should be noted that with essentially the same constructs, the crystal structure of this complex revealed a 1:1 TonB:FhuA ratio (Pawelek et al. 2006) with no density observed for the Pro-rich region. A more recent study has raised the possibility of interactions between the periplasmic linker domain of TonB and the periplasmic siderophore binding protein, FhuD (Carter et al. 2006b). Phage display results map the possible interaction surfaces to residues 28–48, 119–140 and 142–168 in the *E. coli* TonB. The authors suggest that TonB, FhuA and FhuD form a 1:1:1 stoichiometric ternary complex. The mapping of the TonB-FhuD interactions to the amino- (residues 28–48) and carboxy (residues 119–140)-termini of the linker domain of *E. coli* TonB is surprising as there is very little amino acid sequence conservation with other TonBs in these regions to support a general function for these regions. It will be interesting to see if TonBs from other organisms also have this property.

This bioinformatics study represents an important step in demystifying the TonB protein. The overall sequence similarity of TonB proteins is very low, however, by dividing the protein into structural domains we have shown that TonB N-terminal and C-terminal domains can be grouped into clusters with particular characteristics. In addition, we have identified sequence specific features that can potentially be used as a starting point for future TonB studies involving its interactions with OMTs and ExbB/ExbD. To this end an important second bioinformatics step would be to examine OMT sequences in hopes of identifying and classifying structural elements that participate in TonB-OMT interactions using the TonB-BtuB and TonB-FhuA relationship as a template. This study highlights the need for careful scrutiny of bioinformatics studies and the synergy between experimental and in silico findings.

**Acknowledgements** This work was supported by an operating grant from the Canadian Institutes for Health Research (CIHR) to H.J.V.. R.S.P. was supported by Studentship awards from the Alberta Heritage Foundation for Medical Research (AHFMR) and the National Science and Engineering Research Council (NSERC). HJV holds a Scientist award from AHFMR. The NMR equipment used was obtained through grants from the Canada Foundation for Innovation, the Alberta Science and Research Authority (ASRA) and AHFMR. Maintenance of the Bio-NMR centre is supported by CIHR and the University of Calgary.

## References

- Altschul SF, Madden TL, Schaffer AA et al. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25:3389–3402
- Beddek AJ, Sheehan BJ, Bosse JT et al. (2004) Two TonB systems in *Actinobacillus pleuropneumoniae*: their roles in iron acquisition and virulence. *Infect Immun* 72:701–708
- Berger-Bachi B, Rohrer S (2002) Factors influencing methicillin resistance in staphylococci. *Arch Microbiol* 178:165–171
- Birck C, Cha JY, Cross J et al. (2004) X-ray crystal structure of the acylated beta-lactam sensor domain of BlaR1 from *Staphylococcus aureus* and the mechanism of receptor activation for signal transduction. *J Am Chem Soc* 126:13945–13947
- Braun V, Braun M (2002) Active transport of iron and siderophore antibiotics. *Curr Opin Microbiol* 5:194–201

- Brewer S, Tolley M, Trayer IP et al. (1990) Structure and function of X-Pro dipeptide repeats in the TonB proteins of *Salmonella typhimurium* and *Escherichia coli*. *J Mol Biol* 216:883–895
- Carter DM, Gagnon JN, Damlaj M et al. (2006a) Phage display reveals multiple contact sites between FhuA, an outer membrane receptor of *Escherichia coli*, and TonB. *J Mol Biol* 357:236–251
- Carter DM, Miousse IR, Gagnon JN et al. (2006b) Interactions between TonB from *Escherichia coli* and the periplasmic protein FhuD. *J Biol Chem*
- Chang C, Mooser A, Pluckthun A, Wlodawer A (2001) Crystal structure of the dimeric C-terminal domain of TonB reveals a novel fold. *J Biol Chem* 276:27535–27540
- Chimento DP, Kadner RJ, Wiener MC (2005) Comparative structural analysis of TonB-dependent outer membrane transporters: implications for the transport cycle. *Proteins* 59:240–251
- Cuff JA, Barton GJ (2000) Application of multiple sequence alignment profiles to improve protein secondary structure prediction. *Proteins* 40:502–511
- Delaglio F, Grzesiek S, Vuister GW et al. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* 6:277–293
- Eisenhauer HA, Shames S, Pawelek PD, Coulton JW (2005) Siderophore transport through *Escherichia coli* outer membrane receptor FhuA with disulfide-tethered cork and barrel domains. *J Biol Chem* 280:30574–30580
- Evans JS, Levine BA, Trayer IP, Dorman CJ, Higgins CF (1986) Sequence-imposed structural constraints in the TonB protein of *E. coli*. *FEBS Lett* 208:211–216
- Ghosh J, Postle K (2004) Evidence for dynamic clustering of carboxy-terminal aromatic amino acids in TonB-dependent energy transduction. *Mol Microbiol* 51:203–213
- Hanique S, Colombo ML, Goormaghtigh E et al. (2004) Evidence of an intramolecular interaction between the two domains of the BlaR1 penicillin receptor during the signal transduction. *J Biol Chem* 279:14264–14272
- Hardt K, Joris B, Lepage S et al. (1997) The penicillin sensory transducer, BlaR, involved in the inducibility of beta-lactamase synthesis in *Bacillus licheniformis* is embedded in the plasma membrane via a four-alpha-helix bundle. *Mol Microbiol* 23:935–944
- Higgs PI, Larsen RA, Postle K (2002) Quantification of known components of the *Escherichia coli* TonB energy transduction system: TonB, ExbB, ExbD and FepA. *Mol Microbiol* 44:271–281
- Johnson BA, Blevins RA (1994) NMRView—a computer program for the visualization and analysis of NMR data. *J Biomol NMR* 4:603–614
- Jurado RL (1997) Iron, infections, and anemia of inflammation. *Clin Infect Dis* 25:888–895
- Kampfenkel K, Braun V (1992) Membrane topology of the *Escherichia coli* ExbD protein. *J Bacteriol* 174:5485–5487
- Kampfenkel K, Braun V (1993) Topology of the ExbB protein in the cytoplasmic membrane of *Escherichia coli*. *J Biol Chem* 268:6050–6057
- Karlsson M, Hannavy K, Higgins CF (1993) A sequence-specific function for the N-terminal signal-like sequence of the TonB protein. *Mol Microbiol* 8:379–388
- Kay LE, Torchia DA, Bax A (1989) Backbone dynamics of proteins as studied by <sup>15</sup>N inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease. *Biochemistry* 28:8972–8979
- Khursigara CM, De Crescenzo G, Pawelek PD, Coulton JW (2004) Enhanced binding of TonB to a ligand-loaded outer membrane receptor: role of the oligomeric state of TonB in formation of a functional FhuA-TonB complex. *J Biol Chem* 279:7405–7412
- Khursigara CM, De Crescenzo G, Pawelek PD, Coulton JW (2005) Deletion of the proline-rich region of TonB disrupts formation of a 2:1 complex with FhuA, an outer membrane receptor of *Escherichia coli*. *Protein Sci* 14:1266–1273
- Klebba PE (2003) Three paradoxes of ferric enterobactin uptake. *Front Biosci* 8:s1422–s1436
- Koedding J, Howard P, Kaufmann L et al. (2004) Dimerization of TonB is not essential for its binding to the outer membrane siderophore receptor FhuA of *Escherichia coli*. *J Biol Chem* 279:9978–9986
- Koedding J, Killig F, Polzer P et al. (2005) Crystal structure of a 92-residue C-terminal fragment of TonB from *Escherichia coli* reveals significant conformational changes compared to structures of smaller TonB fragments. *J Biol Chem* 280:3022–3028
- Larsen RA, Letain TE, Postle K (2003) In vivo evidence of TonB shuttling between the cytoplasmic and outer membrane in *Escherichia coli*. *Mol Microbiol* 49:211–218
- Larsen RA, Postle K (2001) Conserved residues Ser(16) and His(20) and their relative positioning are essential for TonB activity, cross-linking of TonB with ExbB, and the ability of TonB to respond to proton motive force. *J Biol Chem* 276:8111–8117
- Larsen RA, Wood GE, Postle K (1993) The conserved proline-rich motif is not essential for energy transduction by *Escherichia coli* TonB protein. *Mol Microbiol* 10:943–953
- Letain TE, Postle K (1997) TonB protein appears to transduce energy by shuttling between the cytoplasmic membrane and the outer membrane in *Escherichia coli*. *Mol Microbiol* 24:271–283
- Lubkowski J, Hennecke F, Pluckthun A, Wlodawer A (1999) Filamentous phage infection: crystal structure of g3p in complex with its coreceptor, the C-terminal domain of TolA. *Structure Fold Des* 7:711–722
- Nierman WC, Feldblyum TV, Laub MT et al. (2001) Complete genome sequence of *Caulobacter crescentus*. *Proc Natl Acad Sci USA* 98:4136–4141
- Pawelek PD, Croteau N, Ng-Thow-Hing C et al. (2006) Structure of TonB in complex with FhuA, *E. coli* outer membrane receptor. *Science* 312:1399–1402
- Peacock RS, Weljie AM, Howard SP, Price FD, Vogel HJ (2005) The solution structure of the C-terminal domain of TonB and interaction studies with TonB box peptides. *J Mol Biol* 345:1185–1197
- Postle K, Kadner RJ (2003) Touch and go: tying TonB to transport. *Mol Microbiol* 49:869–882



- Ratledge C, Dover LG (2000) Iron metabolism in pathogenic bacteria. *Annu Rev Microbiol* 54:881–941
- Roof SK, Allard JD, Bertrand KP, Postle K (1991) Analysis of *Escherichia coli* TonB membrane topology by use of PhoA fusions. *J Bacteriol* 173:5554–5557
- Sauter A, Howard SP, Braun V (2003) In vivo evidence for TonB dimerization. *J Bacteriol* 185:5747–5754
- Seliger SS, Mey AR, Valle AM, Payne SM (2001) The two TonB systems of *Vibrio cholerae*: redundant and specific functions. *Mol Microbiol* 39:801–812
- Shultis DD, Purdy MD, Banchs CN, Wiener MC (2006) Outer membrane active transport: structure of the BtuB:TonB complex. *Science* 312:1396–1399
- Stork M, Di Lorenzo M, Mourino S et al. (2004) Two tonB systems function in iron transport in *Vibrio anguillarum*, but only one is essential for virulence. *Infect Immun* 72:7326–7329
- Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL\_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* 25:4876–4882
- Torres AG, Redford P, Welch RA, Payne SM (2001) TonB-dependent systems of uropathogenic *Escherichia coli*: aerobactin and heme transport and TonB are required for virulence in the mouse. *Infect Immun* 69:6179–6185
- Weinberg ED (1984) Iron withholding: a defense against infection and neoplasia. *Physiol Rev* 64:65–102
- Wiener MC (2005) TonB-dependent outer membrane transport: going for Baroque? *Curr Opin Struct Biol* 15:394–400
- Witty M, Sanz C, Shah A et al. (2002) Structure of the periplasmic domain of *Pseudomonas aeruginosa* TolA: evidence for an evolutionary relationship with the TonB transporter protein. *EMBO J* 21:4207–4218
- Zhai Y, Saier MH Jr. (2001) A web-based program for the prediction of average hydropathy, average amphipathicity and average similarity of multiply aligned homologous proteins. *J Mol Microbiol Biotechnol* 3:285–286
- Zhao Q, Poole K (2002) Mutational analysis of the TonB1 energy coupler of *Pseudomonas aeruginosa*. *J Bacteriol* 184:1503–1513