Sequence analysis

AMIN domains have a predicted role in localization of diverse periplasmic protein complexes

Robson Francisco de Souza1,2,†, Vivek Anantharaman3,†, Sandro José de Souza2, L. Aravind3 and Frederico J. Gueiros-Filho1,∗

1Departamento de Bioquímica, Universidade de São Paulo, São Paulo, Brazil, 2Ludwig Institute for Cancer Research, São Paulo Branch, São Paulo, Brazil and 3National Center for Biotechnology Information, NIH, Bethesda, MD 20894, USA

Received on March 29, 2008; revised on August 8, 2008; accepted on August 19, 2008

Advance Access publication August 21, 2008

Associate Editor: Dmitrij Frishman

ABSTRACT

We describe AMIN (Amidase N-terminal domain), a novel protein domain found specifically in bacterial periplasmic proteins. AMIN domains are widely distributed among peptidoglycan hydrolases and transporter protein families. Based on experimental data, contextual information and phyletic profiles, we suggest that AMIN domains mediate the targeting of periplasmic or extracellular proteins to specific regions of the bacterial envelope.

Contact: fgueiros@iq.usp.br

Supplementary information: Supplementary data are available at Bioinformatics online.

1 INTRODUCTION

Bacterial cells are more architecturally complex than previously thought. Several fundamental processes, such as cell division, chromosome replication, chemotaxis, motility and secretion depend on protein assemblies that are restricted to specific subcellular locations (Gitai, 2005). Perhaps the most striking example of this is the cell polarization of rod-shaped bacteria. In addition to being the cellular site of structures such as polar flagella and pili, several proteins have been recently shown to be targeted to the cell poles, where they perform diverse functions (Janakiraman and Goldberg, 2004). The targeting of specific proteins to the cell poles promotes functional compartmentalization in bacteria, which is akin to that achieved by organelles in eukaryotic cells. However, the mechanisms of protein localization in the absence of eukaryotic-style targeting systems and internal membranes are still poorly understood.

Here, we describe a novel protein domain prototyped by the N-terminal domain of AmiC, an N-acetylmuramoyl-L-alanine amidase of Escherichia coli, which localizes to the septal ring during division and plays a key role in the separation of daughter cells (Bernhardt and de Boer, 2003). The septal localization of AmiC has been shown to be dependent on an N-terminal globular domain that was previously only observed in the homologous amidases of Gram-negative bacteria (Bernhardt and de Boer, 2003). Hence, we investigated this domain (hereafter termed AMIN for Amidase N-terminal domain) further to determine its possible function. Our analysis showed that the AMIN domain is present in several protein families besides amidases and suggests that AMIN may represent a general targeting determinant involved in the localization of periplasmic protein complexes.

2 RESULTS AND DISCUSSION

2.1 Sequence analysis and secondary structure prediction

Using as query the AMIN domain (residues 31–180) of E.coli’s AmiC protein (accession number NP_311701), we implemented two search strategies based on PSI-BLAST (Altschul et al., 1997) and detected proteins from several bacterial lineages containing AMIN domains. These homologous domains were systematically collected and used as seeds for exhaustive transitive PSI-BLAST searches which revealed new homologs. Using the KALIGN program (Lassmann and Sonnhammer, 2005, 2006), we prepared a multiple sequence alignment of all homologous domains found by PSI-BLAST and built an Hidden Markov Model (HMM) model of AMIN, which was used in searches with the hmm_search program of the HMMER2 package (Finn et al., 2006) (refer to Material and methods section in the Supplementary Material for the statistical significance thresholds which were used). The combined result of our searches was the identification of numerous proteins with one or more copies of the AMIN domain. In addition to AmiC-like amidases from several proteobacteria, cyanobacteria and a few firmicutes, we identified AMIN in several other proteins, which lacked amidase domains but had other domains related to the physiology of the bacterial envelope.

A multiple alignment of AMIN domain homologs was generated using the output of the KALIGN program, manually adapted to take into account HSPs and alignments from PSI-BLAST and hmm_search, respectively (Fig. 1). The alignment was used as input for the secondary structure prediction program JPRED, which combines the information from the frequencies of residues in...
R.F.de Souza et al.

Fig. 1. Comprehensive multiple sequence alignment of representatives of all families of AMIN domains. Sequences are identified by their accession numbers in Genbank. The shading of residues highlights its degree of conservation. Boxes I to IV display the location of the conserved motifs Rxxx-

the multiple alignment, a HMM derived from it and the PSI-BLAST Position-Specific Scoring Matrix (PSSM) to derive a final consensus predicted structure. The results showed that the AMIN domain is likely to adopt an all \( \beta \)-strand fold with eight \( \beta \)-strands which are universally conserved in all members of the superfamily. Fold prediction searches for the AmiC AMIN domain using the Phyre (Bennett-Lovsey et al., 2008) and HHpred (Soding et al., 2005) servers returned hits to E-set and PapD-like domain superfamilies (Supplementary Material). Despite the weak similarity (13% identity), this suggests that the AMIN domain may adopt an immunoglobulin-like \( \beta \)-sandwich fold, with its eight strands organized in two parallel \( \beta \)-sheets (see Supplementary Fig. S1, panel A). Another possible fold for the domain can be proposed from examination of the consensus conservation pattern. This revealed that starting from the first strand, every two successive strands formed a comparable pair, with the entire domain being composed of a tetrad of such strand pairs. The second strand in each pair tended to conserve a signature of the form Rxxx- (where ‘-’ is usually a negatively charged residue, i.e. E/D). This pattern is particularly strong in second and third strand pairs of the domain (Fig. 1 and Supplementary Material). This observation raised the possibility that the domain arose from tandem tetraplication of an ancestral strand–strand dyad, each bearing the Rxxx- signature. Given the correlated conservation of the oppositely charged residues in the Rxxx-signature, it is likely that the second strands of the dyads interact with each other via the formation of salt bridges. This would predict an arrangement of strands wherein they form a \( \beta \)-helix structure with the even strands antiparallel to each other and held by the salt bridges (Supplementary Fig. S1, panel B). If this prediction is correct, the AMIN domain forms an unusual and novel structural scaffold.

2.2 Domain architecture

To obtain further insights into the possible function of the AMIN domain, we first clustered the proteins containing this domain into distinct families using the BLASTCLUST program and serial threshold cutoffs. We then systematically explored the domain architectures of the AMIN-containing proteins by taking representatives of each family and running PSI-BLAST searches, as well as by querying the library of identified AMIN domain proteins with PSSMs and HMMs for previously described domains (Finn et al., 2006). We found that the AMIN domain is found in a striking array of architectures (Fig. 2), and is only observed in secreted or cell-surface proteins from bacteria. The major types of organization are: (i) Fusion to the N-terminus of the amidase domain in AmiC-like amidases. Some of these proteins might contain more than one copy of the AMIN domain in the form of tandem repeats at the N-terminus of the protein. In firmicutes alone, the amidases contain a further Cu-oxidase N-terminal domain upstream of the amidase domain; (ii) In cyanobacteria, there is a fusion of the AMIN to the C-terminal \( \beta \)-lactamase domain, which is another type of murein hydrolase, which is unrelated to the amidase superfamily; (iii) Mainly in cyanobacteria, there is a fusion of the AMIN domain to a C-terminal \( \alpha \)-N-acetylglucosaminidase domain which hydrolyzes sugar linkages in peptidoglycan. In all the above-described domain architectures, there is a striking syntax of the form: AMIN + enzymatic domain X (where X is any of a range of enzymatic domains that operate on peptidoglycan components); (iv) The AMIN domain is found at the N-terminus of PilQ, which has additional C-terminal domains forming the secretin module such as the STN, secretin-N and secretin domains; (v) In a thematically similar architecture, one or more AMIN domains are combined
AMIN domain localizes periplasmic protein complexes

Fig. 2. Domain architecture (A) and conserved operon structure (B) for representative proteins containing copies of the AMIN domain. Proteins are identified by the organism code and protein accession numbers in Genbank. Organisms are: Anabaena variabilis (Ana), Paenibacillus polymyxa (Ppol), Sinorhizobium meliloti (Smel), Erwinia carotovora (Ecar), Campylobacter jejuni (Cjej), P. aeruginosa (Paer), Magnetococcus sp. MC-1 (Msp), Geobacter sp. FRC-32 (Gsp), δ-proteobacterium MLMS-1 (MldDRAFT) and M. xanthus (Mxan).

We surveyed the genes encoding AMIN proteins across all bacteria to detect conserved gene-neighborhoods (predicted operons). These are suggestive of functional interaction between the gene products: (i) in many proteobacteria, it occurs in an apparent operon with YjeE, a predicted ATPase of the P-loop fold, possibly involved in cell-wall biogenesis; (ii) In both proteobacteria and cyanobacteria, it occurs in an operon with Murl, a glutamate racemase involved in murein synthesis (Doublet et al., 1994). The same operon in cyanobacteria also encodes a polyprenyl synthetase (SDS), which might be required for synthesis of an outer membrane component; and (iii) In the case of PilQ, it is invariably found in the context of a large operon encoding the other pilin biogenesis genes.

2.3 Possible roles of AMIN domain

A combination of contextual information from phyletic patterns, domain architectures and conserved gene-neighborhoods, taken together with the available experimental data can be used to make functional predictions of uncharacterized proteins. We speculate that the role of the AMIN domain is to help localize and orient a range of transport and murein remodeling activities at key points in the periplasmic space. This speculation is based on the following observations: (i) The AMIN domain is almost always found at the N-terminus of a protein irrespective of the architectural context in which it is found. This polarity implies that it might function with a C-terminal STN domain and TPR repeats; (vi) The AMIN domain is combined with another cell-surface transport module—the TonB receptor module, which contains the TonB-dependent receptor domain along with the PLUG domain; (vii) Likewise, in at least one instance the AMIN domain is fused to two α-helical domains of the OEP (outer membrane efflux protein) family that form trimeric channels in the outer membrane of Gram-negative bacteria; (viii) The AMIN domain in proteins from several δ-proteobacteria is also fused to another α-helical charged region related to the coiled-coil segments in TolA protein; (ix) The AMIN domain is also found fused directly to multiple TPR repeats in several α-proteobacteria and some firmicutes; and (x) Finally, solo versions of the AMIN domain are found in several ε- and δ-proteobacteria, cyanobacteria and some actinobacteria.

An examination of the phyletic patterns of the AMIN domain proteins showed that they are found solely in the bacterial superkingdom, with no detectable versions in eukaryotes or currently available archaeal sequences. Amongst bacteria, it is predominantly found in bacteria with Gram-negative envelope: the majority of proteobacterial lineages (including acidobacteria), cyanobacteria, Aquifex and some nitrospirae, which contain one or more proteins with AMIN domains. It is rarely found in firmicutes (e.g. Moorella, Paenibacillus, Alkaliphilus, Pelotomaculum, Car boxydothermus and Clostridium) and actinobacteria (e.g. Corynebacterium, Frankia and Streptomyces) with Gram-positive envelopes.
as some kind of anchor that specifically helps orient the fused C-terminal domains in a particular fashion in the periplasmic space; (ii) The AMIN portion of AmiC is necessary and sufficient to target a protein to the septal region of E.coli’s periplasm (Bernhardt and de Boer, 2003); (iii) The PiIQ protein from Myxococcus xanthus, which contains copies of the AMIN domain, has been shown to localize to the poles of the bacterial cell (Nudleman et al., 2006). A similar situation can be expected in Pseudomonas aeruginosa, whose PiIQ protein also contains AMIN domains: even though PiIQ itself has not been localized, several other proteins involved in the biogenesis of type IV pili have been shown to be polar in this species (Chiang et al., 2005); and (iv) Evidence from the Neisseria PiIQ protein suggests that its N-terminal region, where the AMIN domain is found, is involved in interacting with the periplasm (Frye et al., 2006).

The AMIN domain is preponderant in Gram-negative bacteria suggesting that it evolved in response to the need of these bacteria to organize complex structures in their periplasm. These include transport systems to establish communication with the extracellular milieu and enzyme assemblies devoted to the biogenesis and maintenance of the cell wall and outer membrane. The fusion of AMIN to a range of enzymatic domains, all with catalytic activities related to murein hydrolysis suggests that it participates in cell-wall remodeling. Strikingly, the other set of architectures combine the AMIN domain with domains that form unrelated, but functionally comparable, transport structures across the Gram-negative outer membranes. These include the secretin module in PiIQ, which is involved in transport of type IV pili components across the outer membrane, the TonB-dependent receptors, which are porins that import metal chelators across outer membranes and OEP domains, which are involved in efflux of xenobiotics (Collins and Derrick, 2007; Ferguson and Deisenhofer, 2004; Johnson and Church, 1999). Based on this architectural context, we predict that proteins where the AMIN domain is fused to TPRs are likely to be similar outer membrane transport proteins, with the TPRs forming a shell-like super-structure similar to the OEP or TonA-like α-helical domains.

AMIN-mediated targeting of a protein must involve the recognition of structurally specialized regions in the periplasmic gel. Targeting of AmiC to a septal location is dependent on the cell division protein FtsN (Bernhardt and de Boer, 2003). FtsN has itself been shown to be a peptidoglycan-binding protein via its conserved peptidoglycan-binding domain (Ursinus et al., 2004; Yang et al., 2004). This raises the question as to whether the AMIN domain also directly binds peptidoglycan. In this case, binding would occur to a specific local configuration of murein, dependent on its cross-linking density, thickness or geometry. In support of this model, it is known that the metabolism of polar murein differs from that of the lateral wall, suggesting that peptidoglycan composition will not be the same throughout the cell (de Pedro et al., 1997). Nevertheless, at this point we cannot rule out that AMIN recognizes other components of the periplasm such as proteins or lipids.

3 CONCLUSIONS

We have identified a novel domain likely involved in the targeting of periplasmic or extracellular proteins to specific regions of the bacterial envelope. Understanding the nuts and bolts behind the organization of bacterial cells is still in its infancy. We hope our inferences will stimulate experimentation on this exciting topic.

In addition to confirming our prediction that AMIN is responsible for targeting proteins other than amidases to specific regions of the envelope, other important questions include: what does AMIN recognize? What is the structure of the AMIN domain? Why would some nutrient transporters need to be targeted to specific cellular locations? Finally, we suspect that the restriction of the AMIN domain to bacteria, especially Gram-negative forms, including several pathogens, makes it an attractive target for antibiotic development.

ACKNOWLEDGEMENTS

R.F.de S. thanks Alex Bateman for discussion and for stimulating the submission of this work. L.A. thanks Alex Bateman for facilitating the collaboration by informing him of R.F.de S. et al.’s independent work.

Funding: FAPESP 02/10616-4 (to F.J.G.-F.); National Library of Medicine [Intramural funds. National Institutes of Health (to V.A. and L.A.).]. R.F.de S. was a recipient of a doctoral fellowship from CNPq (Brazil).

Conflict of Interest: none declared.

REFERENCES


