doi:10.1016/j.jmb.2006.02.062 J. Mol. Biol. (2006) 358, 1010–1022

JMB

Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

The C-terminal TPR Domain of Tom70 Defines a Family of Mitochondrial Protein Import Receptors Found only in Animals and Fungi

Nickie C. Chan[1,2], Vladimir A. Likić[2], Ross F. Waller[3] Terrence D. Mulhern[1,2] and Trevor Lithgow[1,2]*

[1]*Department of Biochemistry and Molecular Biology University of Melbourne Parkville 3010, Australia*

[2]*Bio21 Molecular Science and Biotechnology Institute University of Melbourne Parkville 3010, Australia*

[3]*Botany School, University of Melbourne, Parkville 3010 Australia*

**Corresponding author*

In fungi and animals the translocase in the outer mitochondrial membrane (TOM complex) consists of multiple components including the receptor subunit Tom70. Genome sequence analyses suggest no Tom70 receptor subunit exists in plants or protozoans, raising questions about its ancestry, function and the importance of its activity. Here we characterise the relationships within the Tom70 family of proteins. We find that in both fungi and animals, a conserved domain structure exists within the Tom70 family, with a transmembrane segment followed by 11 tetratricopeptide repeat motifs organised in three distinct domains. The C-terminal domain of Tom70 is highly conserved, and crucial for the import of hydrophobic substrate proteins, including those with and those without N-terminal presequences. Tom70 likely arose after fungi and animals diverged from other eukaryote lineages including plants, and subsequent gene duplication gave rise to a paralogue specific to the *Saccharomyces* group of yeasts. In animals and in fungi, Tom70 plays a fundamental role in the import of precursor proteins, by assisting relatively hydrophobic regions of substrate proteins into the translocation channel in the outer mitochondrial membrane. Proteins that function equivalently to Tom70 may have arisen independently in plants and protists.

Keywords: protein import; tetratricopeptide repeat; Tom70; Tom71; hidden Markov models

Introduction

Recent proteomic analyses suggest that mitochondria in any given cell type probably contain 800–1000 different proteins.[1] Ninety-nine per cent of these mitochondrial proteins are coded by nuclear genes, made in the cytosol and imported into the organelle. Their initial recognition and import through the outer mitochondrial membrane is facilitated by the multi-subunit translocase called the TOM (translocase in the outer mitochondrial membrane) complex, consisting of receptors Tom70 and Tom20, and the channel-forming Tom40 and its attendant subunits. Many mitochondrial proteins are synthesised with an N-terminal presequence that provides the targeting information needed for mitochondrial location. After passage through the TOM complex, N-terminal presequences direct the precursor to interact with the TIM23 complex in the inner membrane, and as a result the protein enters the mitochondrial matrix space where a processing peptidase removes the targeting sequence.[2–5] Many of the proteins destined for mitochondria do not have an N-terminal extension and instead rely on targeting information embedded internally, often at or overlapping their transmembrane domains.[2–4,6] Some of the proteins carrying these internal targeting sequences, notably members of the carrier protein family,[7] will be translocated through the outer membrane and passed on to the TIM22 complex that drives the insertion of substrate proteins into the mitochondrial inner membrane.[2,4,6]

Abbreviations used: TOM, translocase in the outer mitochondrial membrane; TIM, translocase in the inner mitochondrial membrane; TPR, tetratricopeptide repeat; HMM, hidden Markov model; CCCP, *m*-chlorophenylhydrazone.

E-mail address of the corresponding author: t.lithgow@unimelb.edu.au

How does the TOM complex recognize and bind both N-terminal and internal targeting sequences? Although two parallel pathways are often discussed, with Tom20 and Tom70 mediating each pathway independently, a model whereby a single, hetero-oligomeric receptor was responsible for recognition of all substrates was previously proposed.[8] According to this model, each molecule of precursor protein interacts through its targeting portion with Tom20, while other parts of the substrate interact with Tom70. Evidence that the Tom70 receptor binds at least some precursor proteins directly came with peptide scans of substrate proteins: recombinant Tom70 selectively binds peptides derived from discrete segments of substrates and a proteolytically stable "core" domain of ~25 kDa is also capable of binding substrates, albeit with reduced avidity.[9] In addition, a segment of Tom70 was recently identified with features similar to the "clamp" domain of the chaperone HOP, and a point mutation in this region of Tom70 prevents the binding of molecular chaperones like Hsp70.[10] Several molecular chaperones in the cytosol transfer a range of substrate proteins to Tom70, suggesting that this receptor subunit plays a fundamental role in protein import into mitochondria.[10,11]

Tom20 and Tom70 can interact with each other by virtue of tetratricopeptide repeats (TPRs), with a site-specific mutation in the TPR segment of Tom20 having no effect on the function of Tom20 but blocking instead the function of Tom70.[12] The TPR motif is degenerate in primary structure, composed of 34 amino acid signatures that reflect structural elements of tight helix-turn-helix packing. No residue within the motif is absolutely conserved and until recently prediction of TPR segments has yielded ambiguous results.[13,14] Structural analysis of Tom20 from animals and fungi showed that a single TPR segment is found in Tom20.[15–18] In the case of Tom70, previous analyses suggested seven TPR motifs in Tom70 from yeast and humans,[9,10,13,19] but at least ten TPR motifs in the Tom70 from rats.[20] With structures recently determined for a range of proteins composed of multiple TPR segments, it is now possible to confidently predict these elements from sequence alone and it is becoming clear that specific features conserved within the TPR segments can assist in defining distinct protein families and clans.

Here we evaluate structural aspects of the Tom70 family of proteins and suggest that they are characterised by 11 TPR motifs. These 11 motifs are organised with three TPRs in an N-terminal "clamp" domain, five in a "core" domain and three in a previously uncharacterised C-terminal domain. Analysis of motifs in the Tom70 family revealed contiguous motifs in the C-terminal domain with a highly conserved sequence signature centred on residues that would interact with substrate. The signature motif provides a diagnostic search tool to find members of the Tom70 family, and shows that the family arose early in the evolution of fungi and animals. A series of Tom70 mutants, designed on the predicted TPR segment structure, shows that Tom70 plays a role in facilitating the import of most, if not all, precursor proteins into mitochondria. This is true for substrate proteins with N-terminal targeting signals as well as for proteins with internal targeting signals. The C-terminal domain is crucial for binding these substrate proteins.

Results

Ancestry of the Tom70 family

In a comprehensive search for Tom70 orthologues, we started with the sequences for the functionally defined Tom70 from *Saccharomyces cerevisiae*, *Neurospora crassa* and rat, retrieving an initial set of 20 sequences related to Tom70 (Supplementary Data, Figure 1). The initial set of sequences was subjected to a pattern search algorithm[21] to elucidate Tom70 motifs. Hidden Markov models (HMMs) were built and proved an effective search tool, able to discriminate a family of 41 Tom70 proteins from the next most similar family of TPR-rich proteins (the Sti1/p60 family[22]) and from further proteins unrelated to Tom70 except for the presence of multiple TPR motifs. Pair-wise identity of the Tom70 sequences is only in the order of 20–30%, yet conserved features are apparent in each sequence: a protein size of between 477 and 639 residues, a single predicted N-terminal signal-anchor (transmembrane) sequence, followed by a region rich in glutamine, glutamate and lysine residues that might represent a disordered segment of polypeptide, followed by a series of characteristic TPR motifs.

A phylogeny of the Tom70 sequences from a range of fungal and animal species is shown in Figure 1. This topology is well supported with very high bootstrap values across the tree. A single gene encoding Tom70 is present in each animal species and in each of the fungi except for some yeasts that possess a Tom70 paralogue (Tom71). The topology suggests a gene duplication having occurred at the base of the *Saccharomyces* clade, including the closely related yeasts *Candida glabrata* and *Naumovia castellii*.[23] This is consistent with evidence of a genome duplication event that occurred after the divergence of the *Saccharomyces* species from the lineage that gave rise to *Kluveromyces lactis* and *Eremothecium gossypii*.[24,25] Further evidence of a Tom70 paralogue arising through such an event is clear from genome analysis of *S. cerevisiae*, where the *TOM70* and *TOM71* genes sit amidst common gene synteny on their respective chromosomes (XIV and VIII[25,26]). Previously, BLAST analysis with Tom70 found bacterial protein relatives, which led to the suggestion that the Tom70 receptor subunit might have an ancient origin in the evolution of mitochondria from bacterial ancestors.[27] Considering the number of diverse bacterial proteins that