Correspondence

The yeast Sar exchange factor Sec12, and its higher organism orthologs, fold as β-propellers

Pierre Chardin[a], Isabelle Callebaut[b,*]

First published online 16 July 2002

Sar belongs to a distinct family of small GTP-binding proteins that also comprises Arf and Arl proteins. It is involved in the membrane recruitment of the COPII coat that drives the formation of vesicles, which exit from the endoplasmic reticulum [1]. Sec12 has been well defined by genetic studies in yeast as an activator of Sar [2]. It is a transmembrane protein with a large cytosolic N-terminal domain possessing guanine nucleotide exchange factor (GEF) activity on Sar. Because Arf, Arl and Sar proteins share a common fold and several characteristic structural properties (a β2–β3 hairpin and N-terminal amphipathic helix), we wondered if Sec12 shares some structural similarities with GEFs for Arfs that contain a 'Sec7' domain with a structure mostly formed of α-helices [3]. No significant similarity could be detected with Sec7 domains by the usual search algorithms such as PSI-BLAST [4]. However, WD40 repeats were detected in the cytosolic domain of Sec12. WD40 repeats are short repeated sequences of about 40 amino acids, originally named after the presence of a conserved central Trp–Asp (W–D) motif, first identified in the Gβ protein. These repeated sequences are arranged in circular β-propeller structures, in which each blade consists of a four-stranded anti-parallel β-sheet (see inset in Fig. 1). On both sides of the β-sheets plate are displayed the loops connecting β-strands, which concentrate the greatest variability and form widely exposed surfaces amenable to interaction with different partners. Some WD40 repeats have very degenerated sequences in which the WD40 hallmarks are not maintained, but can still be classified as WD40 repeats as the hydrophobicity in positions participating in the maintenance of the fold is conserved (see accession number SM0320 in the Smart database [5]).

A search against domain databases such as Smart [5] and Pfam [6] led to the detection of only two WD40 repeats at the C-terminal end of the cytoplasmic domain of yeast Sec12, just before the transmembrane segment. However, most well known proteins folding as β-propellers have six or seven blades, strongly suggesting that the number of WD40 repeats in Sec12 was underestimated. This hypothesis was supported by the detection of four WD40 repeats in the mammalian orthologue, mSec12, also named PREB [2], but the high sequence divergence between yeast and mammalian Sec12 hampers their accurate alignment (it is not reported in a BLAST search) and the delineation of crucial functional amino acids.

A search on the full sequence of yeast Sec12 within the non-redundant database at NCBI using PSI-BLAST (default values) first revealed its closest homologues in other yeasts (Sec12 of *Kluyveromyces lactis*, *Pichia pastoris*, *Candida albicans*, St11 of *Schizosaccharomyces pombe* and Sed4 of *Saccharomyces cerevisiae*) and also detected significant similarities with the *Arabidopsis thaliana* Ste12 sequence. At further iterations, PSI-BLAST detected significant similarities with functionally unrelated WD40 repeat-containing proteins, which allowed us to align accurately five WD40 repeats of yeast Sec12, located at the C-terminus of its cytoplasmic domain. It is worth noting that, using the full sequence as query in PSI-BLAST, no significant alignment of the N-terminal sequence of Sec12 could be obtained with WD40 repeats, nor with the human protein PREB. Limiting the query to this Sec12 N-terminal sequence, PSI-BLAST searches led us to highlight the probable human, fly and worm orthologues of yeast Sec12 (PREB, CG9175 and K02B12.3, respectively) and to detect marginal similarities with WD40 repeats, which were then confirmed at the secondary structure level and refined using hydrophobic cluster analysis (HCA) [7]. In this way, two additional WD40 repeats were detected at the N-terminus of Sec12, leading to the prediction of a total of seven blades in Sec12 family members, and a refined alignment of all the available sequences of the Sec12 family was performed (Fig. 1). Details about this analysis can be found at http://www.lmcp.jussieu.fr/~callebau/Sec12.html.

Surprisingly, the fifth repeat could not be detected in the *Caenorhabditis elegans* sequence (K02B12.3). However, translation of a sequence predicted as an intron and examination of the HCA plots of the different frames led to rapidly highlight clusters typical of strands A, B and C, which can be well aligned with the human PREB and *Drosophila* CG9175 sequences (italics and underlined in Fig. 1). Another possible 5′ splice junction is present, strongly suggesting that the sequence initially predicted as an intron encodes the 'missing' fifth blade of the propeller.

In yeast Sec12, the predicted seven-bladed β-propeller should be rather compact, with only short insertions, and represents most of the cytosolic domain, ending very close to the beginning of the transmembrane helix. The functionally analogous proteins from higher eukaryotes have larger insertions, which, in combination with the high level of divergence, made the alignment difficult. Interestingly, the highest conservation is found in the first two blades, which appear as highly specific of the Sec12 family, suggesting that they might be critical for the function. There is a striking conservation of a four-glycine stretch, followed by another glycine and a strictly conserved asparagine, in a consensus GGGGxxxxGΦxN (where Φ is a hydrophobic amino acid, B–C loop, first blade, star in Fig. 1). It is tempting to speculate that the strictly conserved asparagine predicted to be exposed on one side of the propeller could be one of the major catalytic residues. This sequence is seven amino acids longer than those of classical WD40 repeat B–C loops (which very often correspond to a tight turn), suggesting that it could protrude from the top face of the propeller, like the β-wedge in the β-propeller structure of RCC1, a GEF for Ran proteins [8]. Only three other strictly conserved residues were detected along the alignment of the orthologous sequences. One of these, in the seventh repeat, corresponds to a histidine, which belongs to the structural tetrad (vertical arrow in Fig. 1) and thus should not be exposed to solvent. Interestingly, also in

PII: S0014-5793(02)03068-5

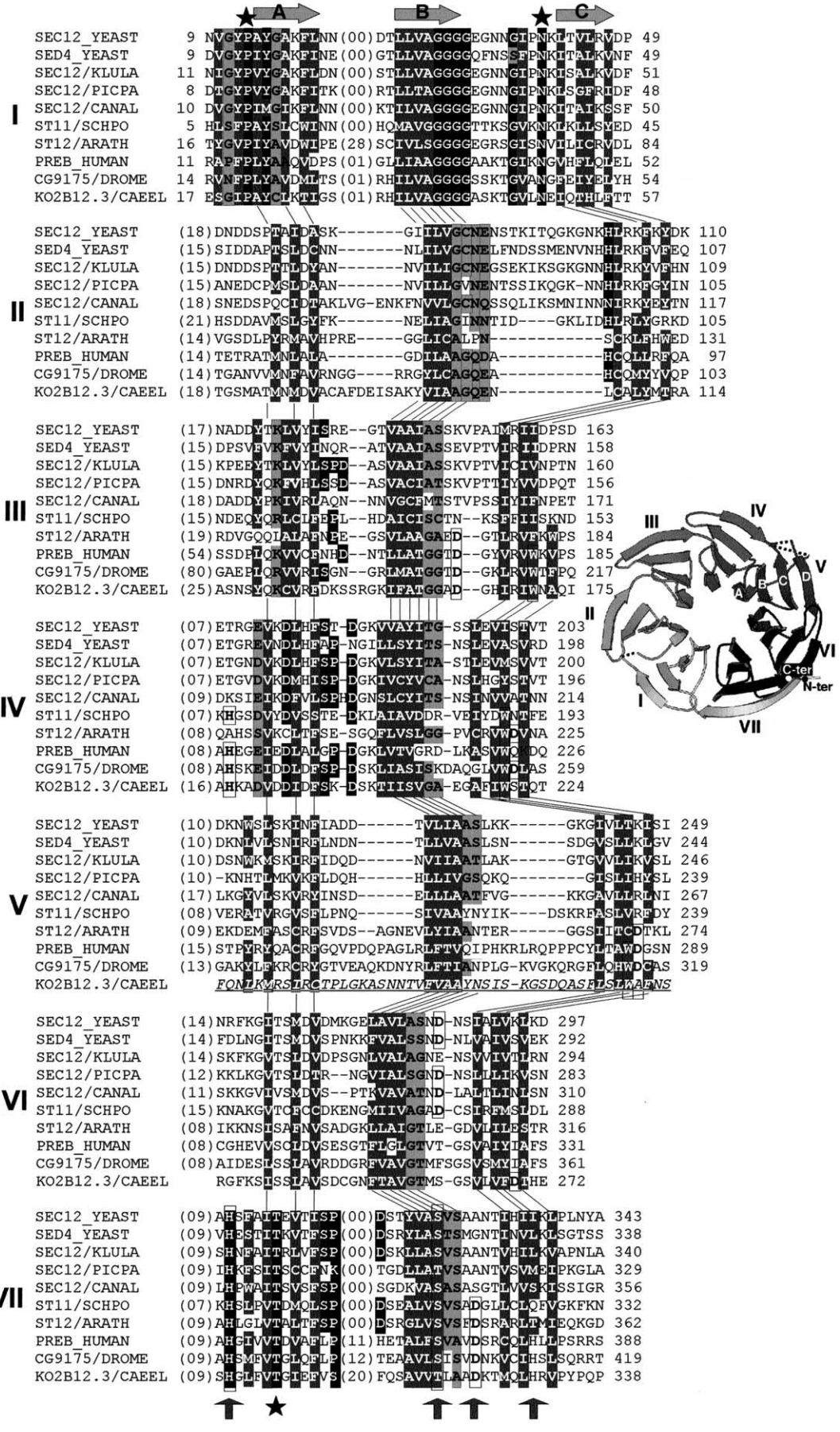
A
B
C
I
II
III
IV
V
VI
VII
SEC12_YEAST
SED4_YEAST
SEC12/KLULA
SEC12/PICPA
SEC12/CANAL
ST11/SCHPO
ST12/ARATH
PREB_HUMAN
CG9175/DROME
KO2B12.3/CAEEL
C-ter
N-ter

Fig. 1. Alignment of several orthologues in the Sec12 family. Sequence identifiers are reported in front of each repeated sequence (labelled I–VII). Positions of strands A–C are indicated and conserved positions, typical of the WD40 repeat signature, can be followed along the different repeats by black lines (mostly positions occupied by hydrophobic amino acids, in addition to two positions occupied by two small amino acids at the end of strands B). Strand D, which is highly variable between the repeats, is not aligned. Distances between strands C and A, including the non-aligned strands D, are indicated within brackets. Sequence identities are indicated white on a black background, similarities are grey-shaded, with hydrophobic amino acids (V, I, L, F, M, Y, W) or amino acids that can substitute them in some circumstances (A, S, T, C) indicated in white, other conserved non-hydrophobic residues are indicated black on a grey background. Amino acids typical of the WD40 repeat structural tetrad are indicated with vertical arrows and boxed. The three strictly conserved amino acids that are predicted to play a key role in the Sec12 function are indicated with stars. The sequence of the fifth repeat (V) of the *C. elegans* K02B12.3 protein (italics and underlined) was deduced from its translated DNA sequences (predicted as an intron), which were compared through HCA with the fifth repeats of other organisms. The position of this sequence in the nucleotidic sequence of the *C. elegans* cosmid K02B12 ranges from 17065 to 17211 (3′–5′ translation). The inset shows a ribbon representation of the three-dimensional structure of the TUP1 β-propeller (PDB identifier 1ERJ), illustrating the numbering of each blade (I–VII) and, within each blade, the strand labelling (A–D). The sequence repeat is shifted relatively to the structural repeat, so that the last strand of the last blade is furnished by the N-terminal sequence of the Tup1 β-propeller. Also in Sec12, the last strand of the last blade (β-strand D, seventh blade) is most likely provided by the N-terminal. Indeed, the first strand A of the first repeat is preceded by a cluster predicted as strand D, which should participate in the last blade including strands A, B and C of the last repeat. GenBank identification numbers: SEC12_YEAST (*S. cerevisiae*): 6324353, SED4_YEAST: 140524, SEC12/KLULA (*K. lactis*): 7578634, SEC12/PICPA (*P. pastoris*): 6746586, SEC12/CANAL (*C. albicans*): 3850135, ST11/SCHPO (*S. pombe*): 173493, ST12/ARATH (*A. thaliana*): 11358859, PREB_HUMAN (*H. sapiens*): 17368974, CG9175_DROME (*D. melanogaster*): 7297075, K02B12.3_CAEEL (*C. elegans*): 17508005. ←

the seventh repeat, a strictly conserved threonine is predicted exposed at the surface of the protein, at the beginning of strand A (D–A loop). This position lies in the vicinity of the repeat I B–C loop, which bears the conserved asparagine discussed above. Thus, the proximity of these two strictly conserved residues, together with that of the third strictly conserved residue (a proline in the first repeat D–A loop) reinforces the hypothesis of their critical role in the Sec12 exchange function. It is worth noting that these positions match the 'supersite' described for β-propellers, particularly involving the N-termini of the inner β-strands (strands A) [9]. However, in RCC1 [8], the Ran binding site is located on the opposite side, involving A–B and C–D loops.

Tup1, the structurally characterised WD40 repeat protein that shares the highest sequence similarities with Sec12, is a co-repressor of transcription in yeast [10] and PREB, the human Sec12 orthologue, had first been described as a prolactin regulatory element binding protein [11]. It is at present unknown how the putative function of PREB in the regulation of transcription might be related to that of a Sar GEF. Since propellers represent a rather common type of fold that has been found in proteins with very different functions, it is too early to speculate on the functional relationship between PREB and Tup1.

Owing to the strong structural similarities between Sar1 and Arfs or Arls, it is somewhat surprising that their respective GEFs are structurally unrelated. However, there are precedents for such structural differences. GEFs for Ras and for Rho are mostly composed of α-helices, but RCC1, the GEF for Ran proteins which share about 35% identity with Ras or Rho, folds as a seven-bladed propeller characterised by a repeated sequence different from the WD40 β-propeller motif [8]. The Sec12 cytosolic domain represents a second example of GEF that folds as a seven-bladed β-propeller.

Acknowledgements: We thank Bruno Antonny and Jean-Paul Mornon for support and critical reading of the manuscript, as well as referees for pertinent suggestions.

References

[1] Antonny, B. and Schekman, R. (2001) Curr. Opin. Cell Biol. 13, 438–443.
[2] Weissman, J., Plutner, H. and Balch, W. (2001) Traffic 2, 465–475.
[3] Cherfils, J. and Chardin, P. (1999) Trends Biochem. Sci. 24, 306–311.
[4] Altschul, S., Madden, T., Schaffer, A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D. (1997) Nucleic Acids Res. 25, 3389–3402.
[5] Letunic, I. et al. (2002) Nucleic Acids Res. 30, 242–244.
[6] Bateman, A. et al. (2002) Nucleic Acids Res. 30, 276–280.
[7] Callebaut, I., Labesse, G., Durand, P., Poupon, A., Canard, L., Chomilier, J., Henrissat, B. and Mornon, J. (1997) Cell. Mol. Life Sci. 53, 621–645.
[8] Renault, L., Kuhlmann, J., Henkel, A. and Wittinghofer, A. (2001) Cell 105, 245–255.
[9] Russell, R.B., Sasieni, P.D. and Sternberg, M.J.E. (1998) J. Mol. Biol. 282, 903–918.
[10] Sprague, E.R., Redd, M.J., Johnson, A.D. and Wolberger, C. (2000) EMBO J. 19, 3016–3027.
[11] Fliss, M., Hinkle, P. and Bancroft, C. (1999) Mol. Endocrinol. 13, 644–657.

*Corresponding author. Fax: (33) 1 44 27 37 85.
E-mail address: callebau@lmcp.jussieu.fr (I. Callebaut).

[a] *Institut de Pharmacologie moléculaire et cellulaire, CNRS UMR 6507, 660 route des Lucioles, Sophia Antipolis, 06560 Valbonne, France*
[b] *Systèmes Moléculaires et Biologie structurale, LMCP, CNRS UMR 7590, Universités Paris 6 et Paris 7, case 115, 4 place Jussieu, 75252 Paris Cedex 05, France*

PII: S0014-5793(02)03068-5