Protein Science (1999), *8*:1658–1667. Cambridge University Press. Printed in the USA.
Copyright © 1999 The Protein Society

The α-subunit of protein prenyltransferases is a member of the tetratricopeptide repeat family

HONG ZHANG[1] AND NICK V. GRISHIN[2]

[1]Center for Advanced Research in Biotechnology and University of Maryland Biotechnology Institute, 9600 Gudelsky Drive, Rockville, Maryland 20850
[2]National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, Maryland 20894

(RECEIVED February 4, 1999; ACCEPTED April 30, 1999)

Abstract

Lipidation catalyzed by protein prenyltransferases is essential for the biological function of a number of eukaryotic proteins, many of which are involved in signal transduction and vesicular traffic regulation. Sequence similarity searches reveal that the α-subunit of protein prenyltransferases (PTα) is a member of the tetratricopeptide repeat (TPR) superfamily. This finding makes the three-dimensional structure of the rat protein farnesyltransferase the first structural model of a TPR protein interacting with its protein partner. Structural comparison of the two TPR domains in protein farnesyltransferase and protein phosphatase 5 indicates that variation in TPR consensus residues may affect protein binding specificity through altering the overall shape of the TPR superhelix. A general approach to evolutionary analysis of proteins with repetitive sequence motifs has been developed and applied to the protein prenyltransferases and other TPR proteins. The results suggest that all members in PTα family originated from a common multirepeat ancestor, while the common ancestor of PTα and other members of TPR superfamily is likely to be a single repeat protein.

Keywords: helix packing; protein evolution; protein–protein interactions; protein prenyltransferases; tetratricopeptide repeat

Protein prenyltransferases (PTs) catalyze the transfer of one or two isoprenyl groups from either farnesyl diphosphate or geranylgeranyl diphosphate to the C-terminal cysteins in a variety of proteins including nuclear lamins, trimeric G protein γ-subunits, Ras, and nearly all Ras-related GTPases (Casey & Seabra, 1996; Zhang & Casey, 1996; Seabra, 1998). Prenylation is essential for the membrane localization and thus the function of these proteins, most of which play critical roles in important cellular processes such as signal transduction and vesicular traffic regulation (Zhang & Casey, 1996). There are three subfamilies of the PTs: protein farnesyltransferase (FT), geranylgeranyltransferase type I (GGT1), and Rab geranylgeranyltransferase (RabGGT or GGT2). Extensive biochemical and structural information is available on PTs, because FT is a prime target for anticancer drugs (Gibbs & Oliff, 1997). All PTs are composed of two tightly associated α- and β-subunits. The catalytic function resides on the β-subunit of PTs (PTβ) and requires a Zn^{2+} ion (Chen et al., 1993). The α-subunit in PTs (PTα) is also required for the activity (Andres et al., 1993), but its role is less clear. The structure of rat FT complexed with farnesyl diphosphate suggests that FTα participate in prenyl diphosphate binding (Long et al., 1998). It has been shown that FT and GGT1 share a common α-subunit (Seabra et al., 1991), while the α-subunit of GGT2 is distinct but homologous, and may include additional domains (Seabra et al., 1992; Armstrong et al., 1993). The crystal structure of the rat FT reveals that the α-subunit contains 15 helices that fold into a right-handed crescent-shaped superhelix (Park et al., 1997). The β-subunit is an α-α barrel of six helical pairs. The concave surface of the α-subunit superhelix embraces the β-subunit at the open end of the barrel around half of its circumference. The sequences of PTα are characterized by the presence of seven repetitive motifs. The first five repeats are easily detectable from the sequence (Boguski et al., 1992), while the sixth and seventh repeats are more divergent. The PTβ sequences are composed of six repetitive motifs that are different from those in PTα (Boguski et al., 1992).

PTs have been identified only in eukaryotes. It would be interesting to know whether they have prokaryotic homologs, which will shed light on evolution of this interesting and unique group of enzymes. Because the α- and the β-subunits have distinct folds, they are likely to originate from different ancestral proteins. In this paper we focus on the PTα. Folds similar to the PTα have been observed in several other proteins, such as bacterial muramidase (Thunnissen et al., 1994) and lipovitellin (Anderson et al., 1998). However, structural consideration alone does not allow establishment of evolutionary relationship between these proteins and PTα, because the structural pattern is very simple (right-handed super-

Reprint requests to: Nick V. Grishin, National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, Maryland 20894; e-mail: grishin@ncbi.nlm.nih.gov.

helix) and repetitive, and might occur independently by convergence. In this case, only similarity at the sequence level would indicate homology. Indeed, statistically significant sequence similarity between PTα and the tetratricopeptide repeat (TPR) motif was detected, which suggests that PTα is a member of the TPR superfamily.

The TPR motif is defined as a degenerate 34-amino acid sequence characterized by eight loosely conserved residues (-W_4-L_7-G_8-Y_{11}-A_{20}-F_{24}-A_{27}-P_{32}-; the subscripts denote the TPR numbering in Sikorski et al., 1990). TPRs are widely spread among all organisms, from Eubacteria to Archaea and Eukarya, and occur in proteins that perform diverse functions such as cell cycle regulation, transcriptional repression, signal transduction, stress response, mitochondrial and peroxisomal protein transport, protein secretion, and DNA replication (Chen et al., 1994; Lamb et al., 1995; Gindhart & Goldstein, 1996; Koonin et al., 1996). Mutagenesis and deletion studies revealed a role for TPRs in mediating specific protein–protein interactions (Lamb et al., 1994; Smith et al., 1995; Tzamarias & Struhl, 1995). Recently, the crystal structure of the TPR domain of protein phosphatase 5 (PP5) has been determined (Das et al., 1998). The PP5-TPR domain contains three TPR motifs and adopts an overall structure of a right-handed superhelix with each TPR composed of a pair of antiparallel α-helices.

Here we describe the sequence similarity between repetitive motifs in PTα and TPR. Our finding that PTα is a TPR protein makes the crystal structure of FT the first structural model of a TPR domain interacting with its protein partner. We compare the structures of FTα and PP5-TPR, especially the interactions within and between TPRs. These analyses provide insight into the mechanism of the TPR-mediated protein recognition. Finally, we propose a general approach to the evolutionary analysis of proteins containing repetitive motifs and apply it to PTα and other TPR proteins.

Results and discussion

Sequence similarity between PTα and TPR proteins

When the sequence of rat FTα (NCBI gene identification number gi|417481, residues 89–377) is used as a query for the gapped BLAST search (Altschul et al., 1997), the sequence gi|2621120 from *Methanobacterium thermoautrotrophicum* is found with a score of 64 and an E-value of 1e-09. This score is higher than that with the yeast FT, RAM2 (gi|266880). *M. thermoautrotrophicum* protein gi|2621120 contains divergent TPR motifs, reminiscent of those in O-linked N-acetylglucosamine transferase (OGT) of eukaryotes (Lubas et al., 1997). When iterative PSI-BLAST search (Altschul et al., 1997) is carried out with the rat FTα and a stringent threshold (E-value of 1e-06, instead of the default value of 1e-03), known TPR proteins, such as CDC27 (gi|231708), NUC2 (gi|2135326), and SSN6 (gi|283218) are found above the threshold after the first iteration. The TPR domain of PP5 with the known structure (PDB entry 1a17) is found on the second iteration with the score of 44 and E-value less than 0.001. Similar results are obtained when a number of other PTα sequences are used as queries for BLAST searches. These results strongly suggest that the PTα is a member of the TPR protein superfamily.

Interestingly, although *M. thermoautrotrophicum* protein gi|2621120 is closest in sequence to the human and *Caenorhabditis elegans* OGT and therefore is annotated as an O-linked N-acetylglucosamine transferase, its sequence consists entirely of TPRs and lacks the OGT catalytic domain. Orthologs of the OGT catalytic domain are not detected in *Archaea*. Therefore, the protein gi|2621120 in *M. thermoautrotrophicum*, termed here TPR Mt, must have a different function.

The TPRs usually exist as tandem arrays of 3–16 motifs in the polypeptide chains that can be composed entirely of TPRs, such as TPR Mt, or have additional domains located either outside the TPR domain, such as in PP5 Hs, or inserted between TPRs, such as in GGT2 Rn and NUC2 Sp (Figs. 1, 2). The multiple sequence alignment of TPRs, including PTα, is shown in Figure 1. The TPR motif has been defined as a helical pair A_iB_i, where A and B are α-helix names, and index i denotes a repeat number along the sequence (Sikorski et al., 1990; Lamb et al., 1995). However, in the current alignment we present repeat units as helical pairs B_iA_{i+1}, in agreement with an earlier definition when repeats in PTα were first described (Boguski et al., 1992). Although different from the conventional TPR definition, it seems more appropriate for PTα because the linker sequence is better conserved between the helices B_iA_{i+1} than between helices A_iB_i. The longest insertion in the B_iA_{i+1} linker is nine residues long (FT Sc, Fig. 1), whereas large insertions and even additional domains are placed between helices A_i and B_i. Indeed, a 22-residue insertion is present between A_2 and B_2 in GGT2 from *C. elegans* (Fig. 1), and GGT2 from animals (rat, human, and *C. elegans*) accommodates a globular domain of more than 100 residues between helices A_6 and B_6 and a leucine-rich repeat domain (Kobe & Deisenhofer, 1994; Kajava et al., 1995; Kajava, 1998) after the last helix, A_8 (Fig. 2).

The TPR sequence motif is highly degenerate (Fig. 1). Although most of the consensus residues are preserved within PTα family, significant differences exist. Namely, sites 7 and 20 in repeats 2 to 5 are occupied by invariant Arg and Glu, respectively, instead of conserved Leu and Ala in typical TPR (Fig. 1). There are also additional invariant residues in PTα outside the TPR consensus sequence (Fig. 1). To understand the structural and functional significance of these variations, we compared the crystal structures of FTα (Long et al., 1998; Protein Data Bank (PDB) entry 1ft2) and PP5-TPR (Das et al., 1998; PDB entry 1a17).

Structural analysis of TPRs

The crystal structures of FTα and PP5-TPR show that an individual TPR folds into a pair of antiparallel α-helices (helices A_i and B_i), with TPR consensus residues involved in the packing of these two helices, in agreement with the sequence-based prediction (Sikorski et al., 1990). Adjacent TPRs are packed together in a parallel arrangement resulting in a superhelical pattern (TPR superhelix) (Fig. 3A,C). The concave surface of the superhelix is formed by the helices A in each TPR, while the helices B packed on the outside. The overall structures of FTα and PP5-TPR show substantial differences (Fig. 3A–D), resulting in root-mean-square deviation (RMSD) of 4.7 Å between C_α atoms of 97 residues as calculated by DALI (Holm & Sander, 1993). To understand the structural basis for this large deviation, we studied interactions between neighboring helices A_iB_i, B_iA_{i+1}, A_iA_{i+1}, and B_iB_{i+1}. The results of comparison of all possible helical pairs in FTα and PP5-TPR are shown in Table 1.

In FTα, the conformation of all B_iA_{i+1} helix pairs (i = 1 to 7) is well conserved, as indicated by the small averaged RMSD (1.18 Å) and standard deviation (SD) (0.25 Å) among 21 possible superpositions (Table 1, section I). The conformations of other three classes of helix pairs, including the TPR motif A_iB_i, are more variable, as indicated by the higher RMSD (2.36 Å) and SD

(1.19 Å) (Table 1, section I). The pairwise comparisons show that the first TPR A_1B_1 is rather similar to the seventh TPR A_7B_7 (RMSD 1.7 Å), while the middle five A_iB_i pairs ($i = 2, 3\ldots,6$) are similar to each other (RMSD about 1.0 Å). These two groups of helical pairs in FTα are different from each other (RMSD more than 3 Å). The case of PP5-TPR reveals the same trends as in FTα. Although there are not enough superpositions to establish sound statistics, we see that the helix pair B_iA_{i+1} has the most conserved conformation as indicated by the smallest SD in combination with small RMSD (Table 1, section II).

The conformations of all helical pair combinations in FTα were then compared to those in PP5-TPR. The results (Table 1, section III) show that between the two structures, the conformation of B_iA_{i+1} is also best conserved (averaged RMSD 1.19 Å, with SD 0.33 Å, very similar to those within FTα), while conformations of the other three helix pairs are more different (averaged RMSD more than 2.5 Å; see also Fig. 3B,D). The structural comparison of individual TPR motifs A_iB_i in the two structures reveals that the A_7B_7 of FTα is most similar to A_iB_i in PP5-TPR (RMSD of 1.34, 1.26, 1.04 Å to TPR1, 2, and 3 in PP5), A_1B_1 is less similar (RMSD about 2 Å), and the middle five motifs are very different (RMSD more than 3 Å).

In summary, our results indicate that the mutual arrangement of the helices B_i and A_{i+1} is highly conserved in all TPR, regardless of the particular structure they come from. On the contrary, the conformations of the TPR helical pairs A_iB_i are variable not only between FTα and PP5-TPR, but within FTα as well. This phenomenon could be explained by the differences in packing between helices in the two structures (Fig. 4). We compared the helix packing in FTα and PP5-TPR with those in the typical parallel and antiparallel coiled coil structures (Table 1, sections IV and V).

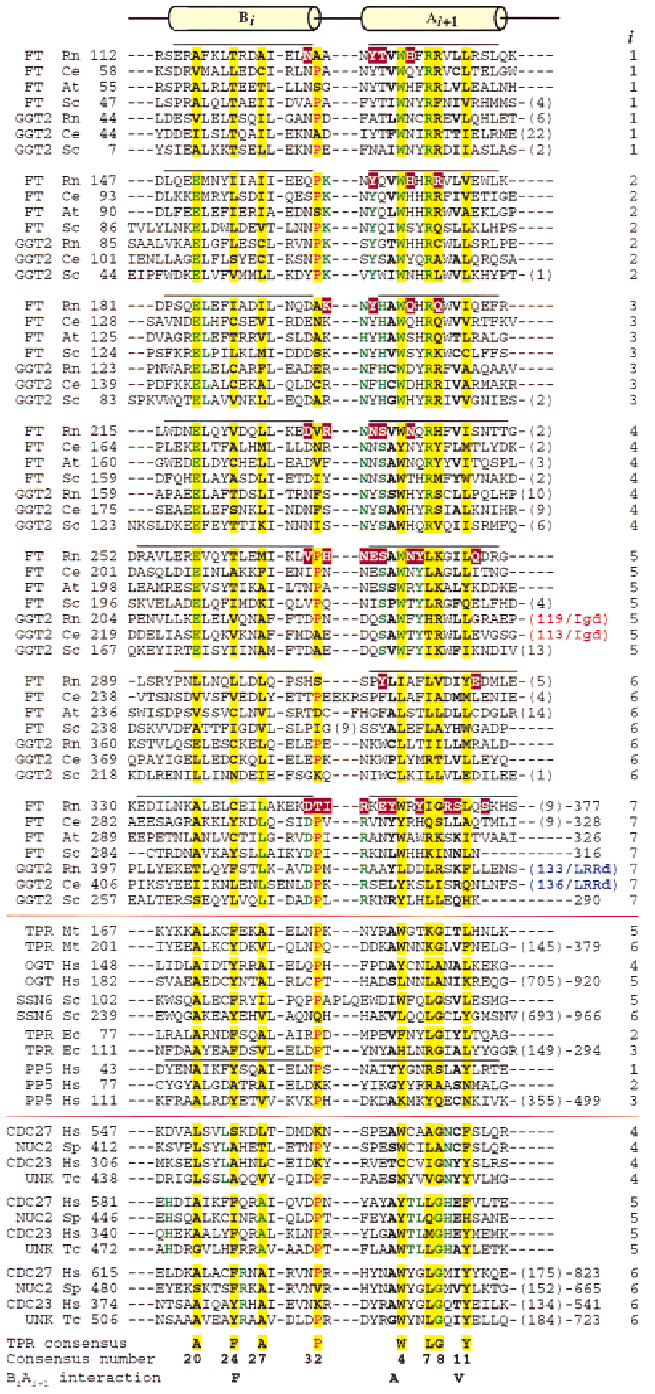


Fig. 1. Sequence alignment of PT-α with other TPR proteins. The alignment includes the representative proteins from PT family (above the first red line), CDC27/23 family (below the last red line), and several divergent TPR proteins (between the red lines). Sequence names consist of a protein name and a two-letter species name abbreviation. The protein names used are: FT, a farnesyltransferase/geranylgeranyltransferase type I α-subunit; GGT2, a geranylgeranyltransferase type II α-subunit; TPR, TPR protein; OGT, O-linked GlcNAc transferase; SSN6, SSN6 protein; PP5, Set/Thr protein phosphatase 5; CDC27, cell division cycle protein 27; NUC2, nuclear scaffold-like protein; CDC23, cell division cycle protein 23; UNK, unknown. Species name abbreviations are: At, thale cress (*Arabidopsis thaliana*); Ce, worm (*C. elegans*); Ec, (*Escherichia coli*); Hs, human (*Homo sapiens*); Mt, (*M. thermoautotrophicum*); Rn, rat (*Rattus norvegicus*); Sc, yeast (*Saccharomyces cerevisiae*); Sp, fission yeast (*Schizosaccharomyces pombe*); Tc, (*Trypanosoma cruzi*). The NCBI gene identification numbers (gi numbers) for each sequence are: FT Rn, 417481 (PDB entry 1ft2); FT Ce, 2736325; FT At, 3142698; FT Sc, 266880; GGT2 Rn, 730316; GGT2 Ce, 3800984; GGT2 Sc, 728961; TPR Mt, 2621120; OGT Hs, 2266994; SSN6 Sc, 283218; TPR Ec, 1176172; PP5 Hs, 3212250 (PDB entry 1a17); CDC27 Hs, 231708; NUC2 Sp, 1709403; CDC23 Hs, 3283051; UNK Tb, 3063543. Each alignment is broken into TPR repeats, and the repeats are aligned. All seven repeats are shown for PTα; the three most conserved consecutive repeats are shown for CDC27/23, the three repeats with known spatial structure are shown for PP5, and only two consecutive repeats, which are most similar to PT repeats, are shown for other TPR proteins. The sites, which belong to the consensus of TPR repeat, are shaded with yellow, and amino acids are shown in bold. Proline residues in the position 32 are shown in red. The TPR consensus residues are shown below the alignment and the numbers under TPR consensus indicate the position in the repeat as in Sikorski et al. (1990). The sites with residues involved in B_iA_{i+1} interaction are also shown in bold, and their consensus residues are given at the very bottom. The sites, invariant within the whole family (prenyltransferase or CDC23/27), are shown in green bold letters. The α-helices in known structures are shown as lines above the corresponding sequence. The secondary structural diagram above the alignment shows the positions of helices B_i and A_{i+1}. The number to the right of each block indicated the repeat number *i*. The residues in FT Rn involved in the interactions with the β-subunit (within 3.5 Å distance of any residue in β-subunit in PDB entry 1ft2) are printed in white and shaded by red. The numbers of the first residue in every block are shown after the sequence names. The numbers in parenthesis indicate how many residues in an insertion are not shown. The numbers at the end of the sequence indicate the total number of amino acids in the protein. Igd and LRRd label the positions of an inserted globular domain and a leucine rich repeats domain, respectively.

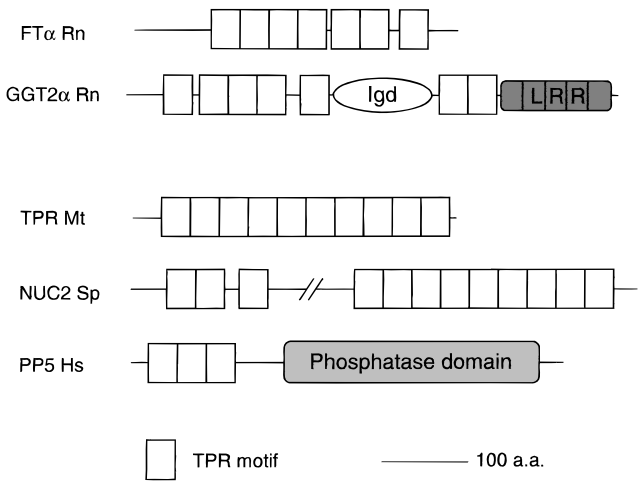


Fig. 2. Domain organization of some TPR proteins. Two PTα are shown above. Sequence and species name abbreviations are explained in the caption to Figure 1. TPR are shown as white rectangles. Igd stands for inserted globular domain; LRR is a leucine reach repeat domain (includes five repeats). The figure is to scale except the region with two slashes.

These comparisons demonstrate that the B_iA_{i+1} packing in FTα, as well as in PP5-TPR, is very similar to that in the typical antiparallel coiled coil. Such packing is observed, for example, in colicin Ia (Wiener et al., 1997) and is characterized by the "ridges-into-grooves" interface (Chothia et al., 1981). This packing is relatively sequence independent. The differences in amino acid sequence cause only moderate variations in all the structures compared (RMSD about 1.2 Å; see Table 1, section IV).

In contrast, most of the helices A_i and B_i in FTα (i = 2 to 6) are much further apart than the helices in the typical antiparallel coiled coil (Fig. 4A). Side-chain–to–side-chain packing dominates the interface between these helical pairs (Fig. 4A). The residues at the TPR position 7 and 20, occupied by highly conserved hydrophobic residues in classical TPR proteins, are Arg and Glu, respectively, in the middle four repeats in FTα. These charged residues are mostly inaccessible to solvent and form interhelix salt bridges between helices A_i and B_i (Fig. 4A). They also interact with neighboring hydrophobic residues through hydrophobic part of their long side chains. These polar substitutions and the invariant Trp at position 4 largely dictate the relative disposition of helices A_i and B_i.

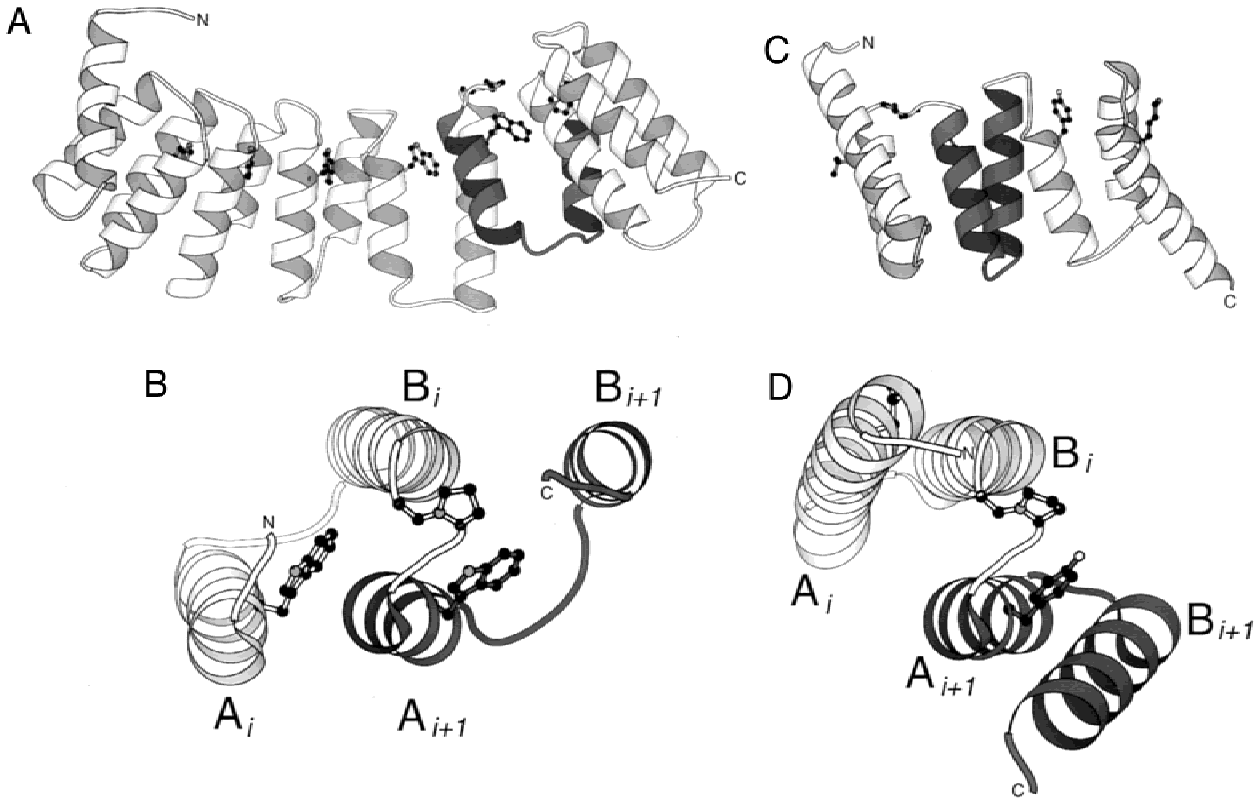


Fig. 3. Ribbon presentation of TPR structures. **A:** Overall structure of Rat FTα (residues A90–A369, PDB entry 1ft2). **B:** Four neighboring helices (residues A232–A306) from FTα. **C:** Overall structure of PP5-TPR (residues 19–148, PDB entry 1a17). **D:** Four neighboring helices (residues 20–93) from PP5-TPR. The four neighboring helices in **B** and **D** are selected from each structure with the helical pair B_iA_{i+1} superimposed and separated for clarity. One of the TPR repeats ($A_{i+1}B_{i+1}$) is shaded. Conserved Pro residue in the turn between helices B_i and A_{i+1}, and the side chains of conserved Trp residues in FTα and corresponding residues in PP5 are shown in ball-and-stick presentation. The N- and C-termini of the structural segments are labeled by letters N and C, respectively. Figures 3, 4, and 5 are drawn by BOBSCRIPT (Esnouf, 1997), a modified version of MOLSCRIPT (Kraulis, 1991).

Table 1. *Analysis of RMSD in TPR repeats and other pairs of helices*[a]

Superimposed structures	I FTα vs. FTα			II PP5 vs. PP5			III FTα vs. PP5			IV FTα vs. coil			V PP5 vs. coil		
	Mean	SD	*N*	Mean	SD	*N*	Mean	SD	*N*	Mean	SD	*N*	Mean	SD	*N*
A_iB_i	2.36	1.19	21	0.75	0.26	3	2.98	0.90	21	3.56	0.96	49	1.95	0.14	21
B_iA_{i+1}	**1.18**	**0.25**	21	0.87	0.11	3	**1.19**	**0.33**	21	**1.24**	**0.34**	49	0.86	0.19	21
A_iA_{i+1}	2.49	1.35	21	1.34	0.35	3	2.50	0.53	21	2.64	0.51	14	1.40	0.39	6
B_iB_{i+1}	2.14	0.92	15	0.83	—	1	2.60	0.51	12	3.73	0.43	12	3.25	0.20	4

[a]The rows are attributed to the class of a helical pair and columns summarize the information on a group of superimposed structures. The RMSD values (in Å) are given for the mean and the standard deviation (SD) of superpositions among the helical pairs within each class and group. *N* stands for the number of superpositions performed. The standard deviation of the mean (standard error) can be estimated as $SD/N^{0.5}$. Structure of colicin A (PDB entry 1cii) is used as an example of antiparallel coiled coil for comparisons of B_iA_{i+1} and A_iB_i. The structure of leu-zipper GCN4 (PDB entry 2zta) is used as an example of parallel coiled coil for comparisons of A_iA_{i+1} and B_iB_{i+1}.

In contrast to FTα, in PP5-TPR the helices A_i and B_i (Fig. 4C) are closer to each other than those in typical antiparallel coiled coil, resulting in slightly larger deviations (RMSD 1.95 Å) when compared to colicin (Table 1, section V). Several highly conserved small residues (such as glycines at positions 8 and alanines at 20 and 27) are absolutely critical for this tight packing.

The conformation of two groups of parallel helix pairs A_iA_{i+1} and B_iB_{i+1} is also different between FTα and PP5, as shown in Figures 3B and 3D. A_iA_{i+1} in PP5-TPR resembles typical parallel coiled coil such as the leucine zipper (RMSD 1.4 Å; Table 1, section V), in which the two helices form a left-handed coil with the "knobs-into-holes" packing of side chains (Crick, 1953; O'Shea et al., 1991). The helices A_iA_{i+1} in FTα, however, form a right-handed coil resulting in larger deviation from the leucine zipper (RMSD 2.64Å; Table 1, section IV; Fig. 3B). The invariant Trp at position 4 in helix A_i of FTα interacts with both helices B_i and A_{i+1}, and is completely buried in the hydrophobic core formed by the three helical bundle $A_iB_iA_{i+1}$. Other important interactions between helices A_iA_{i+1} in FTα include hydrogen bonding between an Asn side chain (N199 or N233) near the N terminus of A_i and a main-chain carbonyl near the N terminus of helix A_{i-1}. These Asn residues are invariant among all PTs (Fig. 1), emphasizing the importance of these interactions for the structure and function of the protein. Helices B_iB_{i+1} are arranged in a similar manner to A_iA_{i+1} in FTα, but do not interact with each other in PP5-TPR (Fig. 3B,D).

Mechanism of TPR-mediated protein–protein interactions

The crystal structure of FT provides a structural model of interactions between a TPR domain and its protein partner, the β subunit of FT. All residues in FTα that interact with FTβ are located

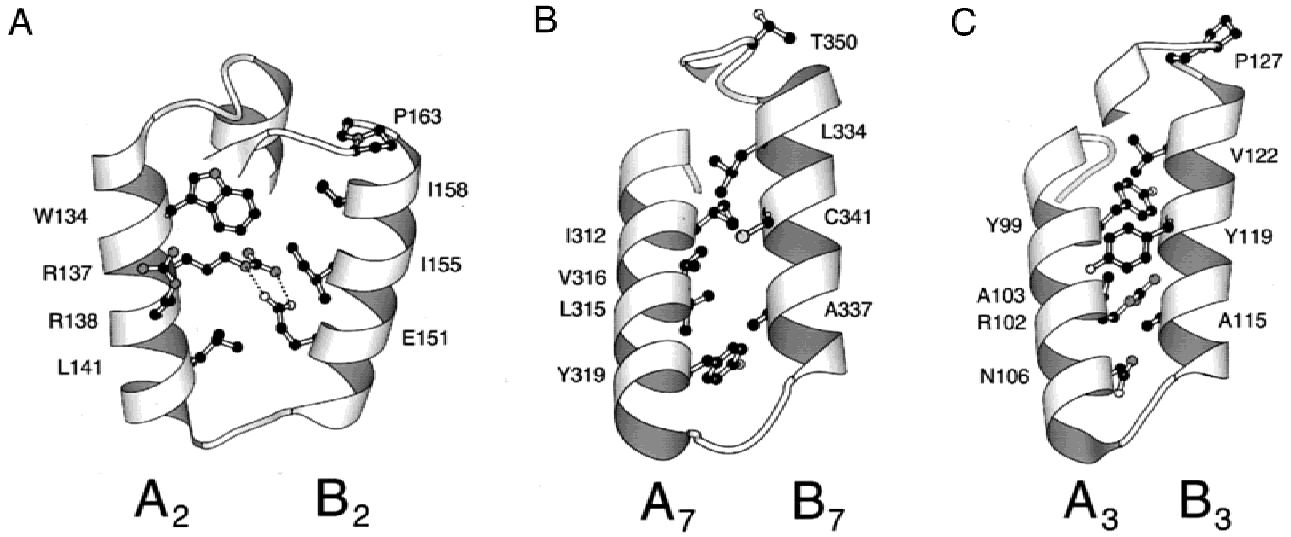


Fig. 4. The packing of consensus residues in TPR structures. Side chains of all eight consensus residues in a repeat A_iB_i are displayed and labeled. Each helix in a repeat is labeled by its name and repeat number. **A:** The packing in the typical repeat of FT (1ft2, residues A121–A166). Hydrogen bonds between R137 and E151 are shown in dotted lines. **B:** The packing in the terminal (7th) repeat of FT (1ft2, residues A307–A351). **C:** The packing in a typical TPR (1a17, residues 91–132).

on the concave inner surface of the TPR superhelix, mostly on one side of the helices A_i, but also in the loop regions connecting helices B_i and A_{i+1} (Figs. 1, 5). This extensive interface buries about 3,300 Å of solvent-accessible surface (Park et al., 1997), which is typical for an oligomeric protein (Janin et al., 1988).

The packing dictated by the unique PTα consensus sequences leads to the unusual conformation of FTα superhelix, which is well suited for the interaction with the large α-α barrel of FTβ. First, the replacement of the Leu and Ala at positions 7 and 20 for Arg and Glu, respectively, is largely responsible for the deviation from the typical TPR motif packing. Second, invariant residues outside the eight TPR consensus sites in the PT family (colored green in Fig. 1) have important structural and functional roles. Some of these residues are involved in stabilizing the unique conformation of the TPR superhelix as discussed above (e.g., N199, N233). Others are directly involved in the interactions with the β-subunit (Figs. 1, 5). Additionally, an invariant K164 is involved in the substrate binding and thus is critical for the enzyme activity (Andres et al., 1993; Long et al., 1998).

The interaction surface in FTα is in principle similar to that proposed for PP5 TPR (Das et al., 1998). However, the predicted TPR superhelix model by Das et al. (1998) has a smaller curvature and a larger twist than those in the FTα superhelix. As a result, the continuous groove formed by the helices A_i in this TPR model is suited for accommodating a single α-helix. In FTα, this interacting surface forms extensive contacts with a protein subunit consisting of 437 residues. It is likely that the residues in direct contact with the TPR protein partner should contribute to the interaction specificity. Mutagenesis data indicate that TPR consensus residues are crucial for the intermolecular interactions as well (Lamb et al., 1994). A single mutant in CDC27p Sc (G613D) results in a greatly reduced ability to interact with CDC23p, but has no effect on interactions with wild-type CDC27p and CDC16p (Lamb et al., 1994). In FTα all TPR consensus residues are involved in the packing between two TPR helices A_i and B_i (Fig. 4), but not directly in contact with FTβ. Our analysis suggests that varying residues in TPR consensus sites would change the relative arrangement of helices A_i and B_i, which in turn would alter the overall shape of the TPR superhelix and thus its interacting surface.

Sequence conservation pattern (Fig. 1) suggests that most TPR protein domains must have structure more similar to the PP5-TPR superhelix than to FTα. However, the high level of degeneracy in the TPR consensus sequences indicate that there could be a wide range of varying conformation of TPR superhelix, as has been demonstrated in the PP5-TPR and FTα. As a result, different TPR proteins could interact with a wide range of protein targets of different sizes and properties (Lamb et al., 1995; Tzamarias & Struhl, 1995).

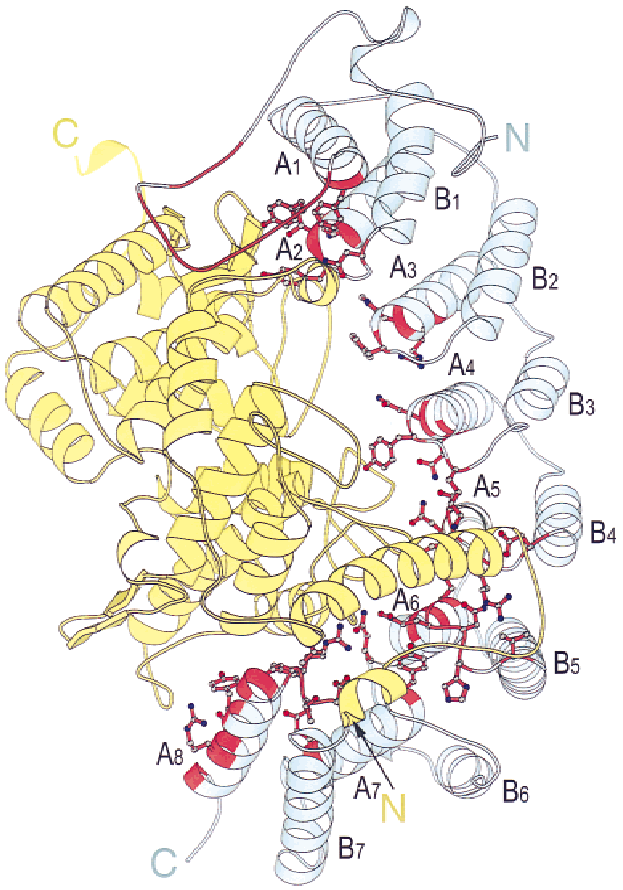


Fig. 5. The interactions between α- and β-subunits in rat FT. The α-subunit is colored cyan, the β-subunit yellow. The residues in FTα that are within 3.5 Å from any atom of FTβ are colored red. Amino acid side chains in FTα that are in contact with FTβ are shown in ball-and-stick presentation. The N- and C-termini of both subunits are labeled in corresponding color. The helices in FTα are labeled according to Figure 1.

The evolutionary history of PTα

The analysis of the evolutionary relationship between proteins containing repetitive motifs is complicated because it is problematic to obtain adequate sequence alignments. In principle, any repeat can be aligned with any other homologous repeat. For the meaningful phylogenetic analysis, it is important to find the correct register of the repeats between the protein sequences.

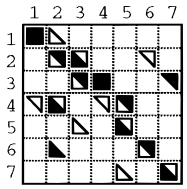
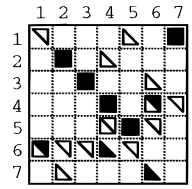
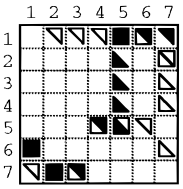
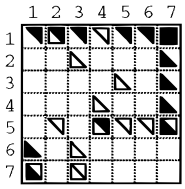
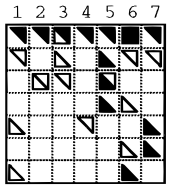
It has been pointed out by Kobe and Deisenhofer (1994) that two evolutionary scenarios are possible for proteins with repeats: (1) the common ancestor of the present-day proteins already contained multiple repeats or (2) the common ancestor was a single repeat protein, and the multiple repeats in present-day proteins originated independently. These two different evolutionary pathways can be distinguished by comparing sequence identity (or similarity) between repeats WITHIN one protein (*self-identity*) to that between corresponding repeats BETWEEN different proteins (*cross-identity*), and also to that between any but corresponding repeats BETWEEN different proteins (*shifted-identity*). *Cross-identity* greater than *shifted-identity* and *self-identity* of both proteins indicates that the two proteins most likely evolved from a common ancestor that already contained multiple repeats. *Self-identity* of both proteins higher than *cross-identity* and *shifted-identity* implies that multiplication occurred independently in each protein unless substitutions in different repeats are correlated. If "molecular clock" hypothesis is not valid, then *self-identity* for one protein in a pair of proteins that are related by the second scenario might be lower than the *cross-identity* and *shifted-identity*. The second scenario also allows the number of the repeats to vary between different protein families. In reality, various combinations of the two scenarios are possible. Additionally, the picture might be complicated by the possibility of gene conversion among repeats. However, the repeat conversion event is equivalent to the deletion of one repeat copy and duplication of another, which

replaces the deleted one. In this sense conversion is equivalent to the repeat duplication, because the trace of the deleted repeat is lost. Therefore, if gene conversion occurred in evolution of TPR repeats, our methods will detect the divergence point of the last conversion event that took place. If conversion events are rare, there will be a significant heterogeneity in similarity between individual repeats in a protein, because the repeats that are related by conversion would be more similar to each other. Our analysis (see below) have shown that this is not the case. If repeat conversions are very frequent, then the repeats will be very similar to each other, and we will not be able to find the repeat-specific conserved regions between orthologous proteins. These conserved regions exist in most TPR proteins we have analyzed. Therefore, frequent gene conversion is also unlikely for most TRP proteins.

To understand the evolutionary relationship between PTα and other TPR proteins, we analyzed the *self-identity*, *cross-identity*, and *shifted-identity*. The results are shown in Table 2. Analogously defined similarity scores instead of identities lead to the same conclusions (data not shown). Within the PTα family, there are two subfamilies: FT/GGT1α and GGT2α. In these two subfamilies, the cross-identity (23.1%) is significantly higher than the shifted identity (16.6%) or the self-identity (~16%) (Table 2, section I). Therefore, the FTα and GGT2α originated from an ancestor that already contained multiple TPR motifs as a present-day PTα (first scenario). Circular permutation analysis reveals that when the first and the last (seventh) repeats in PTα are aligned, the shifted identities are higher than in other cases (~18% vs. ~15%), which implies a more complicated evolutionary history in which the seven TPR motifs in PTα might have heterogeneous origins. The same first evolutionary scenario is also conceivable for the two subfamilies of typical TPR proteins: CDC23 and CDC27, as indicated by the significantly higher cross-identity over self-identity and shifted identity (26.8% vs. 14.5%) (Table 2, section II).

When the identities are compared between PTα and CDC23/27 families, the cross-identity (~11%) is lower than the self-identity (14.3% in CDC23/27 and 16.0% in PTα) and statistically equal to the shifted-identity (~11%) (Table 2, section III). Additionally, the number of TPRs in these protein families is different (Fig. 2). It is, therefore, likely that they evolved mainly through the second evolutionary scenario, i.e., they have a common single ancestral TPR that diverged later in each family. Dot-plot analysis reveals additional information about the individual repeats in the two families

Table 2. *Analysis of identity percentage in TPR repeats* [a]

	I			II			III			IV			V		
Groups	GGT2—FT			CDC27—CDC23			PT—CDC27/23			PT—different TPR			PT—MtTPR		
Group numbers	1	2		1	2		1	2		1	2		1	2	
Number of sequences	4	10		6	4		14	10		14	5		14	2	
	Mean	SD	*N*	Mean	SD	*N*	Mean	SD	*N*	Mean	SD	*N*	Mean	SD	*N*
Analysis of individual repeats															
Self group 1	16.3	7.1	84	14.3	5.7	126	**16.0**	8.3	294	**16.0**	8.3	294	16.0	8.3	294
Self group 2	15.8	8.7	210	14.2	7.7	84	**14.3**	6.6	210	**28.9**	18.1	105	**56.2**	13.4	42
Cross	**23.1**	8.1	280	**26.8**	11.6	168	10.6	4.6	980	13.1	6.5	490	**17.4**	6.0	196
Shifted	16.6	7.9	1,680	14.5	6.1	1,008	10.9	5.3	5,880	12.9	6.2	2,940	**17.2**	6.6	1,176
Circular permutation analysis for the entire sequence															
Cross	**22.9**	1.7	40	**26.7**	2.7	24	10.6	1.4	140	13.1	3.3	70	**17.4**	2.3	28
Shift 1 repeat	18.8	1.4	40	14.5	2.0	24	11.0	1.7	140	12.9	3.5	70	**17.7**	2.1	28
Shift 2 repeats	14.8	1.5	40	15.4	2.3	24	11.7	1.5	140	12.8	3.1	70	**17.7**	2.2	28
Shift 3 repeats	14.0	2.0	40	13.7	1.5	24	10.7	1.7	140	12.8	3.3	70	**16.7**	2.3	28
Shift 4 repeats	16.2	2.3	40	13.3	1.5	24	11.1	1.8	140	12.9	2.8	70	**17.1**	2.4	28
Shift 5 repeats	16.8	1.8	40	13.5	1.9	24	10.1	1.7	140	12.8	3.1	70	**16.8**	2.0	28
Shift 6 repeats	18.3	1.2	40	16.5	1.7	24	10.5	1.6	140	13.2	3.0	70	**17.5**	1.8	28
Mean for shifts	16.5	1.7	6	14.5	1.2	6	10.9	0.5	6	12.9	0.1	6	**17.2**	0.4	6
Dot-plot: repeat vs. repeat															

[a]Each alignment includes seven repeats. The "different TPR" group includes sequences from TPR Mt (gi|2621120), OGT Hs (gi|2266994), SSN6 Sc (gi|283218), TPR Ec (gi|1176172), and CDC27 Hs(gi|231708). The standard deviation (SD) indicates variability of identity percentage between repeats and/or between sequences in a group. *N* is the number of comparisons performed. The standard deviation of the mean (standard error) can be estimated as $SD/N^{0.5}$. The "mean for shifts" row contains the mean and SD from the mean values of shifts 1 to 6. In dot-plots the rows refer to repeat in group 1, and columns refer to the repeat in group 2. The black and white triangles stand for the largest and the second largest identity in a row (column), respectively. Pointed down (◺) and up (◹) triangles refer to the rows and colums, respectively.

(Table 2, bottom). The fifth TPR (residues 615–649 of CDC27 Hs in Fig. 1) in CDC23/27 family is usually more similar to the TPRs of PTα, while the first and seventh repeats in PTα are more similar to the repeats in CDC23/27.

Similar results are obtained when the PTα family is compared to a group of diverse TPR sequences (Table 2, section IV), including TPR Mt (gi|2621120), OGT Hs (gi|2266994), SSN6 Sc (gi|283218), TPR Ec (gi|1176172), and CDC27 Hs(gi|231708) (protein name abbreviations are explained in Figure 1 legend, NCBI gi numbers are given in parenthesis). The average self-identity in these TPR proteins is much higher than that in PTα (28.9 vs. 16%), while the shifted and cross-identity are low (~13%). Dot-plot analysis shows that the first repeat in PTα is somewhat more similar to the TPR repeats in other TPR proteins. This result is consistent with the structural consideration that the first repeat in FTα has a conformation more similar to the typical TPRs such as those in PP5-TPR.

Even comparison of PTα with their closest sequence neighbors, such as TPR proteins from *M. thermoautrotrophicum* (gi|2621120, gi|2621106) does not reveal that their common ancestor shared more than a single repeat with them (Table 2, section V). However, two unusual facts are noticeable. First, the self-identity in *M. thermoautrotrophicum* TPRs is very high (56.2%). Second, the value of the cross-identity that is about the same as the shifted-identity, is higher than the self-identity among PTα (17 vs. 16%). This is indicative of substitution rate differences between lineages and breaking down of the molecular clock hypothesis. Indeed, these data indicate that highly similar repeats in *M. thermoautrotrophicum* TPRs might be closer to the ancestral repeat of all TPR proteins, including PTα. This might be the reason for the higher similarity between *M. thermoautrotrophicum* TPRs and PTα, compared to that between typical TPRs and PTα.

To verify the conclusions from the analysis of *self-*, *cross-*, and *shifted*-identity percentages, we used evolutionary tree reconstruction methods. Based on the multiple sequence alignment of PTα, a phylogenetic tree was constructed using the PHYLIP package as described in Methods (Fig. 6A). This unrooted tree shows clear separation between FT/GGT1 on one side and GGT2 on the other. Every branching point is supported by high bootstrap values. The sequences of animal GGT2 proteins that contain additional domains, such as Igd and LRRd (Figs. 1, 2) are grouped together. The tree construction is based entirely on TPR sequences without using information from the additional domains. Because these additional domains are present in such diverged animals as rat and worm, and not present in yeast GGT2 as well as in all FTs, we hypothesize that they were inserted in the GGT2 sequence during or before the divergence of animals, at least before the separation of the worm and vertebrate lineages. Curiously, plant GGT2 sequences are not yet known.

Generally, to root the tree, a divergent but rather similar homolog is used. Because PTα sequences contain repeats, and homology between repeats within one sequence reflects evolutionary events, preceding the separation between FT and GGT2, we can use this information to root the PTα tree. The self-identity in PTα higher than cross-identity with other TPR proteins ensures that the repeats in PT sequences are the closest homologous sequences. Additionally, we can use other TPR proteins to place the root. Each repeat is too short as an individual unit (34 amino acids) to use for tree reconstruction; therefore, we utilized the entire sequence aligned with the circularly permutated self (see Methods). The tree that includes representative FT, GGT2, OGT, CDC23, and CDC27 sequences is shown in Figure 6B. It is clear that sequences of PTα subfamilies (FTα and GGT2α) are grouped with circularly permuted selves and are separated from the other TPR sequences with the bootstrap value 85%. The same is true for the CDC23 and CDC27 subfamilies and for the sequence of OGT. Therefore, the tree supports the conclusion derived from the analysis of identity percentages that the common ancestor of all these proteins was a single repeat protein (second scenario) and each of the subfamilies in PT and CDC27/23 families had a multirepeat ancestor (first scenario).

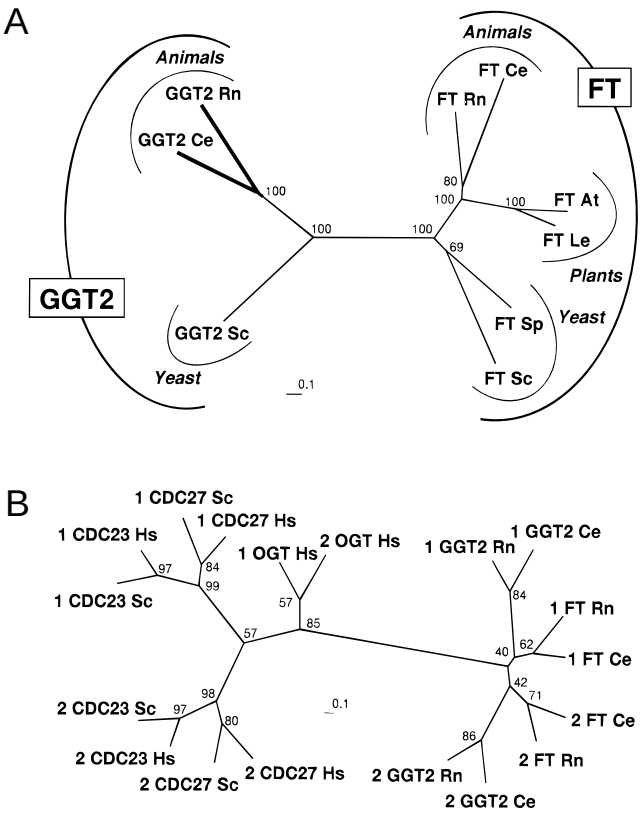


Fig. 6. Phylogenetic trees for PTα and TPR proteins. Sequence name abbreviations are according to Figure 1. Additional species names include Le, tomato (*Lycopersicon esculentum*), and gi numbers for the sequences not shown in the alignment on Figure 1 are: FT Le, 1815666; CDC27 Sc, 584897. Branch lengths are drawn to scale with evolutionary distances. Bootstrap values, based on 1,000 replications, are shown. **A:** An unrooted tree of PT α-subunit. Thicker lines indicate sequences containing Igd and LRRd. Sequences of several mammalian PTs very closely related to rat sequence, namely FT from mouse (2497463), bovine (266752), human (1346694), and GGT2 from human (2497464) are not shown for clarity. **B:** The tree of PTα and other TPR proteins. The tree is built from the alignment, which included each sequence twice (1 and 2): each sequence (1) was aligned with the circularly permuted self (2) (see Methods).

Methods

Sequence analysis

Sequence similarity searches were performed using gapped BLAST and PSI-BLAST programs (Altschul et al., 1997) on the nonredundant (NR) protein sequence database of the National Center for

Biotechnology Information (NIH, Bethesda, Maryland). Multiple alignments were constructed with the ClustalW program (Thompson et al., 1994) and adjusted manually at certain regions to fit PSI-BLAST local alignments and to avoid gaps in the secondary structure elements. The alignments were broken into individual TPR repeats and the repeats were self-aligned to match TPR consensus, PSI-BLAST local alignments and structural equivalents when structure is available. The resulting alignments were used to calculate average identity percentage and similarity score between a pair of protein families. Identity percentage in a segment of a sequence pair alignment is defined as the number of identical residues in the alignment divided by the number of sites without gaps. Blosum62 matrix (Henikoff & Henikoff, 1992) was used to assign the similarity scores to every pair of residues.

A program in C language was written by the authors to perform the calculation. Namely, for two protein families X and Y of n and m sequences each, respectively, and k TPR repeats (defined as B_iA_{i+1}) in each sequence, the mean and the variance of identity percentage (similarity score) were calculated for the following cases.

First, individual repeats were analyzed. *Self-identity*, measured as an identity percent between a repeat j_1 in a sequence s from a family X (or Y) and a repeat j_2 ($j_2 \neq j_1$) in the same sequence s, was calculated for every repeat in every sequence and averaged within a family (X or Y). This resulted in averaging of $nk(k-1)/2$ and $mk(k-1)/2$ measurements for families X and Y, respectively. *Cross-identity* between families X and Y, measured as an identity percent between a repeat j_1 in a sequence s from a family X and corresponding repeat j_1 in a sequence r from a family Y, was calculated in every repeat for every pair of sequences from families X and Y and then averaged. This resulted in knm measurements. *Shifted-identity* between families X and Y, measured as an identity percent between a repeat j_1 in a sequence s from a family X and a repeat j_2 ($j_2 \neq j_1$) in a sequence r from a family Y, was calculated in every repeat for every pair of sequences from families X and Y and averaged. This resulted in $knm(k-1)$ measurements.

Second, the entire sequences were analyzed for the possibility of circular permutations. Cross- and shifted-identities for shifts in $1,\ldots,k$ repeats were calculated for the entire sequence and averaged between sequences instead of averaging between repeats. For example, the shifted identity for shifting u repeats between the sequence s from the family X and sequence r from the family Y is defined as the identity in the alignment of sequence s with a circularly permuted sequence r in which the beginning of repeat u is N-terminal and aligns with repeat 1 of sequence s, the first repeat in r follows the last repeat and the $u-1$ repeat is C-terminal and is aligned with the last repeat k of s. The mean value of identities for shifts 1 to k was calculated. The dot-plots (repeat vs. repeat) were constructed from the average identity percentage of repeat j_1 in a sequence s from family X and repeat j_2 in a sequence r from family Y. Averaging is performed for sequence pairs, not repeats.

Structure analysis

Structural comparisons of α subunit of farnesyl transferase (1ft2 A; Long et al., 1998), three TPR repeats from PP5 (1a17; Das et al., 1998), coiled coil segments of colicin A (1cii; Wiener et al., 1997), and Leu-zipper (1zta A B; O'Shea et al., 1991) were performed with the InsightII package (BIOSYM). The helix core common to every TPR repeat was defined as 14 residues in helix A and 14 residues in helix B (112 main-chain atoms in total). The following residues, in a form of "helix name(residue numbers)," were used for FTα: A1(96–109), B1(113–126), A2(131–144), B2(148–161), A3(166–179), B3(182–195), A4(200–213), B4(216–229), A5(234–247), B5(256–269), A6(274–287), B6(292–305), A7(309–322), B7(334–347), A8(354–367). The following residues were used for 1a17: A1(28–41), B1(44–57), A2(62–75), B2(78–91), A3(96–109), B3(112–125), and A4(130–143). Seven overlapping helical segments, structurally equivalent to TPR repeats, were defined in an antiparallel coiled coil region of 1cii: helix1 is from (267–7k) to (280–7k), helix2 is from (388+7k) to (401+7k), $k = 0, 1, \ldots, 6$. Two overlapping helical segments were defined in a parallel coiled coil of 2zta: helix 1 is chain A from (12–7k) to (25–7k), helix 2 is chain B from (14–7k) to (27–7k), $k = 0, 1$.

For each of the four classes of helical pairs in TPR repeats, namely A_iB_i, B_iA_{i+1}, A_iA_{i+1}, and B_iB_{i+1} (for 1ft2, $i = 1,\ldots,7$, $i \neq 7$ for the class B_iB_{i+1}; and for 1a17, $i = 1,\ldots,3$, $i \neq 3$ for the class B_iB_{i+1}), RMSD between all superpositions of each helical pair to all other nonidentical helical pairs within the class was calculated. For each class the data were averaged in three groups: 1ft2 vs. 1ft2 (a pair from 1ft2 superimposed with a pair from 1ft2), 1a17 vs. 1a17 (a pair from 1a17 superimposed with a pair from 1a17) and 1a17 vs. 1ft2 (a pair from 1a17 superimposed with a pair from 1ft2). For example, for the helical pair B_iA_{i+1} in 1ft2 vs. 1ft2 group, there were $7 \times 6/2 = 21$ superpositions (seven repeats); and for the helical pair B_iB_{i+1} and 1a17 vs. 1ft2 there were $2 \times 6 = 12$ superpositions (two and six pairs of this class presented in 1a17 and 1ft2, respectively). Within each class and group variances were calculated.

For each antiparallel helical pairs B_iA_{i+1} and A_iB_i, the RMSD of its superposition with every helix1–helix2 pair of 1cii was calculated and averaged in two groups: 1a17 vs. 1cii and 1ft2 vs. 1cii. For each parallel helical pairs A_iA_{i+1} and B_iB_{i+1} the RMSD of its superposition with every helix1–helix2 pair of 2zta was calculated and averaged in two groups: 1a17 vs. 2zta and 1ft2 vs. 2zta.

Evolutionary tree construction

The alignments of PTα were used for the unrooted tree reconstruction. To root the tree and to approach the question on evolution of individual TPR repeats, the selected sequences of PTs were aligned with the sequences of other TPR proteins. A second identical alignment was templated onto the original one with the repeat j aligned with the repeat $j+1$ and the last repeat was aligned with the first. The resulting alignment contained each sequence twice: sequence aligned with a circularly permutated self. The alignments were not edited to remove regions with gaps. Evolutionary trees were constructed with the PHYLIP package (Felsenstein, 1996). SEALS package (Walker & Koonin, 1997) was used to reformat alignments to PHYLIP format. The distance method (protdist program) was used with the PAM matrix, and neighbor-joining algorithm (neighbor program) was utilized to construct the tree from the distance matrix. Bootstrap analysis (seqboot, consence programs) was employed to validate statistical significance of branching.

Note added in proof

After submission of this paper, the first plant sequence of a Rab geranylgeranyltransferase (from *Arabidopsis thaliana*, gi|4220541) was released. Similar to animal Rab GGT, it contains Igd and LRR domains in addition to TPRs.

Acknowledgments

We thank Eugen Koonin and Osnat Herzberg for the critical reading of the manuscript. H.Z. was a Howard Hughes Medical Institute associate with Johann Deisenhofer at University of Texas Southwestern Medical Center at Dallas when this work was initiated and is currently supported by the NIH Grant GM7890-01 to Osnat Herzberg.

References

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res 25*:3389–3402.

Anderson TA, Levitt DG, Banaszak LJ. 1998. The structural basis of lipid interactions in lipovitellin, a soluble lipoprotein. *Structure 6*:895–909.

Andres DA, Goldstein JL, Ho YK, Brown MS. 1993. Mutational analysis of alpha-subunit of protein farnesyltransferase. Evidence for a catalytic role. *J Biol Chem 268*:1383–1390.

Armstrong SA, Seabra MC, Sudhof TC, Goldstein JL, Brown MS. 1993. cDNA cloning and expression of the alpha and beta subunits of rat Rab geranylgeranyl transferase. *J Biol Chem 268*:12221–12229.

Boguski MS, Murray AW, Powers S. 1992. Novel repetitive sequence motifs in the alpha and beta subunits of prenyl-protein transferases and homology of the alpha subunit to the MAD2 gene product of yeast. *New Biol 4*:408–411.

Casey PJ, Seabra MC. 1996. Protein prenyltransferases. *J Biol Chem 271*:5289–5292.

Chen MX, McPartlin AE, Brown L, Chen YH, Barker HM, Cohen PT. 1994. A novel human protein serine/threonine phosphatase, which possesses four tetratricopeptide repeat motifs and localizes to the nucleus. *EMBO J 13*:4278–4290.

Chen WJ, Moomaw JF, Overton L, Kost TA, Casey PJ. 1993. High level expression of mammalian protein farnesyltransferase in a baculovirus system. The purified protein contains zinc. *J Biol Chem 268*:9675–9680.

Chothia C, Levitt M, Richardson D. 1981. Helix to helix packing in proteins. *J Mol Biol 145*:215–250.

Crick FHC. 1953. The packing of α-helices: Simple coiled coils. *Acta Crystallogr 6*:689–697.

Das AK, Cohen PW, Barford D. 1998. The structure of the tetratricopeptide repeats of protein phosphatase 5: Implications for TPR-mediated protein–protein interactions. *EMBO J 17*:1192–1199.

Esnouf RM. 1997. An extensively modified version of MolScript that includes greatly enhanced coloring capabilities. *J Mol Graph Model 15*:133–138.

Felsenstein J. 1996. Inferring phylogenies from protein sequences by parsimony, distance, and likelihood methods. *Methods Enzymol 266*:418–427.

Gibbs JB, Oliff A. 1997. The potential of farnesyltransferase inhibitors as cancer chemotherapeutics. *Annu Rev Pharmacol Toxicol 37*:143–166.

Gindhart JG Jr, Goldstein LS. 1996. Tetratrico peptide repeats are present in the kinesin light chain. *Trends Biochem Sci 21*:52–53.

Henikoff S, Henikoff JG. 1992. Amino acid substitution matrices from protein blocks. *Proc Natl Acad Sci USA 89*:10915–10919.

Holm L, Sander C. 1993. Protein structure comparison by alignment of distance matrices. *J Mol Biol 233*:123–138.

Janin J, Miller S, Chothia C. 1988. Surface, subunit interfaces and interior of oligomeric proteins. *J Mol Biol 204*:155–164.

Kajava AV. 1998. Structural diversity of leucine-rich repeat proteins. *J Mol Biol 277*:519–527.

Kajava AV, Vassart G, Wodak SJ. 1995. Modeling of the three-dimensional structure of proteins with the typical leucine-rich repeats. *Structure 3*:867–877.

Kobe B, Deisenhofer J. 1994. The leucine-rich repeat: A versatile binding motif. *Trends Biochem Sci 19*:415–421.

Koonin EV, Tatusof RL, Rudd KE. 1996. *Escherichia coli* function and evolutionary implication of genome scale computer-aided protein sequence analysis. In: Gustafson JP, Flavell RB, eds. *Genomics of plants and animals, 21st Stadler Genetics Symposium.* New York: Plenum Press. pp 177–210.

Kraulis PJ. 1991. MOLSCRIPT: A program to produce both detailed and schematic plots structures. *J Appl Crystallogr 24*:946–950.

Lamb JR, Michaud WA, Sikorski RS, Hieter PA. 1994. Cdc16p, Cdc23p and Cdc27p form a complex essential for mitosis. *EMBO J 13*:4321–4328.

Lamb JR, Tugendreich S, Hieter P. 1995. Tetratrico peptide repeat interactions: To TPR or not to TPR? *Trends Biochem Sci 20*:257–259.

Long SB, Casey PJ, Beese LS. 1998. Cocrystal structure of protein farnesyltransferase complexed with a farnesyl diphosphate substrate. *Biochemistry 37*:9612–9618.

Lubas WA, Frank DW, Krause M, Hanover JA. 1997. O-Linked GlcNAc transferase is a conserved nucleocytoplasmic protein containing tetratricopeptide repeats. *J Biol Chem 272*:9316–9324.

O'Shea EK, Klemm JD, Kim PS, Alber T. 1991. X-ray structure of the GCN4 leucine zipper, a two-stranded, parallel coiled coil. *Science 254*:539–544.

Park HW, Boduluri SR, Moomaw JF, Casey PJ, Beese LS. 1997. Crystal structure of protein farnesyltransferase at 2.25 angstrom resolution. *Science 275*:1800–1804.

Seabra MC. 1998. Membrane association and targeting of prenylated Ras-like GTPases. *Cell Signal 10*:167–172.

Seabra MC, Goldstein JL, Sudhof TC, Brown MS. 1992. Rab geranylgeranyl transferase. A multisubunit enzyme that prenylates GTP-binding proteins terminating in Cys-X-Cys or Cys-Cys. *J Biol Chem 267*:14497–14503.

Seabra MC, Reiss Y, Casey PJ, Brown MS, Goldstein JL. 1991. Protein farnesyltransferase and geranylgeranyltransferase share a common alpha subunit. *Cell 65*:429–434.

Sikorski RS, Boguski MS, Goebl M, Hieter P. 1990. A repeating amino acid motif in CDC23 defines a family of proteins and a new relationship among genes required for mitosis and RNA synthesis. *Cell 60*:307–317.

Smith RL, Redd MJ, Johnson AD. 1995. The tetratricopeptide repeats of Ssn6 interact with the homeo domain of alpha 2. *Genes Dev 9*:2903–2910.

Thompson JD, Higgins DG, Gibson TJ. 1994. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res 22*:4673–4680.

Thunnissen AM, Dijkstra AJ, Kalk KH, Rozeboom HJ, Engel H, Keck W, Dijkstra BW. 1994. Doughnut-shaped structure of a bacterial muramidase revealed by X-ray crystallography. *Nature 367*:750–753.

Tzamarias D, Struhl K. 1995. Distinct TPR motifs of Cyc8 are involved in recruiting the Cyc8–Tup1 corepressor complex to differentially regulated promoters. *Genes Dev 9*:821–831.

Walker DR, Koonin EV. 1997. SEALS: A system for easy analysis of lots of sequences. *ISMB 5*:333–339.

Wiener M, Freymann D, Ghosh P, Stroud RM. 1997. Crystal structure of colicin Ia. *Nature 385*:461–464.

Zhang FL, Casey PJ. 1996. Protein prenylation: Molecular mechanisms and functional consequences. *Annu Rev Biochem 65*:241–269.