Correspondence

NOVEL REPETITIVE SEQUENCE MOTIFS IN THE α AND β SUBUNITS OF PRENYL-PROTEIN TRANSFERASES AND HOMOLOGY OF THE α SUBUNIT TO THE *MAD2* GENE PRODUCT OF YEAST

TO THE EDITOR:

Protein prenylation is an important post-translational modification affecting many diverse aspects of cell biology. Identified targets for prenylation include $p21^{ras}$ proteins, fungal mating factors, the γ subunits of G proteins, and nuclear lamins, and prenylation mediates the targeting to and anchoring of these proteins to cellular membranes (Maltese, 1990; Rine and Kim, 1990; Der and Cox, 1991; Glomset et al., 1991). Prenylation involves thioether linkages of isoprenoid groups to "C-a-a-X"[1] boxes and other carboxy-terminal, cysteine-containing motifs and is catalyzed by heterodimeric enzymes consisting of α and β subunits (Reiss et al., 1991). cDNAs for several β subunits have been cloned, including rat β-farnesyltransferase and its yeast homolog the *RAM1* (*DPR1*) gene product (Goodman et al., 1988; Chen et al., 1991a). The *Saccharomyces cerevisiae* genes *BET2* and *CAL1* (*CDC43*) are also members of this sequence family (Ohya et al., 1991; Rossi et al., 1991). He et al. (1991) have recently cloned an α-subunit gene, *RAM2*, from *S. cerevisiae* and mammalian α-subunit cDNAs have also been identified (Kohl et al., 1991; Chen et al., 1991b). We would like to describe some interesting structural features of α- and β-subunit sequences and their functional implications. We would also like to report that α subunits are evolutionarily related to the *MAD2* gene product which is involved in feedback control of mitosis in budding yeast (Li and Murray, 1991).

Multiple alignment analysis of the yeast RAM2, rat FT-α, and yeast Mad2 sequences revealed significant internal repeats (Boguski et al., 1992; Boguski, in press) which are further characterized here. These repeats have an approximate length of 34 residues and somewhat variable spacing (Fig. 1A). Individual repeat units have diverged in sequence but retain an invariant tryptophan and a number of other well-conserved residues (Fig. 1A). Tryptophan occupies a unique place in protein structure and evolution because it is the least abundant and least mutable amino acid (Dayhoff, 1978) and therefore the repetitive Trp residues are undoubtedly important for the function of FT-α. What might this function be? Tryptophan-containing repeats (otherwise unrelated to FT-α sequences) have been identified in several protein families (Duronio et al., 1992). In *myb* proto-oncogene products, Trp repeats may contribute to the DNA-binding domain by stabilizing a helix–turn–helix motif (Kaneri-Ishii et al., 1990). More pertinent to the present situation, perhaps, are the Trp repeats found in the superfamily of G_β-related sequences (Duronio et al., 1992) and the TPR gene family (Goebl and Yanagida, 1991) where they presumably mediate protein–protein interactions (see below).

In addition to landmark Trp residues, both FT-α and TPR repeats also contain highly conserved proline residues and the spacing between Pro and Trp is the same in both sequence families (Goebl and Yanagida, 1991). Predictions based on the entire set of fifteen, 34-residue FT-α repeats (Fig. 1) suggest a supersecondary structure in which the tetrapeptide Pro-X-Asn-Tyr mediates a reverse turn between α-helix and β-strand components. It should be possible to test aspects of this repetitive structure model using limited proteolysis as Hirano et al. (1990) have done for the TPR family.

Interestingly, FT-β subunit homologs also show an internal repeat but this motif is structurally and evolutionarily distinct from FT-α sequences (Fig. 1B). FT-β repeats have a length of 44–45 residues and are somewhat more divergent with respect to each other than FT-α repeats (Fig. 1A)—a fact which may be related to the functional diversity of β subunits (see below). Nevertheless, several conserved sequence patterns are apparent. These include a centrally located Gly-Gly-Phe-Gly-Gly-X-Pro motif flanked by subdomains containing cysteine, histidine, and cysteine residues at positions 13–15, 33–36, and 35–40, respectively (Fig. 1B). Although structural inferences are less consistent than those for FT-α repeats, the GGFGG.P motif seems to indicate a surface loop or turn. Consistent with this idea is the fact that gaps (representing hypothetical insertion/deletion mutations) are tolerated in this area of the repeat unit (Fig. 1B).

The biochemical and structural properties of prenyltransferase subunits are only beginning to be understood. At a minimum, however, the α subunit must contain a site that binds to β subunits and it may be that Trp residues mediate heterodimerization via aromatic–aromatic interactions (Singh and Thornton, 1985). Specifically, one can postulate hydrophobic bonds between tryptophan side chains and the conserved phenylalanines in FT-β repeats (Fig. 1B). These Phe residues are flanked by glycine residues—a situation reminiscent of the "knob and hole" model for protein–protein interaction (Hirano et al., 1990; Goebl and Yanagida, 1991). Another possibility for the role of Trp (and/or Tyr) residues (Fig. 1A) is that the aromatic side chains contribute to a binding pocket for prenyl groups (Reiss et al., in press). Hydrophobic ligand-binding pockets are important features of many enzymes and transport proteins (Peitsch and Boguski, 1991).

The Cys-His-Cys motif in FT-β repeats suggests a role for metal-binding sites, disulfide bonds, or both in the

[1]This is an abbreviation signifying a tetrapeptide whose sequence is Cysteine-(aliphatic residue)-(aliphatic residue)-(any residue).

THE NEW BIOLOGIST, Vol. 4, No. 4 (April), 1992: pp 408-411

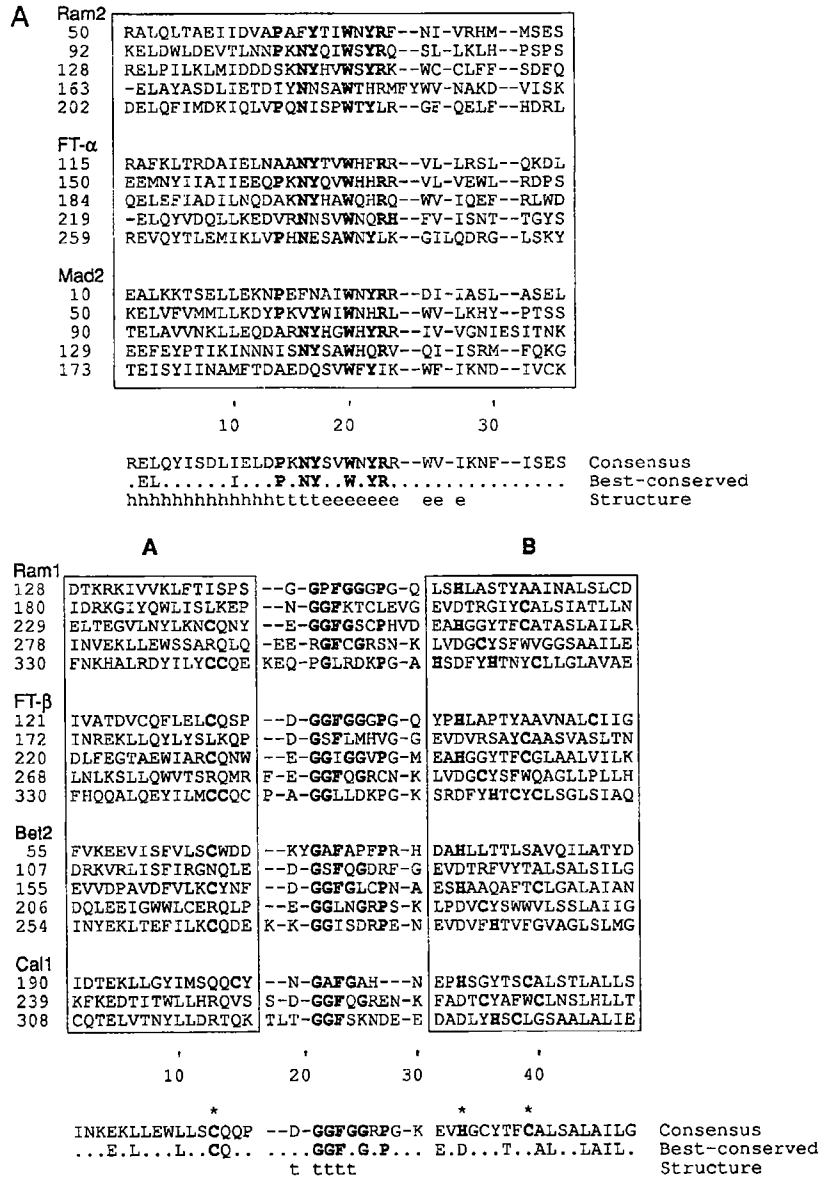

Figure 1. Alignments of internal repeats from prenyltransferase subunits.

For both the α- and β-subunit families, repeat boundaries were estimated from inspection of multiple alignment plots (Fig. 5 in Boguski et al. (1992)), from the positions of gaps in a global alignment of the three sequences (Fig. 6 in Boguski et al. (1992), and from the locations of statistically significant ungapped homology blocks identified by the MACAW program (Schuler et al., 1991; Boguski, in press). Once the repeats were thus defined, they were treated as separate sequences and optimally aligned (Lipman et al., 1989); structure prediction was then carried out using the evolutionary comparison strategy of Crawford et al. (1987). As described in the text, highly significant similarities among Ram2, rat FT-α, and Mad2 were identified by database searching and the complete results are available from M.S.B. on request (email: boguski@ncbi.nlm.nih.gov). Quantitative sequence "profiles" (Gribskov et al., 1990) of the aligned repeats are also available. (A) Aligned internal repeats in FT-α-related sequences. A core block of 24 aligned residues with no gaps constitutes the most highly conserved unit. Although small insertions in three of the 15 repeats disrupt the pattern distal to this, a total of 34–35 residues seems to be the actual length of the repeat unit. Residues numbers at the left refer to the complete sequences of the proteins. The "best-conserved" residues are based on a frequency of 45%. For the secondary structure prediction, "h" = α helix, "t" = turn, and "e" = β strand. The rat (Chen et al., 1991b) and bovine (Kohl et al., 1991) FT-α sequences are 84% identical and thus only the rat sequence is shown. (B) Aligned internal repeats in FT-β-related sequences. The average length of these repeats is 45 residues. Residue numbers, consensus definitions, and structure symbols are as in A. β-subunit repeats appear to be composed of two subdomains (labeled A and B) separated by a "linker" region encompassing the conserved GGFGG.P motif. The "A" subdomain contains the first cysteine residue of the Cys-His-Cys triad with somewhat variably spaced histidine and cysteine residues occurring in the "B" subdomain. Cal1 (Cdc43) is the most distantly related of the β subunits (Boguski, in press) and also the most divergent β subunit with respect to intrasequence repeats (Boguski, in press; Boguski et al., 1992).

THE NEW BIOLOGIST, Vol. 4, No. 4 (April 1992: 408-411)

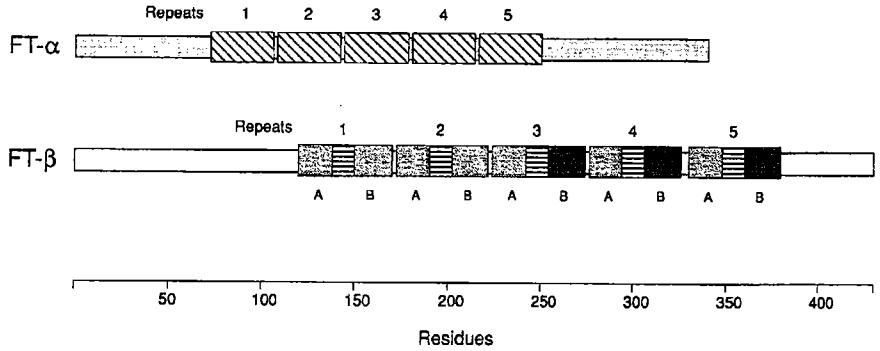

Figure 2. Schematic diagram of internal repeats in prenyltransferase subunits.

The approximate sizes and locations of internal repeats in the FT-α family of sequences are represented by rectangles filled with oblique lines. Individual repeats in the FT-β family are represented by three connected rectangles which illustrate their more complex structure. The "A" and "B" subdomains are depicted as gray rectangles separated by a horizontally lined "linker" region encompassing the conserved GGFGG.P motif (see text). Repeats 3B, 4B, and 5B are shaded more darkly to indicate that they are the best-conserved homology blocks among all four known FT-β-related sequences (Boguski, in press).

structure and function of this protein. Farnesyltransferase activity shows an absolute requirement for Zn^{2+}, probably at the peptide-binding site on the β subunit (Reiss et al., in press). This suggests that one or more of the Cys-His-Cys repeats represent a new type of zinc coordination motif, by analogy with the well-known, repetitive "zinc finger" domains that have been observed in many eukaryotic transcription factors (Klevit, 1991). However, nonrepetitive Cys/His zinc-binding sites have also been described in several enzymes of known crystal structure (Creighton, 1984). For example, liver alcohol dehydrogenase (Brookhaven identification code 6ADH) contains an active-site zinc cation bound by a Cys-His-Cys triad.

The relatively greater degree of sequence divergence among FT-β repeats, compared with FT-α repeats (Boguski et al., 1992), may be one aspect of their functional differentiation with respect to substrate and protein target specificity. In yeast, the same α subunit (Ram2) combines with various β subunits (Ram1, Cal1, Bet2) to form farnesyltransferase or geranylgeranyl transferase type I or type II activities (Kohl et al., 1991). These proteins differ not only in the type of isoprenoid group used, but also in the particular carboxy-terminal peptides they recognize. The identification of another putative α subunit in yeast (see below) suggests additional combinatorial possibilities.

Searches were carried out on the latest releases of the various databases using the BLAST Network Service and, surprisingly, Ram2 and bovine FT-α were found to be related to the *MAD2* gene product with chance probabilities of $< 10^{-8}$ (Altschul et al., 1990). No similarity to other Trp-containing repeat families was detected and thus this implies that Mad2 may be functionally similar to FT-α subunits. FT-α, Ram2, and Mad2 are about 28% identical globally,[2] but more intense localized similarities were identified by the BLAST algorithm (Boguski, in press) and these conserved local homology blocks coincide with the repetitive sequence domains which account for >60% of the total lengths of these proteins (Fig. 2). Thus the repeats themselves are expected to be important for function.

Isoprenoid compounds (farnesyl and geranylgeranyl pyrophosphates) are derived from the cholesterol/mevalonate pathway and indirect evidence for an involvement of these compounds in cell cycle control and DNA replication has been available for some time (Quesney-Huneeus et al., 1979; Habenicht et al., 1980; Maltese, 1990). However, our observation that the *MAD2* gene product is an FT-α homolog represents the first specific indication that protein prenylation may play a role in mitosis, specifically in the feedback controls that link the exit from mitosis to the completion of mitotic spindle assembly.

In summary, we have characterized two novel repetitive sequence motifs in prenyltransferase subunits (Fig. 2). These repeats have distinct structural and functional implications and are unrelated to each other, thus representing an interesting case of two families of repeated sequences that have apparently coevolved to form a heterodimeric enzyme. Furthermore we have shown that the yeast *MAD2* gene product is evolutionarily related to FT-α subunits. Experiments are in progress to test the functional implications of this homology.

Note Added in Proof

We have identified the human homolog of FT-α in the collection of "Expressed Sequence Tags" from a hippocampus cDNA library (M. Adams et al. Nature 355:632-634, 1992). The corresponding clone, EST00211, is available directly from the American Type Culture Collection (12301 Parklawn Drive, Rockville, MD 20852). FT-α and Bet2 homologs are also present in the Caenorhabiditis elegans cDNA library (R. Waterston et al. Nature Genetics, in press). These clones are nos. cm2e12 and cm20c6, respec-

[2]As a point of reference, this is about the same degree of sequence identity as observed among yeast and mammalian GAP "catalytic" domains (Wang et al., 1991).

tively and mapping data indicates that the FT-β homolog (cm20c6) is a member of a dispersed, multigene family present on chromosomes 2, 3 and 4 (acedb - A C. elegans Database, R. Durbin and J. Thierry-Mieg).

Acknowledgment

We thank Drs. Joseph Goldstein and Phil Green for helpful discussions and for providing data prior to publication.

REFERENCES

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990): Basic local alignment search tool. J Mol Biol 215:403-410

Boguski MS (in press): Computational sequence analysis revisited: New databases, software tools and the research opportunities they engender. J Lipid Res

Boguski MS, Hardison RC, Schwartz S, Miller W (1992): Analysis of conserved domains and sequence motifs in cellular regulatory proteins and locus control regions using new software tools for multiple alignment and visualization. New Biol 4:247-260

Chen W-J, Andres DA, Goldstein JL, Russell DW, Brown MS (1991a): cDNA cloning and expression of the peptide-binding β subunit of rat $p21^{ras}$ farnesyltransferase, the counterpart of yeast DPR1/RAM1. Cell 66:327-334

Chen W-J, Andres DA, Goldstein JL, Brown MS (1991b): Cloning and expression of a cDNA encoding the α subunit of rat $p21^{ras}$ protein farnesyltransferase. Proc Natl Acad Sci USA 88:11368-11372

Crawford IP, Niermann T, Kirschner K (1987): Prediction of secondary structure by evolutionary comparison: Application to the alpha subunit of tryptophan synthase. Proteins 2:118-129

Creighton TE (1984): Proteins: Structures and Molecular Properties, New York, Freeman, p 818

Dayhoff MO (1978): Atlas of Protein Sequence and Structure, Washington, DC, National Biomedical Research Foundation

Der CJ, Cox AD (1991): Isoprenoid modification and plasma membrane association: Critical factors for Ras oncogenicity. Cancer Cells 3:331-340

Duronio RJ, Gordon JI, Boguski MS (1992): Comparative analysis of the β transducin family with identification of several new members including PWP1, a nonessential gene of Saccharomyces cerevisiae that is divergently transcribed from NMT1. Proteins Struct Funct Genet 13:41-56

Glomset J, Gelb M, Farnsworth C (1991): The prenylation of proteins. Curr Opinion Lipidol 2:118-124

Goebl M, Yanagida M (1991): The TPR snap helix: A novel protein repeat motif from mitosis to transcription. Trends Biochem Sci 16:173-177

Goodman LE, Perou CM, Fujiyama A, Tamanoi F (1988): Structure and expression of yeast DPR1, a gene essential for the processing and intracellular localization of ras proteins. Yeast 4:271-281

Gribskov M, Luthy R, Eisenberg D (1990): Profile analysis. In Doolittle RF (eds). Methods In Enzymology, San Diego, Academic Press, pp 146-159

Habenicht AJR, Glomset JA, Ross R (1980): Relation of cholesterol and mevalonic acid to the cell cycle in smooth muscle and Swiss 3T3 cells stimulated to divide by platelet-derived growth factor. J Biol Chem 255:5134-5140

He B, Chen P, Chen S-Y, Vancura KL, Michaelis S, Powers S (1991): RAM2, an essential gene of yeast, and RAM1 encode the two polypeptide components of the farnesyltransferase that prenylates a-factor and Ras proteins. Proc Natl Acad Sci USA 88:11373-11377

Hirano T, Kinoshita N, Morikawa K, Yanagida M (1990): Snap helix with knob and hole: Essential repeats in S. pombe nuclear protein nuc2+. Cell 60:319-328

Kaneri-Ishii C, Sarai A, Sawazaki T, Nakagosh H, He D-N, Ogata K, Nishimura Y, Ishii S (1990): The tryptophan cluster: A hypothetical structure of the DNA-binding domain of the myb protooncogene product. J Biol Chem 265:19990-19995

Klevit RE (1991): Recognition of DNA by Cys_2His_2 zinc fingers. Science 253:1367

Kohl NE, Bichl RE, Schaber MD, Rands E, Soderman DD, He B, Moores SL, Pompliano DL, Ferro-Novick S, Powers S, Thomas KA, Gibbs J (1991): Structural homology among mammalian and Saccharomyces cerevisiae isoprenyl-protein transferases. J Biol Chem 266:18884-18888

Li R, Murray AW (1991): Feedback control of mitosis in budding yeast. Cell 66:519-531

Lipman DJ, Altschul SF, Kececioglu JD (1989): A tool for multiple sequence alignment. Proc Natl Acad Sci USA 86:4412-4415

Maltese WA (1990): Posttranslational modification of proteins by isoprenoids in mammalian cells. FASEB J 4:3319-3326

Ohya Y, Goebl M, Goodman LE, Peterson-Bjorn S, Friesen JD, Tamanoi F, Anraku Y (1991): Yeast CAL1 is a structural and functional homologue to the DPR1 (RAM) gene involved in ras processing. J Biol Chem 266:12356-12360

Peitsch MC, Boguski MS (1991): The first lipocalin with enzymatic activity. Trends Biochem Sci 16:363

Quesney-Huneeus V, Wiley MH, Siperstein MD (1979): Essential role for mevalonate synthesis in DNA replication. Proc Natl Acad Sci USA 76:5056-5060

Reiss Y, Seabra MC, Armstrong SA, Slaughter CA, Goldstein JL, Brown MS (1991): Nonidentical subunits of $p21^{H-ras}$ farnesyltransferase. J Biol Chem 266:10672-10677

Reiss Y, Brown MS, Goldstein JL (in press): Divalent cation and prenyl pyrophosphate specificities of the protein farnesyltransferase from rat brain, a zinc metalloenzyme. J Biol Chem

Rine J, Kim S-H (1990): A role for isoprenoid lipids in the localization and function of an oncoprotein. New Biol 2:219-226

Rossi G, Jiang Y, Newman AP, Ferro-Novick S (1991): Dependence of Ypt1 and Sec4 membrane attachment on Bet2. Nature 351:158-161

Schuler GD, Altschul SF, Lipman DJ (1991): A workbench for multiple alignment construction and analysis. Proteins Struct Funct Genet 9:180-190

Singh J, Thornton JM (1985): The interaction between phenylalanine rings in proteins. FEBS Lett 191:1-6

Wang Y, Boguski MS, Riggs M, Rodgers L, Wigler M (1991): Sar1, a gene from Schizosaccharomyces pombe encoding a protein that regulates ras1. Cell Regul 2:453-465

Mark S. Boguski[1,*]
Andrew W. Murray[2]
Scott Powers[3]
[1]National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894.
[2]Departments of Biochemistry and Biophysics and Physiology, University of California, San Francisco, CA 94143.
[3]Department of Biochemistry, University of Medicine and Dentistry, Robert Wood Johnson Medical School, Piscataway, NJ 08854.
December 20,1991; revised February 5, 1992

*To whom correspondence should be addressed.