

Schierbeek, A. J., Swarte, M. B. A., Dijkstra, B. W., Vriend, G., Read, R. J., Hol, W. G. J., & Drenth, J. (1989) *J. Mol. Biol.* 206, 365-380.

Schmitt, B., & Cohen, R. (1980) *Biochem. Biophys. Res. Commun.* 93, 709-712.

Shepherd, G., & Hammes, G. G. (1977) *Biochemistry* 16, 5234-5241.

Spencer, M. E., Darlison, M. G., Lewis, H. M., & Guest, J. R. (1984) *Eur. J. Biochem.* 141, 361-374.

Stephens, P. E., Darlison, M. G., Lewis, H. M., & Guest, J. R. (1983) *Eur. J. Biochem.* 133, 481-489.

Stepp, L. R., & Reed, L. J. (1985) *Biochemistry* 24, 7187-7191.

Stepp, L. R., Bleile, D. M., McRorie, D. K., Pettit, F. H., & Reed, L. J. (1981) *Biochemistry* 20, 4555-4560.

Takenaka, A., Kizawa, K., Hata, T., Sato, S., Misaka, E.-J., Tamura, C., & Sasada, Y. (1988) *J. Biochem. (Tokyo)* 103, 463-469.

Texter, F. L., Radford, S. E., Laue, E. D., Perham, R. N., Miles, J. S., & Guest, J. R. (1988) *Biochemistry* 27, 289-296.

Thekkumkara, T. J., Ho, L., Wexler, I. D., Pons, G., Liu, T.-C., & Patel, M. S. (1988) *FEBS Lett.* 240, 45-48.

Wagenknecht, T., Francis, N., & DeRosier, D. J. (1986) *Biochem. Biophys. Res. Commun.* 135, 802-807.

Wagenknecht, T., Francis, N., DeRosier, D. J., Hainfeld, J. F., & Wall, J. S. (1987) *J. Biol. Chem.* 262, 877-882.

Westphal, A. H., & de Kok, A. (1990) *Eur. J. Biochem.* 187, 235-239.

Wexler, I. D., Hemalatha, S. G., & Patel, M. S. (1991) *FEBS Lett.* 282, 209-213.

Williams, C. H., Jr. (1976) in *The Enzymes* (Boyer, P. D., Ed.) 3rd ed., Vol. 13, pp 89-173, Academic Press, New York.

Wu, T.-L., & Reed, L. J. (1984) *Biochemistry* 23, 221-226.

Yeaman, S. J. (1989) *Biochem. J.* 257, 625-632.

Yang, H., Hainfeld, J. F., Wall, J. S., & Frey, P. A. (1985) *J. Biol. Chem.* 260, 16049-16051.

Yang, Y.-S., & Frey, P. A. (1986) *Biochemistry* 25, 8173-8178.

Yang, Y.-S., & Frey, P. A. (1989) *Arch. Biochem. Biophys.* 268, 465-474.

Accelerated Publications

Exon Organization of the Human FKBP-12 Gene: Correlation with Structural and Functional Protein Domains[‡]

Anthony G. DiLella* and Robert J. Craig

Department of Molecular Genetics, SmithKline Beecham Pharmaceuticals, King of Prussia, Pennsylvania 19406

Received June 14, 1991

ABSTRACT: FKBP-12, the major T-cell binding protein for the immunosuppressive agents FK506 and rapamycin, catalyzes the interconversion of the cis and trans rotamers of the peptidyl-prolyl amide bond of peptide and protein substrates. The function of rotamase activity in cells and the role of FKBP-12 in immunoregulation is uncertain. In this paper we report the cloning and characterization of the human chromosomal FKBP-12 gene and four processed FKBP-12 pseudogenes. The FKBP-12 gene is 24 kilobases in length and contains five exons. The protein-coding region of the gene is divided into four exon modules that correlate with the structural and functional domains of the protein. The novel structure of FKBP-12 resulting from the topology of the antiparallel β -sheet is the topological crossing of two loops that are encoded by separate exons. Separate exons also encode the antiparallel β -sheet and α -helical region that define the drug-binding pocket and enzyme activity site of FKBP-12. The exon organization of the FKBP-12 gene also provided insight into the genetic evolution of the immunophilin family. Knowledge of the FKBP-12 gene structure will enable inactivation of this gene by homologous recombination in cells to provide a model to study the role of FKBP-12 in immunoregulation and normal cellular processes.

Early events in T-lymphocyte activation result from the stimulation of the T-cell receptor signal transmission pathway by a specific antigen (Crabtree, 1989). Immunosuppressive agents that specifically inhibit this pathway have been used to prevent graft rejection of organ and bone marrow transplants (Schreiber, 1991). FK506 and CsA are structurally unrelated immunosuppressive drugs that block the T-cell receptor signal transmission pathway in T-cell activation by inhibiting expression of the same set of lymphokine genes, such

as IL-2 (Tocci et al., 1989; Mattila et al., 1990; Emmel et al., 1989). Although FK506 and CsA each inhibit T-cell activation by a similar mechanism, they bind to distinct and abundant T-cell cytoplasmic receptors, termed immunophilins (Schreiber, 1991). The CsA-binding protein is cyclophilin (Handschoenmacher, 1984), while the major FK506-binding protein in T cells is the 12-kDa cytosolic protein FKBP-12 (Harding et al., 1989; Siekierka et al., 1989; Fretz et al., 1991). Rapamycin, an immunosuppressive agent structurally related to FK506 (Schreiber, 1991), also specifically binds to FKBP-12 (Fretz et al., 1991) but inhibits a later stage of T-cell activation by blocking the lymphokine receptor signal trans-

[‡]The nucleotide sequence in this paper has been submitted to GenBank under Accession Number J05340.

mission pathway (Tocci et al., 1989; Dumont et al., 1990). FKBP-12 and cyclophilin are also distributed in nonlymphoid cells and are phylogenetically conserved (Siekierka et al., 1990; Koletsy et al., 1986), suggesting a fundamental biochemical role for these proteins. Both proteins are peptidyl-prolyl isomerases that catalyze the interconversion of the cis and trans rotamers of the peptidyl-prolyl amide bond of peptide and protein substrates; this rotamase activity is specifically inhibited by their respective immunosuppressive ligands (Harding et al., 1989; Siekierka et al., 1989; Fischer et al., 1989; Takahashi et al., 1989). However, recent evidence suggests that inhibition of rotamase activity is an insufficient requirement for mediating the immunosuppressive actions of these drugs on T-cell activation (Bierer et al., 1990a,b). It is speculated that the immunophilin-drug complex functions as the biological effector; this complex may interact with distinct molecules in the signaling pathways of T-cell activation. The biological significance of rotamase activity within cells and the precise role of immunophilins in immunoregulation is uncertain. Also, the mechanisms by which FK506 and rapamycin inhibit the activity of a common receptor, FKBP-12, yet inhibit different T-cell activation signaling pathways remain unclear. The understanding of these events is further complicated by the presence of several low-abundance T-lymphocyte immunophilins that also have affinity for FK506 and rapamycin (Fretz et al., 1991). One of these proteins has recently been shown to share significant homology with FKBP-12 (Jin et al., 1991).

Inactivation of the FKBP-12 gene by homologous recombination in T cells will help us understand the role of FKBP-12 as mediator of the biological actions of FK506 and rapamycin. These studies require prior knowledge of the FKBP-12 gene structure. We report the characterization of the human chromosomal gene encoding FKBP-12. We used a FKBP-12 cDNA probe (Standaert et al., 1990) to screen a genomic library to characterize the productive gene as well as four processed FKBP-12 pseudogenes. The human FKBP-12 gene is 24 kilobases (kb) in length and contains five exons. An unusual feature of this gene is the presence of an intron in the 3'-untranslated region. The remaining introns split the gene into exon modules that correlate with the drug-binding and novel loop-crossing domains of the FKBP-12 protein (Michnick et al., 1991; Van Duyne et al., 1991; Moore et al., 1991; Rosen et al., 1991). The exon organization of the FKBP-12 gene also provides insight into the genetic evolution of the immunophilin family.

MATERIALS AND METHODS

Southern Analysis. Restriction enzyme digests of human genomic DNA were electrophoresed on agarose gels and transferred to nitrocellulose filters. ^{32}P -labeled FKBP-12 cDNA was hybridized to DNA transfers as described (DiLella et al., 1990). After hybridization, the transfer membranes were washed at 68 °C in 0.5× SSC (1× SSC = 0.15 M NaCl and 15 mM sodium citrate, pH 7).

Cloning of the Human FKBP-12 Gene. A human genomic DNA library constructed in bacteriophage EMBL-3 was purchased from Clontech Laboratory. Screening of the library with ^{32}P -labeled FKBP-12 cDNA and restriction enzyme mapping of clones were carried out as described (DiLella et al., 1990).

Nucleotide Sequencing. Restriction fragments that hybridized to the FKBP cDNA probe were subcloned from the bacteriophage clones. Sequence data were generated from subclones by the dideoxy chain-termination method (Sanger et al., 1977) using synthetic oligonucleotide primers corre-

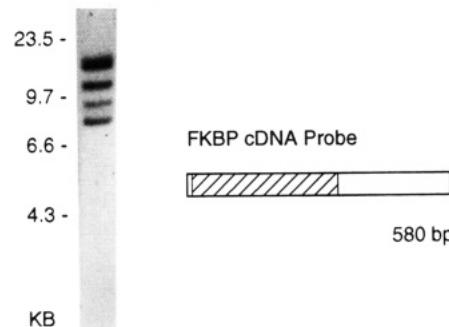


FIGURE 1: Southern hybridization analysis of genomic DNA. Genomic DNA was digested with *Eco*RI and hybridized to the FKBP-12 cDNA probe under stringent conditions. The hatched and open boxes in FKBP-12 cDNA (Standaert et al., 1990) represent the coding and untranslated regions, respectively.

sponding to various regions of FKBP-12 cDNA and derived DNA sequences.

Nucleotide Sequence Analysis. The University of Wisconsin GCG programs (Devereux et al., 1984) were used to search for homologous sequences and align the sequences of the FKBP-12 gene with related pseudogenes.

RESULTS

Figure 1 shows a Southern blot hybridization analysis of *Eco*RI-digested human genomic DNA using a 580-bp FKBP-12 cDNA probe. The hybridization pattern may be explained either by the presence of multiple exons or by the presence of related genes or pseudogenes in the human genome. In contrast, Southern analysis carried out with a cyclophilin cDNA probe demonstrated more than 20 copies of genomic sequences homologous to cyclophilin (A. G. DiLella, unpublished results; Maki et al., 1990). The single-copy gene encoding the major T-cell form of cyclophilin, numerous processed cyclophilin pseudogenes, and genes encoding cyclophilin isoforms contribute to the genomic complexity of cyclophilin (Haendler & Hofer, 1990; Price et al., 1991; Levy et al., 1991). To determine the genomic complexity of FKBP-12, we screened a human genomic DNA library under low stringency using the FKBP-12 cDNA probe. Three overlapping recombinant clones spanning 34 kb were first analyzed (Figure 2A). Restriction enzyme mapping and hybridization of clones to the cDNA probe established the 5'-3' orientation of the clones.

To establish the detailed structure of the FKBP-12 gene, we focused our analysis on two hybridizing fragments of the gene (Figure 2A). Fragment I (derived from λ 211) and fragment II (derived from λ 910) were subcloned for sequence analysis. The exon-intron boundaries were defined by sequencing the appropriate regions of these subcloned genomic DNA fragments (Figure 2B) and aligning them with an FKBP-12 cDNA sequence that closely approximates the size of the corresponding mRNA detected in T cells (Maki et al., 1990). On the basis of comparison with the cDNA, the FKBP-12 chromosomal gene contained the 69 nucleotides of the 5'-untranslated region, 324 nucleotides of the protein-coding region, and the 1120-nucleotide 3'-untranslated region. The only difference we observed between the FKBP-12 chromosomal gene and cDNA sequences were G to A and T to C nucleotide polymorphisms in the 3'-untranslated region, corresponding to cDNA nucleotide positions 407 and 573, respectively. The detailed structure of the FKBP-12 gene is shown at the bottom of Figure 2A. The gene contained five exons and four introns that span about 24 kb. Clones λ 211 and λ 228 overlapped within intron 2. Restriction enzyme analysis of fragments derived from the overlapping regions of

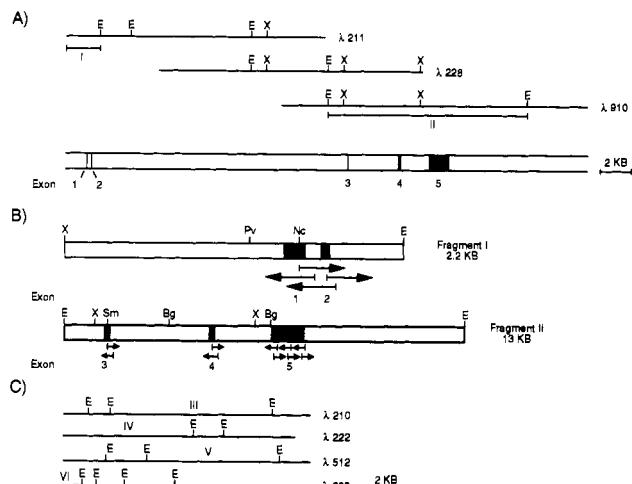


FIGURE 2: Physical map of the human FKBP-12 gene and related pseudogenes. (A) The map of three overlapping genomic DNA inserts from λ clones that were used to characterize the gene are shown. The fragments used for DNA sequence analysis are labelled I and II. The detailed structure of the FKBP-12 gene in the 5' to 3' orientation is shown (bottom). Exons and introns are denoted by solid and open boxes, respectively. (B) Sequencing strategy used to determine the exon-intron junctions of the gene. Fragments I and II were subcloned for sequence analysis. The arrows indicate the direction of sequencing for each fragment with exon- and intron-specific synthetic oligonucleotide primers. (C) The maps of four genomic inserts derived from λ clones that were used to characterize the FKBP-12 pseudogenes are shown. The fragments used for DNA sequence analysis are labelled III–VI. The precise location of the pseudogenes within these fragments were not determined. Restriction enzyme abbreviations: E, EcoRI; X, XbaI; Pv, PvuII; Nc, NcoI; Sm, SmaI; Bg, BglII.

these clones established the extent of overlap and ruled out the possibility of any gaps between these clones (data not shown).

Introns 1–3 divide the protein-coding portion of the FKBP-12 gene (Figure 3). Exon 1 (109 bp) contained the entire 5'-untranslated region reported for the cDNA. The 3'-untranslated region is interrupted by intron 4, such that exon 4 contains 126 bp of protein-coding sequence, the termination codon, and 36 bp of 3'-untranslated sequences. The most 3' exon (exon 5, Figure 3) is 1085 bp and contains the remaining portion of the 3'-untranslated region. The polyadenylation site (Figure 3, arrow) was inferred from the FKBP-12 cDNA clone and is preceded 13 bases by the common polyadenylation signal AATAAA. Figure 3 also shows the limited sequences at the 5' and 3' boundaries of each intron. The exon-intron boundaries were consistent with established consensus splice sites (Mount, 1982). In addition, we observed two clusters of exons containing the peptide-coding region of the gene: (1) exons 1 and 2 are contained in 257 bp at the 5' end and (2) exons 3 and 4 are contained in 3.3 kb at the 3' end. The two exon clusters are separated by a large intron of 16 kb.

Four additional genomic clones isolated from the same library were found to be distinct on the basis of restriction enzyme mapping (Figure 2C). EcoRI fragments III, IV, V, and VI that hybridized to the FKBP-12 cDNA probe were derived from λ 210, λ 222, λ 512, and λ 223, respectively, and were subcloned for sequence analysis. Figure 4 shows a 5'- and 3'-end sequence comparison of these clones with the FKBP-12 gene. The clones contained sequences that were highly homologous to the FKBP-12 cDNA but lacked the introns characteristic of the FKBP-12 gene. Comparison of the FKBP-12 open reading frame with the corresponding sequences of the related genes demonstrated that λ 210, λ 222, λ 223, and λ 512 contained altered reading frames and premature stop codons due to insertions, deletions, and point

Exon (bp)	Intron (kb)	Exon	Intron	Exon
1 (109)	1 (0.1)	GAC G Gly 13	<u>gt</u> gagt.....gtctcc <u>tcag</u>	GG CGC Gly 13
2 (48)	2 (1.6)	ACC G Gly 29	<u>gt</u> gagt <u>cg</u> gg.....c <u>ac</u> ac	GG ATG Gly 29
3 (113)	3 (3)	GCC CAG Gln 66	<u>gt</u> at <u>gt</u> ct <u>tc</u>ct <u>ttt</u> ca <u>ca</u> g	ATG AGT Met 67
4 (165)	4 (1.8)	TTCTTG 435	<u>gt</u> a <u>agg</u> aa <u>at</u>t <u>cccc</u> aa <u>ca</u> g	ATCTGCC 436
5 (1085)			ATGGAGG..... <u>AATAAA</u> AGTGC <u>TTT</u> T <u>ATG</u> CCGG <u>TTT</u> T <u>CT</u> -3' 449 1501	

FIGURE 3: Exon-intron junctions of the FKBP-12 gene. Nucleotide sequences of the exons and exon-intron junctions were determined from genomic subclones. Intron junctions were positioned by applying the gt-ag splice rule (Mount, 1982). The numbers shown at the splice junctions denote the positions of the corresponding amino acid codons interrupted by introns 1–3 and 3' untranslated nucleotides interrupted by intron 4, as deduced from the cDNA sequence. The polyadenylation signal is underlined and the polyadenylation site is indicated by an arrow.

mutations (Figure 4A). Such modifications preclude the coding of a functional FKBP-12 protein. In addition, the homology between the FKBP-12 gene and related λ 210, λ 222, and λ 512 clones ceased near the points corresponding to the 5' and 3' ends of the FKBP-12 cDNA, and these boundaries were flanked by direct repeats in the related genes (Figure 4A,B). Sequences in the corresponding regions of λ 223 were not determined. The data demonstrated that λ 210, λ 222, λ 223, and λ 512 contained processed FKBP-12 pseudogenes (Vanin, 1984; Wagner, 1986).

The solution structure of human FKBP-12 (Michnick et al., 1991) and the crystal structure of the human FKBP-12/FK506 complex (Van Duyne et al., 1991) were recently determined. Figure 5 shows the α -carbon backbone of the FKBP-12 protein and indicates the intron positions in relation to the protein domains. The structure of FKBP-12 is characterized by a large amphiphilic β -sheet composed of five antiparallel strands with +3, +1, -3, +1 loop connectivity (Michnick et al., 1991). β -strands 1–5 of the sheet are composed of amino acid residues 1–8, 21–30, 47–50, 72–78, and 97–108, respectively. An α -helix composed of residues 60–66 is connected to the fifth and second β -strands of the sheet by two loops. The intron positions marked turns in the FKBP-12 structure that defined domain boundaries of the protein, each containing one or two β -strand and α -helical elements. The exons encoding these elements were observed to be as follows: exon 1 encodes β -strand 1, exon 2 encodes β -strand 4, exon 3 encodes β -strand 5 and the α -helix, and exon 4 encodes β -strands 2 and 3. A unique feature of FKBP-12 resulting from the topology of the β -sheet is a topological crossing of the loops Ser⁹–Gly²⁰ and Leu⁵¹–Gln⁷⁰ (Michnick et al., 1991), corresponding to the positions of introns 1 and 3, respectively. In addition, the β -sheet of FKBP-12 has a right-handed twist that wraps around the α -helical region to form a shallow hydrophobic core where the rotamase activity and drug-binding site reside (Van Duyne et al., 1991). The α -helical domain that serves as the FK506-binding platform is encoded by exon 3. Residues 37–43 that form a positive-charged loop at the open end of the binding pocket are also contained in this domain and are thought to be involved in the biological activity of FKBP-12. Hence, distinct exons encode amino acids that specify the function and unique conformation of FKBP-12.

DISCUSSION

We determined the structural organization of the human chromosomal FKBP-12 gene by characterizing λ clones con-

A) 5'-Analysis

1 50
 210 cactgcactc cagccctgggt gacaaagaga ggctctgtat caaaaaataa
 222 gctaaqaatt tactccctagt gcaggacggg gcgtgcgggg gtcccatgt
 512 aagtctatta caaccccaac aagactcaga gaggaggcaa aattataatgt
 51 100
 210 a-aaat-aat -aaa---a- -gt----c- ---t-----
 222 ---ag-tga tg-ac-a-ca gga-gag- -----tg--- ---a-t-a-
 512 -gttaa-a-c cg---aaag- g---t---a- -a-t-----t-
 FKBP GTGCCGAGC ACGCCGAG GTAGTAG.CA GAGCCGTGGA ACCGCCGCCA
 101 150
 210 -----a-----a-----ta-----a-c-----a-
 222 ---c-c-c a-----a-t-----t-----c-----a-
 223 -----a-----t-----a-t-----t-----c-----a-
 512 -----g a-----c-----t-----a-----a-----c-----c-
 FKBP GGTGCGCTGTT GGTCCACGCC GCCCGTCGCG CGGCCGCC GCTCAGCGC
 151 200
 210 g-----g-----g-----g-a-t
 222 -----at-t c-----c-a -----g-t -----tt---
 223 -a-t-----t-----g-----a-----a-----t-
 512 a-----t-----g-----a-----a-----c-----c-
 FKBP CGCCGCCGCC ATGGGACTGC AGGTGAAAC CATCTCCCCA GGAGACGGC
 201 250
 210 -----g-g-a-----
 222 at-----gg-----
 223 g-----t-----a-----g-----t-----t-----t-
 512 a-----c-----t-----g-----a-----a-----c-----c-
 FKBP GCACCTTCCC CAAGCGCGC CAGACCTGCG TGGTGCATA CACCGGGATG
 251 300
 210 -----c-----
 222 -----t-----c-----t-----aa-----tg-----
 223 -----g-----a-----c-----tc-----t-----tt-----
 512 -----t-----t-----a-----t-----t-----t-----
 FKBP CTTGAAGATG GAAAGAAATT TGATTCCCTCC CGGGACAGAA ACAAGCCCTT
 301 350
 210 -----t-----t-----a-----
 222 -----g-g-----a-----c-----g-ag-----g-t-----
 223 -----t-----a-----c-----c-----a-----
 512 -----t-----t-----c-----c-----a-----
 FKBP TAAGTTATG CTAGGCAAGC AGGAGGTGAT CCGAGGCTGG GAAGAAGGG
 351 400
 210 -----tt-----
 222 -----a-----
 223 -----t-----a-----
 512 -----t-----c-----tg-----accggccat-----ttagt-----c-----
 FKBP TTGCCAGAT GAGTGTGG... GTCAAGAG CCAAACGTAC
 401 450
 210 -----c-----c-----
 222 -g-----t-----c-----tca-----ag-----
 223 a-----c-----c-----a-----
 512 -----t-----c-----c-----a-----
 FKBP TATATCTCCA GATTATGCC ATGGGCCAC TGGGCACCCA GGCATCATCC
 451 500
 210 -----c-----c-----
 222 -----t-----a-----ca-----g-----g-----g
 223 -----t-----a-----
 512 -----c-----a-----
 FKBP CACCAACATGCC CACTCTCGTC TTGATGTGG AGCTTCTAAA ACTGGAATGA

B) 3'-Analysis

1 50
 210 ---c-----a-----a-----a-----c-----a-----
 222 -----a-----a-----a-----t-----c-----a-----
 512 -----a-----t-----t-----a-----gac-t-----a-----
 FKBP TAATTGACAG TTTCATTGAG AGGTGCTGTT TGTAGACTTA A.CACCCAG
 51 100
 210 -----tg-cg- g-----cg-gc- c-----cg-c-----a-----a-
 222 -----t-----t-----t-----t-----c-----c-----a-----
 512 -----g-----t-----t-----t-----c-----c-----a-----
 FKBP GAAAGCCAG CCAT.CATG ACAAACTCTT GAATGTTCTC TAAAGAAAT
 101 150
 210 tt-----aa gc-aag-tgg gt-----a-----t-----g-----gg-----c-----gtcgt-act
 222 -----a-----a-----c-----c-----ca-----a-----
 512 a-----t-----t-----a-----t-----g-----g-----c-----c-
 FKBP GATGCTGGTC ATCGCAGCTT CAGCATCTCC TGTTTTGAG TGCTTGCTC
 151 200
 210 agc-----a-c-----a-----tc-----c-----a-----aaaat-----a-----aaaat-----tag
 222 t-----a-----a-----c-----t-----t-----t-----g-----
 512 t-----a-----t-----a-----t-----a-----g-----g-----c-----c-
 FKBP CCTCTGCTGA TCTCA.. GTT TCCTGGCTT TCCTCCCTCA GCCCCCTCTC
 201 250
 210 ccggg-a-ga t-----g-----a-----ct-----a-----ccc-----agctactcg-----ga-----gctga-----g
 222 -----t-----a-----a-----ac-----c-----g-----a-----t-----t-----
 512 -----c-----t-----a-----c-----g-----a-----t-----a-----c-----c-
 FKBP A.CCCCTTGTG CTGTCCTGTT TAGTGTGTTT GTGAGA.. AA TCGTTGCTG.
 251 300
 210 --gaa-----g-----t-----tg-----a-----tggg-----ggca-----gt-----gtg-----ggcc-----g-----gat-----
 222 -----c-----t-----c-----a-----a-----t-----c-----a-----c-----a-----
 512 -----t-----a-----c-----c-----a-----g-----t-----t-----g-----g-----c-----c-
 FKBP CACCCCTTCCC CCAGCACCAT TTATGAGTCT CAAGTTTAT TATTGCAATA
 301 350
 210 -----t-----c-----a-----a-----aaa-----aaa-----a-----t-----
 222 -----t-----a-----c-----caa-----caa-----a-----cc-----g-----g-----g-----g-----c-----c-
 512 -----t-----t-----a-----c-----a-----c-----a-----t-----t-----a-----
 FKBP AAAAGTGTCTT ATGCCGGCTT TTCTCAGCTC TGTGTCA... .TGGTGGTTA
 351 400
 210 a-----c-----t-----aaa-----a-----a-----a-----a-----t-----
 222 a-----c-----t-----a-----c-----gataaa-----aaac-----aa-----aca-----a-----t-----
 512 a-----a-----c-----a-----g-----a-----t-----a-----c-----t-----c-----c-
 FKBP TTTTCAGGTG CCTCCCTGTC CATTGGTGC AGGGGGATGG GGTTGGCAG

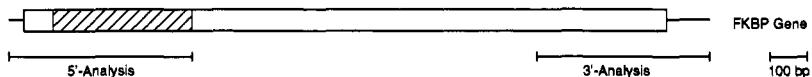


FIGURE 4: Sequence comparison of the FKBP-12 gene with four processed pseudogenes. The regions of the FKBP-12 gene used for 5'-analysis (A) and 3'-analysis (B) are indicated by brackets at the bottom. The hashed and open boxes in the FKBP-12 gene represent the coding and untranslated cDNA regions, respectively; the solid lines represent the 5' and 3' genomic sequences derived from the flanking regions of the FKBP-12 gene. The sequences representing the FKBP-12 gene are indicated in full and are compared to the corresponding pseudogene sequences derived from λ 210, λ 222, λ 223, and λ 512. Conserved nucleotides are indicated with (-) and gaps are marked (...). The complete open reading frame is underlined in the FKBP-12 gene sequence and the start and stop codons are double underlined in (A). Arrows indicate the boundaries of the FKBP-12 cDNA sequence. The direct repeats flanking the pseudogenes are also underlined.

taining 34 kb of contiguous genomic DNA from the corresponding genetic locus. In addition, four processed FKBP-12 pseudogenes were identified. It appeared that the restriction fragments detected by Southern blotting of genomic DNA using the FKBP-12 cDNA as probe are represented in the λ clones, suggesting that there is a single-copy gene coding for FKBP-12. Although FKBP-12 is suggested to be the common immunophilin responsible for buffering the actions of FK506 and rapamycin in T cells, several lower abundance human T-cell immunophilins that have affinity for FK506 and rapamycin exist (Fretz et al., 1991). One of these immunophilins, FKBP-13, is a 13-kDa protein that has 51% nucleotide sequence identity and 43% amino acid sequence identity to FKBP-12 (Jin et al., 1991), suggesting that these two proteins

evolved from a common ancestral gene. Virtually all of the homology between these proteins resides within the C-terminal region, corresponding to residues 26–108 in FKBP-12. It has been reported that this region of FKBP-12 is involved in drug binding and rotamase activity (Van Duyne et al., 1991) and is encoded by exons 3 and 4. It is interesting that no significant homology exists between the two proteins in the region corresponding to the N-terminal 26 amino acids of FKBP-12. For FKBP-12, this nonhomologous region is encoded by exons 1 and 2, which are clustered at the 5' end of the gene (Figure 2A). There is evidence that introns function in gene assembly (Dorit et al., 1990). It is possible that the N-terminal regions of FKBP-12 and FKBP-13 may have arisen by recruitment of unrelated exons carrying functional or structural domains.

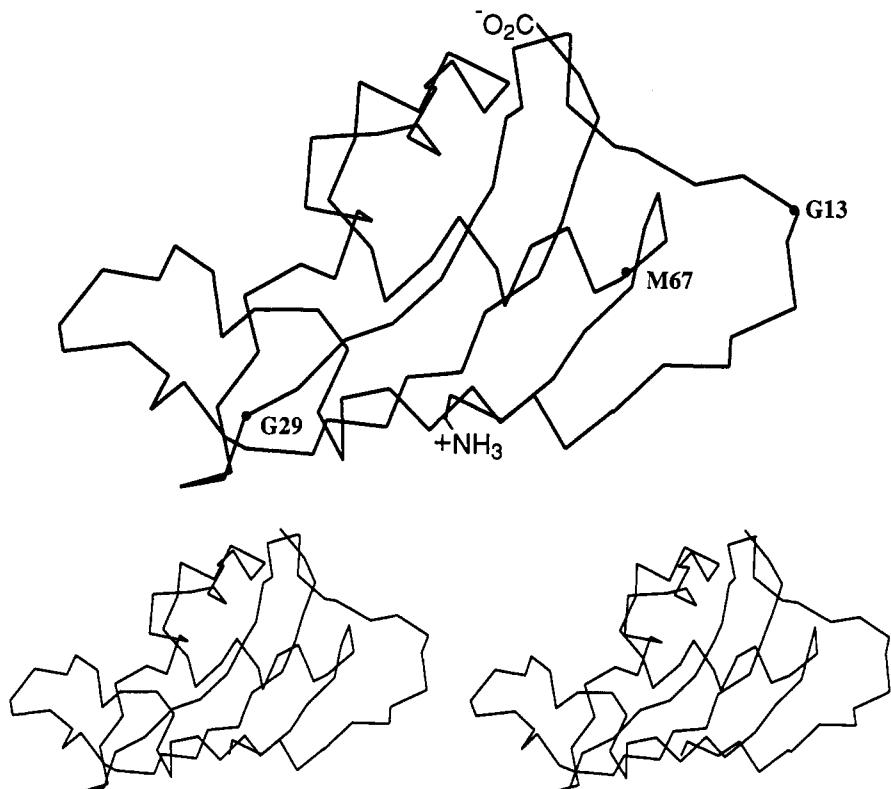


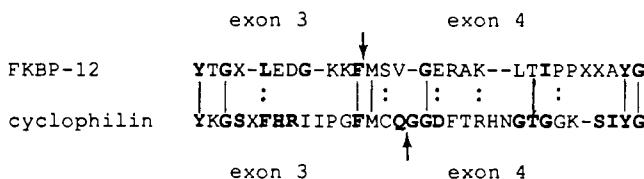
FIGURE 5: FKBP-12 protein structure showing the location of introns that interrupt the corresponding coding sequences in the gene. The α -carbon trace of the FKBP-12 structure, as previously determined (Michnick et al., 1991; Moore et al., 1991), is shown. Numbered residues indicate the intron positions relative to the protein structure.

The large 16-kb intron detected in the region of the FKBP-12 gene separating exons 1 and 2 from the downstream coding exons could be a vestige of genetic reorganization, which would also involve the recruitment of signals that might be involved in the transcriptional regulation of these genes. The observation that an FKBP-12-like protein in the bacterium *Neisseria meningitidis* also lacks significant homology with the N-terminal region of human FKBP-12 (Standaert et al., 1990) lends support to this hypothesis. Future comparison of the exon organization of FKBP-13 and other human FK506-binding immunophilins with the FKBP-12 gene will provide insight into the evolution and structure/function relationships of this protein family.

It is viewed that genes are assembled by exon building blocks that code for functional domains, folding regions, or structural elements within proteins (Dorit et al., 1990). However, this does not seem to apply to all cases (Young & Sylvester, 1989). To determine the case for the FKBP-12 gene, we compared the boundaries of each coding exon in the gene to the domains of the protein. Correlations between the structural and functional domains of the FKBP-12 protein and the location of introns were striking. FKBP-12 structure and function seems to be assembled from discrete exon-coded peptides; intron positions are not random. The structure of FKBP-12 reveals a unique topological feature that results in the crossing of two loops that connect four strands of a five-stranded antiparallel β -sheet (Michnick et al., 1991). Loop crossings were thought to be prevented in antiparallel β -sheets due to difficulties in efficient side-chain packing and were previously unobserved (Michnick et al., 1991). For FKBP-12, the unique loop crossing feature is stabilized by a number of backbone-backbone and backbone-side chain hydrogen bonds and van der Waals contacts (Michnick et al., 1991). It is interesting that one loop is encoded by exons 1 and 2 at the 5' end of the gene and the other loop is encoded by exons 3 and 4 at the

3' end, suggesting that intron 2 mediated an exon shuffling event to yield this unique topology. Furthermore, exons 1 and 2 encode the N-terminal region that is divergent in FKBP-12 and FKBP-13. It will be interesting to determine whether exons in other FKBP-12-related genes encode amino acid residues that form and help stabilize this unique topology.

FKBP-12 and cyclophilin are functionally related proteins that possess rotamase activity and bind drugs that prevent T-cell activation by inhibiting similar signaling pathways. FKBP-12 is distinct from cyclophilin on the basis of differences in immunosuppressant binding specificity, size, and amino acid sequence. However, a region of similarity between FKBP-12 and cyclophilin was recently suggested. Wiederrecht et al. (1991) showed that FKBP-12 reveals two regions (Tyr²⁷-Phe³⁷ and Met⁶⁷-Gly⁸⁴) that, when combined, are conserved with respect to a single region (Tyr⁴⁸-Gly⁸⁰) in cyclophilin. The respective regions are highly conserved in divergent species. Consensus amino acid sequences pertaining to these regions were generated from homologues of FKBP-12 and cyclophilin and were aligned by Wiederrecht et al. (1991) as follows:



Residues in boldface type represent absolute identity among the homologues of FKBP-12 and cyclophilin, X represents no conservation, dashes represent gaps introduced to align the sequences, vertical bars mark identity between residues, and colons represent a conserved change between the sequences. The exon organization of the human FKBP-12 (Figure 2) and cyclophilin (Haendler & Hofer, 1990) genes correlate with this conservation. The arrows in the sequence alignment in-

dicate intron positions relative to FKBP-12 and cyclophilin proteins. These introns interrupt exons 3 and 4 for both the FKBP-12 and cyclophilin genes at approximately the same position in the corresponding protein sequences. The three residue length variation at the positions of the two introns may result from splice junction sliding (Craik et al., 1983). The fact that there is not only a conserved region within FKBP-12 and cyclophilin but also an apparently conserved exon-intron boundary in the genes coding for these regions suggests that these domains evolved from a common ancestral protein and may play a role in catalysis.

The natural ligands of the immunophilins are not known; hence, the biological significance of rotamase activity within cells and the precise role of these proteins in immunoregulation is uncertain. Having established the molecular structure of the FKBP-12 gene, these important issues can now be addressed by inactivation of the FKBP-12 gene by homologous recombination in cells. The Jurkat T-cell line provides a model system for this study since requirements for lymphokine expression appear to be similar to those in primary T cells (Weiss & Imboden, 1987). Such studies are in progress.

ACKNOWLEDGMENTS

We thank Dr. S. Schreiber for the FKBP cDNA and the α -carbon trace of FKBP-12 shown in Figure 5. We also thank Dr. P. Young and Dr. S. Schreiber for helpful discussions on the FKBP protein structure and Dr. G. Sathe and F. Watson for synthesis of oligonucleotide sequencing primers.

REFERENCES

Bierer, B. E., Somers, P. K., & Wandless, T. J. (1990a) *Science* 250, 556-559.

Bierer, B. E., Mattila, P. S., & Standaert, R. F. (1990b) *Proc. Natl. Acad. Sci. U.S.A.* 87, 9231-9235.

Crabtree, G. R. (1989) *Science* 243, 355-361.

Craik, C. S., Rutter, W. J., & Fletterick, R. (1983) *Science* 220, 1125.

Devereux, J., Haeberli, P., & Smithies, O. (1984) *Nucleic Acids Res.* 12, 387-395.

DiLella, A. G., Page, D. C., & Smith, R. G. (1990) *New Biol.* 2, 49-56.

Dorit, R. L., Schoenbach, L., & Gilbert, W. (1990) *Science* 250, 1377-1382.

Dumont, F. J., Staruch, M. J., & Koprak, S. L. (1990) *J. Immunol.* 144, 251-258.

Emmel, E. A., Verweig, C. L., & Durand, D. B. (1989) *Science* 246, 1617-1620.

Fischer, G., Wittmann-Liebold, B., & Lang, K. (1989) *Nature* 337, 476-478.

Fretz, H., Albers, M. W., & Galat, A. (1991) *J. Am. Chem. Soc.* 113, 1409-1411.

Haendler, B., & Hofer, E. (1990) *Eur. J. Biochem.* 190, 447-482.

Handschumacher, R. E., Harding, M. W., & Rice, J. (1984) *Science* 226, 544-547.

Harding, M. W., Galat, A., & Uehling, D. E. (1989) *Nature* 341, 758-760.

Jin, Y. T., Albers, W., & Bierer, B. E. (1991) *Proc. Natl. Acad. Sci. U.S.A.* 88, 6677-6681.

Koletsky, A., Harding, M., & Handschumacher, R. (1989) *J. Immunol.* 137, 1054-1059.

Levy, M. A., Brandt, M., & Livi, G. P. (1991) *Transplant. Proc.* 23, 319-322.

Maki, N., Sekiguchi, F., & Nishimaki, J. (1990) *Proc. Natl. Acad. Sci. U.S.A.* 87, 5440-5443.

Mattila, P. S., Ullman, K. S., & Fiering, S. (1990) *EMBO J.* 9, 4425-4433.

Michnick, S. W., Rosen, M. K., & Wandless, T. J. (1991) *Science* 252, 836-839.

Moore, J. M., Peattie, D. A., & Fitzgibbon, M. J. (1991) *Nature* 351, 248-250.

Mount, S. M. (1982) *Nucleic Acids Res.* 10, 459.

Price, E. R., Zydowsky, L. D., & Jin, M. (1991) *Proc. Natl. Acad. Sci. U.S.A.* 88, 1903-1907.

Rosen, M., Michnick, S. W., & Karplus, M. (1991) *Biochemistry* 30, 4774-4789.

Sanger, F., Nicklen, S., & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A.* 74, 5463.

Schreiber, S. L. (1991) *Science* 251, 283-287.

Siekierka, J. J., Hung, S. H. Y., & Poe, M. (1989) *Nature* 341, 755-757.

Siekierka, J. J., Weiderrecht, G., & Greulich, H. (1990) *J. Biol. Chem.* 265, 21011-21015.

Standaert, R. F., Galet, A., & Verdine, G. L. (1990) *Nature* 346, 671-674.

Takahashi, N., Hayano, T., & Suzuki, M. (1989) *Nature* 337, 473-475.

Tocci, M. J., Matkovich, D. A., & Collier, K. A. (1989) *J. Immunol.* 143, 718-729.

Van Duyne, G. D., Standaert, R. F., & Karplus, P. A. (1991) *Science* 252, 839-842.

Vanin, E. F. (1984) *Biochim. Biophys. Acta* 782, 231-241.

Wagner, M. (1986) *Trends Genet.* 2, 134-137.

Weiss, A., & Imboden, J. B. (1988) *Adv. Immunol.* 41, 1-38.

Wiederrecht, G., Brizuela, L., & Elliston, K. (1991) *Proc. Natl. Acad. Sci. U.S.A.* 88, 1029-1033.

Young, P. R., & Sylvester, D. (1989) *Protein Eng.* 2, 545-551.