The biosynthetic gene cluster for the polyketide immunosuppressant rapamycin

(*Streptomyces*/peptide synthetase/polyketide synthase/FK506)

TORSTEN SCHWECKE*, JESÚS F. APARICIO*, ISTVÁN MOLNÁR*, ARIANE KÖNIG*, LAKE EE KHAW*,
STEPHEN F. HAYDOCK*, MARKIYAN OLIYNYK*†, PATRICK CAFFREY*, JESÚS CORTÉS*,
JOHN B. LESTER*, GÜNTER A. BÖHM‡, JAMES STAUNTON‡, AND PETER F. LEADLAY*§

*Department of Biochemistry and Cambridge Centre for Molecular Recognition, University of Cambridge, Tennis Court Road, Cambridge CB2 1QW, United Kingdom; and ‡University Chemical Laboratory and Cambridge Centre for Molecular Recognition, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, United Kingdom

Communicated by Christopher T. Walsh, Harvard Medical School, Boston, MA, May 8, 1995

ABSTRACT The macrocyclic polyketides rapamycin and FK506 are potent immunosuppressants that prevent T-cell proliferation through specific binding to intracellular protein receptors (immunophilins). The cloning and specific alteration of the biosynthetic genes for these polyketides might allow the biosynthesis of clinically valuable analogues. We report here that three clustered polyketide synthase genes responsible for rapamycin biosynthesis in *Streptomyces hygroscopicus* together encode 14 homologous sets of enzyme activities (modules), each catalyzing a specific round of chain elongation. An adjacent gene encodes a pipecolate-incorporating enzyme, which completes the macrocycle. The total of 70 constituent active sites makes this the most complex multienzyme system identified so far. The DNA region sequenced (107.3 kbp) contains 24 additional open reading frames, some of which code for proteins governing other key steps in rapamycin biosynthesis.

Polyketides are a large and highly diverse group of natural products that includes antibiotics, antitumor compounds, and immunosuppressants. The specific binding of polyketides to prevent T-cell proliferation was reported in 1992 by Schreiber (1) and Rosen and Schreiber (2). These polyketide metabolites are produced by successive condensation of simple carboxylic acid units (primarily acetate and propionate) as for fatty acid biosynthesis (3), except that the β-keto function introduced during each elongation cycle may be reduced only partially or not at all. Macrocyclic polyketides are produced principally by *Streptomyces* and related filamentous bacteria, through the action of so-called type I modular polyketide synthases (PKSs), multienzymes in which different sets (modules) of enzymic activities catalyze each successive round of elongation, as first shown for the erythromycin-producing PKS (4–6). Characterization and genetic engineering of such systems to produce "hybrid" products (7) are particularly challenging because of the large size of the genes and their products and because the factors that control the specificity of chain extension are still largely unknown (7, 8).

Rapamycin (Fig. 1) is a macrocyclic polyketide from *Streptomyces hygroscopicus* that, in addition to its antifungal (13) and antitumor (14) properties, is a potent immunosuppressant (15). Like the structurally related FK506, rapamycin is of interest for the clinical treatment of autoimmune disease (16) and in the prevention of rejection of organ and skin allografts (15, 17). In spite of their similar polyketide backbone, these immunosuppressants act in radically different ways on T cells, FK506 by inhibiting the production of interleukin 2 (1, 2) and rapamycin by preventing the proliferative response to inter-

leukin 2 bound at the interleukin 2 receptor (18). The engineered biosynthesis of altered rapamycins would also be of great interest for the study of these signaling processes. We have therefore undertaken a detailed study of the organization of the rapamycin biosynthetic genes in *S. hygroscopicus*.¶

MATERIALS AND METHODS

Bacterial Strains and Cloning Vectors. *S. hygroscopicus* NRRL 5491 was used as the source of DNA in the construction of the genomic libraries. *Escherichia coli* strain NM538 was used as a host for λEMBL3, and *E. coli* strain XL1-Blue MR was used for pWE15 cosmid (Stratagene) recombinant derivatives. *E. coli* strain TG1 recO served as a host for plasmid subcloning in pUC118 and pUC18 (BRL).

DNA Procedures. Phage, cosmid, and plasmid preparations, DNA restriction digests, size fractionations, DNA-fragment isolations, ligation reactions, and gel electrophoresis were done by standard procedures (19). *E. coli* transformation was done by following the procedure of Hanahan (20). DNA templates for sequencing reactions were purified by Qiagen columns (Qiagen, Chatsworth, CA).

A library of size-fractionated genomic DNA in pUC18 was screened using probes derived from the *eryA* PKS genes from *Saccharopolyspora erythraea* (6). A positive clone (pR19), containing both PKS and "late" genes, was used as the starting point for "chromosome walking." Phage and cosmid clones were obtained that provided multiple coverage over a contiguous 130-kb region of the *S. hygroscopicus* chromosome, and individual inserts were subcloned with plasmid vectors. Sequencing templates were obtained by random subcloning of fragments generated by controlled partial *Hae* III digestions.

DNA Sequencing and Analysis. Automated DNA sequencing was done on double-stranded DNA templates by the dideoxynucleotide chain-termination method (21) with an Applied Biosystems model 373A sequencer. Subclone junctions were verified and occasional sequence gaps were filled by direct sequencing of phage or cosmid clones using synthesized internal oligonucleotide primers. Each nucleotide was sequenced a minimum of three times on both strands. Compilation of the sequence was done using the Staden package (22). DNA and protein sequence homology searches of data bases were done by using the BLAST (23) and FASTA (24) programs. Sequences were analyzed by using the University of Wisconsin Genetics Computer Group (25) programs.
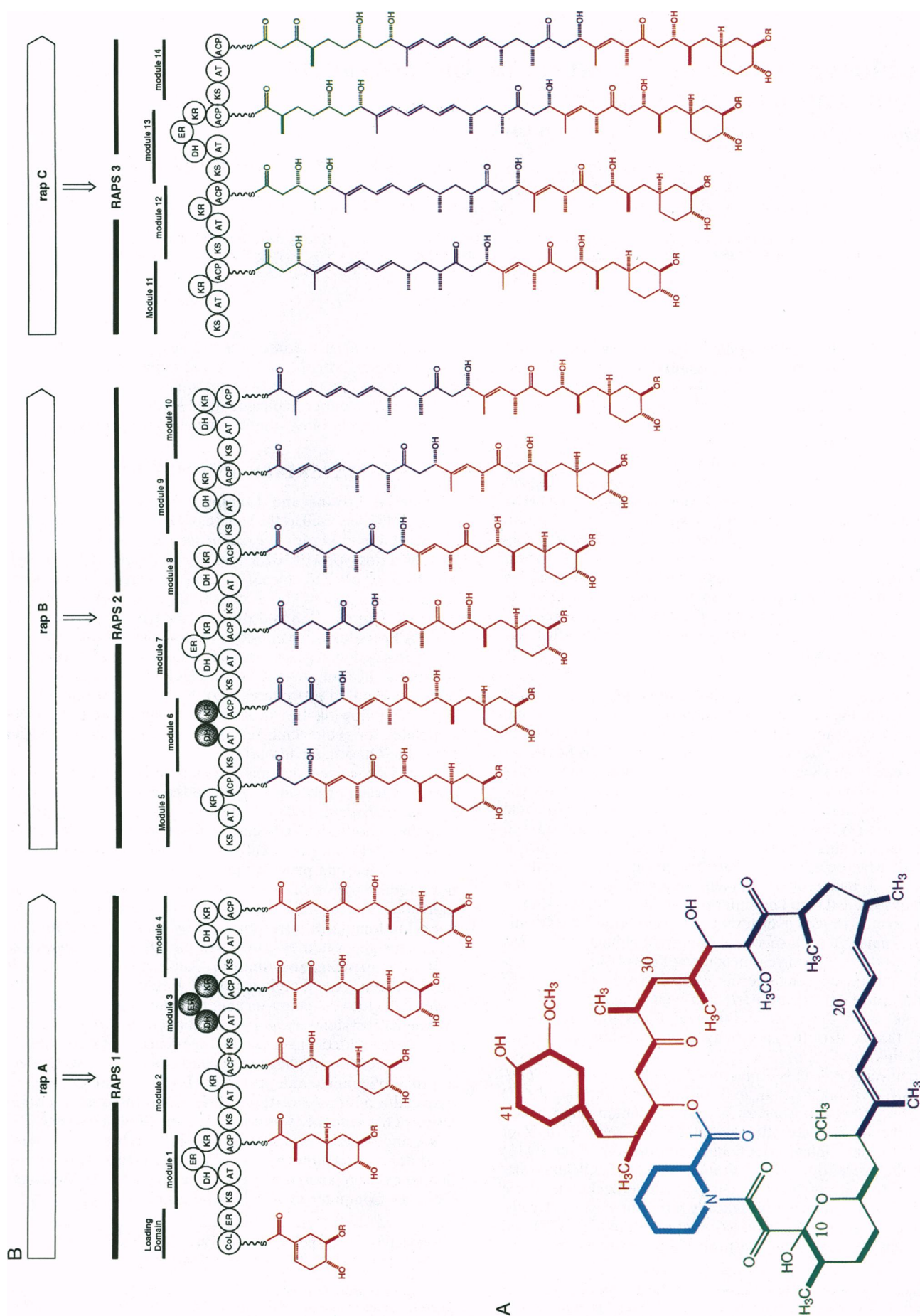

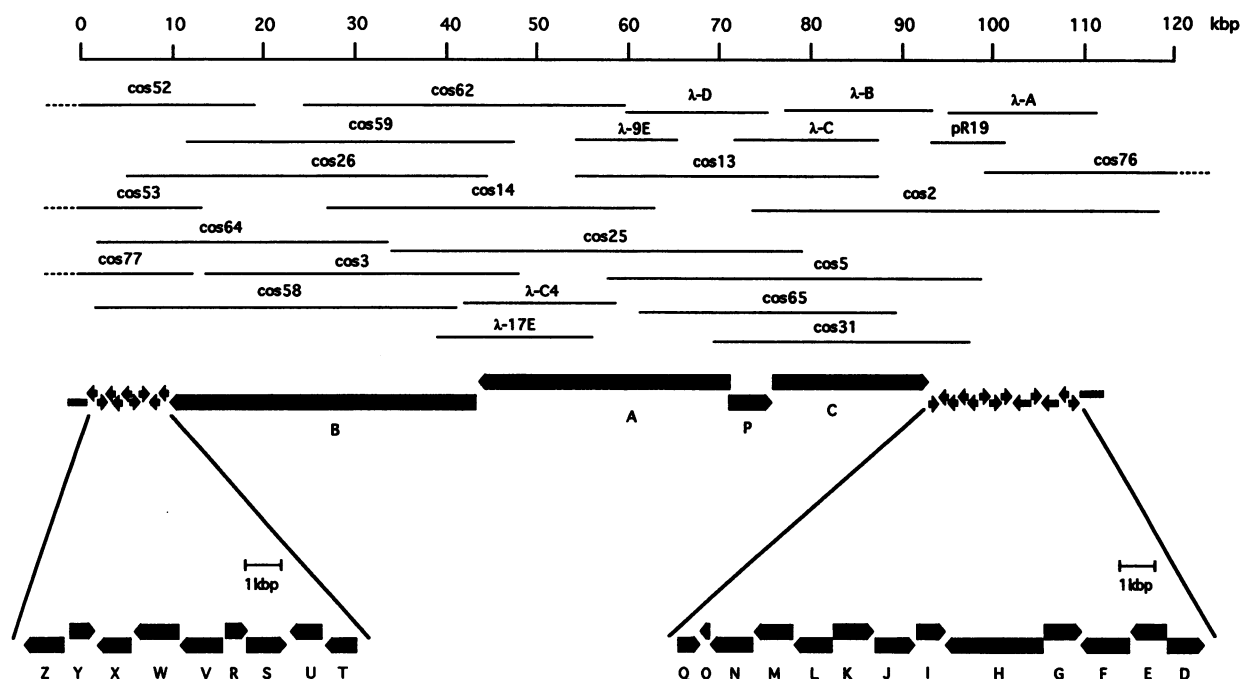FIG. 1. (*Legend appears at the bottom of the opposite page.*)

FIG. 2. Organization of the rapamycin biosynthetic gene cluster. The direction of transcription and relative sizes of the ORFs deduced from analysis of the nucleotide sequence are indicated. Letters under arrows relate to ORFs and *rap* gene products of Table 1.

Protein Expression and Analysis. The pipecolate-incorporating enzyme was expressed in *E. coli* by using a pT7 expression system (26) and purified to homogeneity. The activity was measured by the method of Lee and Lipmann (27).

RESULTS AND DISCUSSION

The rapamycin biosynthetic genes were identified by hybridization using DNA from the PKS genes of *Saccharopolyspora erythraea* that govern the biosynthesis of the macrocyclic polyketide antibiotic erythromycin A (4, 5). A total of 107.3 kbp of contiguous DNA was cloned and sequenced, and the deduced gene organization within this region is shown in Fig. 2. The most striking feature is the presence of three extraordinarily large open reading frames (ORFs), *rapA, rapB,* and *rapC. rapA, rapB,* and *rapC* encode, respectively, the multienzymes Raps1 ($\approx$900 kDa), which contains the first four modules for polyketide chain extension; Raps2 ($\approx$1.07 MDa), which contains the next six modules containing the activities required to continue chain elongation up to C-16 (Fig. 1); and Raps3 ($\approx$660 kDa), which contains four modules required to complete the polyketide portion of the rapamycin ring (Table 1). In total, the three giant proteins contribute 70 catalytic functions, making this the largest and most complex multienzyme system identified so far. The integration of six modules within a single protein (Raps2) is also unprecedented. Previously studied modular PKSs possess either two modules per protein (4, 5) or up to three (28). In modules three and six there are predicted active sites for reduction, which are not reflected in the ultimate structure of rapamycin and which may be nonfunctional (28). It is possible that temporary reduction is necessary at C-26 and C-32 to provide conformational flexibility, so that the growing chain does not fold back on itself and inhibit further elongation.

Chain initiation on the rapamycin PKS may require an ATP-dependent carboxylic acid: CoA ligase. Such an activity

was identified at the N terminus of Raps1, adjacent to an enoyl reductase activity. The only precedent for such an adenylate-forming domain is a sequence of unknown function from *Bacillus subtilis* (29). Previous work (9, 10) has suggested that the substituted cyclohexanecarboxylic acid starter unit for the polyketide immunosuppressants is derived by reduction of shikimate, and our results indicate that at least one reductive step may occur after the CoA ligase has linked the starter unit to the Raps1 multienzyme.

The rapamycin PKS also has an unusual mechanism for chain termination and cyclization: embedded between the PKS genes and translationally coupled to *rapC* is a gene (*rapP*) with striking sequence similarity to activation domains of nonribosomal peptide synthetases (30). *rapP* encodes a polypeptide of 170 kDa, which is exactly the size of the analogous enzyme purified from *S. hygroscopicus* var. *ascomyceticus*, which inserts pipecolic acid into the FK506 analogue immunomycin (31). We propose that the polyketide chain of rapamycin is transferred from a thioester linkage on Raps3 directly to the amino group of an enzyme-bound pipecolyl moiety, which in turn is attacked by the C-34 hydroxyl group to form the macrolactam ring.

Lying to the right of the PKS genes in Fig. 2 are 13 other ORFs, for some of which there is significant sequence similarity with known proteins in published data bases. Several of the identified genes can be plausibly assigned roles in rapamycin biosynthesis. The products of *rapM* and *rapQ* resemble methyltransferases (32, 33) and are presumably required for O-methylation of hydroxyl groups on either C-39, C-27, or C-16. The clusters of both FK506 and immunomycin contain such methyltransferases (34). Likewise, the gene products of *rapJ* and *rapN* resemble cytochrome P450 enzymes (35) and may hydroxylate C-9 and C-27 (and possibly also C-26 or C-32). The product of *rapL* shows significant sequence identity with ornithine cyclodeaminase (36), which implicates it in the cyclization of lysine to provide pipecolic acid. DNA sequencing

FIG. 1. (*A*) Structure of rapamycin. Bold lines indicate the biosynthetic origin of the carbon atoms (9–12). (*B*) Domain organization and biosynthetic intermediates. Color code is as follows: Raps1 product in red, Raps2 product in dark blue, Raps3 product in green, and pipecolic acid-incorporating enzyme product in light blue. Shaded domains are possibly inactive (see text).

Table 1.   Deduced functions of ORFs in the gene cluster

Polypeptide	Amino acids, no.	Proposed function and sequence similarities detected					
Raps1	8,563	PKS					
Loading domain		CoA ligase			ER		
Module 1		KS	AT(P)	DH	ER	KR	ACP
Module 2		KS	AT(A)			KR	ACP
Module 3		KS	AT(P)	DH	ER	KR*	ACP
Module 4		KS	AT(P)	DH		KR	ACP
Raps2	10,222	PKS					
Module 5		KS	AT(A)			KR	ACP
Module 6		KS	AT(P)	DH		KR*	ACP
Module 7		KS	AT(P)	DH	ER	KR	ACP
Module 8		KS	AT(A)	DH		KR	ACP
Module 9		KS	AT(A)	DH		KR	ACP
Module 10		KS	AT(P)	DH		KR	ACP
Raps3	6,260	PKS					
Module 11		KS	AT(A)			KR	ACP
Module 12		KS	AT(A)			KR	ACP
Module 13		KS	AT(P)	DH	ER	KR	ACP
Module 14		KS	AT(A)				ACP
ORF D	383	?					
ORF E	465	?					
ORF F	454	(Membrane transport protein?)					
ORF G	330	(Helix-turn-helix DNA-binding protein?)					
ORF H	872	(Putative regulator of cholesterol oxidase)					
ORF I	260	?					
RapJ	386	Cytochrome P450					
ORF K	341	?					
RapL	343	Lysine cyclodeaminase					
RapM	317	Methyltransferase					
RapN	404	Cytochrome P450					
RapO	78	Ferredoxin					
RapP	1,541	Pipecolate incorporating enzyme					
RapQ	211	Methyltransferase					
RapT	264	(Ketoreductase/dehydrogenase)					
ORF U	200	?					
ORF S	399	(Sensory protein kinase)					
ORF R	220	(Response regulator)					
ORF V	437	(Membrane transport protein?)					
ORF W	459	?					
ORF X	235	(ABC-transporter)					
ORF Y	204	(Regulators of antibiotic transport complexes?)					
ORF Z	389	?					

Potential PKS catalytic activities are indicated as follows: ACP, acyl carrier protein; KS, β-ketoacyl-ACP synthase; AT(A), acyltransferase incorporating an acetate extender unit; AT(P), acyltransferase incorporating a propionate extender unit; KR, β-ketoacyl-ACP reductase; DH, β-hydroxyacyl-thioester dehydratase; ER, enoyl reductase. Activities listed for Raps proteins are colinear with the sequence of each ORF as listed from left to right and from top to bottom. Asterisks indicate that the enzyme activity is possibly nonfunctional (see text). Detected sequence similarities where the proposed function remains uncertain are shown in parentheses. Partial or low-level sequence similarities are indicated by a question mark.

beyond ORF D has uncovered genes that appear to be required for cysteine biosynthesis, indicating the limit of the rapamycin cluster (A.K., T.S., J.F.A., and I.M., unpublished results).

On the left of the PKS genes in Fig. 2 are nine ORFs, for most of which no obvious function can be suggested. *rapT*, immediately downstream of *rapB*, encodes a discrete enzyme with significant sequence similarity to ketoreductases found in aromatic PKSs and fatty acid synthases. Beyond ORF Z lie genes that appear to be part of an operon for carbohydrate utilization, indicating the other probable limit of the rapamycin cluster (T.S., J.F.A., and I.M., unpublished results).

The cloned and sequenced DNA can be used for targeted gene disruption and the complementation of mutants, to provide final confirmation of the identity of the gene cluster, although the *S. hygroscopicus* strain has so far proved highly resistant to transformation. The evidence that the rapamycin

gene cluster has indeed been cloned includes the following: (*i*) the presence of 14 modules of a type I PKS, as required for the addition of the seven acetate and seven propionate units of the rapamycin chain; (*ii*) the unconventional loading domain; (*iii*) the correct oxidation level predicted for 12 of the 14 modules, assuming a colinear arrangement of genes and activities as in the erythromycin-producing PKS (4, 5); and (*iv*) the lack of a gene encoding a conventional thioesterase off-loading domain. The latter is replaced in the cluster by the *rapP* gene, encoding an enzyme activating and incorporating pipecolic acid. To confirm the substrate specificity of the *rapP* gene product, the gene was overexpressed in *E. coli* and purified to homogeneity (unpublished work). The enzyme specifically catalyzed the L-pipecolic acid-stimulated exchange between pyrophosphate and ATP (A.K. and T.S., unpublished work), and it is therefore designated pipecolic acid-incorporating enzyme. The entirely analogous enzyme from *S. hygroscopicus* var. *ascomyceticus* has
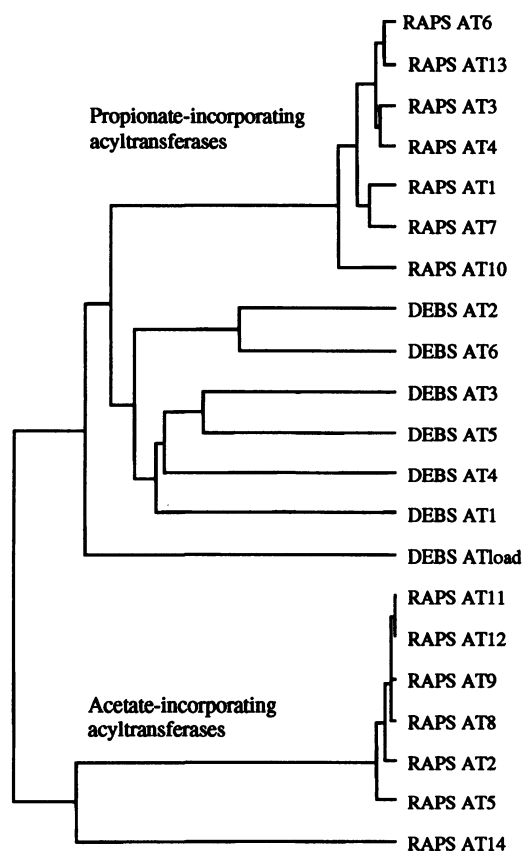
FIG. 3. Sequence similarity between acyltransferase (AT) domains that load either methylmalonyl-CoA (propionate extender units) or malonyl-CoA (acetate extender units) in modular type I PKSs for erythromycin and rapamycin biosynthesis. ATload is the domain that loads the starter unit from propionyl-CoA in the erythromycin PKS (37). The dendrogram was constructed with the program PILEUP (25). Numbering of rapamycin PKS modules corresponds to that in Table 1.

been convincingly linked to production of the FK506-analogous immunomycin (31). The polyketide backbone of rapamycin and FK506 are identical in the region formed by *rapC* and its counterpart FKB-AI[||] apart from the degree of reduction at C-13 (hydroxyl group in FK506 vs. methylene in rapamycin). In agreement with this, *rapC* and FKB-AI house the same enzymic activities in the same order and differ only in the extra β-hydroxyacyl-thioester dehydratase and enoyl reductase required to reduce the -OH group at C-13 to methylene in rapamycin. The apparent translational coupling of *rapA* and *rapB* and the very high sequence identity between rapamycin PKS modules suggest that the PKS genes for rapamycin and FK506 may have diverged from a common ancestor through recent recombination events. Finally, analysis of the sequences of the acyltransferase domains from each of the 14 modules and their comparison with the sequences (4, 5) of the methylmalonyl-CoA:ACP acyltransferases from the erythromycin-producing PKS reveal (Fig. 3) that the acyltransferase specificity of a module for an acetate or a propionate extender unit can be unambiguously predicted from its sequence (S.F.H., J.F.A., and I.M., unpublished work). The assignment of modules to condensation cycles made on the basis of the reductive activities present (Table 1) is in agree-

ment with the deduced specificity of each acyltransferase in Fig. 3.

The complete gene sequencing of a modular type I PKS for an immunosuppressant, together with most (if not all) of the other genes required for biosynthesis, is not only a prerequisite for engineering of such systems, it also promises further insight into the molecular signals that regulate the biosynthesis of a major class of antibiotics and other bioactive compounds. The sophistication of the enzymology involved also holds important lessons for our understanding of the natural evolution of biological catalysis.

1. Schreiber, S. L. (1992) *Cell* 70, 365–368.
2. Rosen, M. K. & Schreiber, S. L. (1992) *Angew. Chem. Int. Ed. Engl.* 31, 384–400.
3. Hopwood, D. A. & Sherman, D. H. (1990) *Annu. Rev. Genet.* 24, 37–66.
4. Cortés, J., Haydock, S. F., Roberts, G. A., Bevitt, D. J. & Leadlay, P. F. (1990) *Nature (London)* 348, 176–178.
5. Donadio, S., Staver, M. J., McAlpine, J. B., Swanson, S. J. & Katz, L. (1991) *Science* 252, 675–679.
6. Bevitt, D. J., Cortés, J., Haydock, S. F. & Leadlay, P. F. (1992) *Eur. J. Biochem.* 204, 39–49.
7. Katz, L. & Donadio, S. (1993) *Annu. Rev. Microbiol.* 47, 875–912.
8. Marsden, A. F. A., Caffrey, P., Aparicio, J. F., Loughran, M. S., Staunton, J. & Leadlay, P. F. (1994) *Science* 263, 378–380.
9. Paiva, N. L., Demain, A. & Roberts, M. F. (1993) *Enzyme Microb. Technol.* 15, 581–585.
10. Wallace, K. K., Reynolds, K. A., Koch, K., McArthur, H. A. I., Brown, M. S., Wax, R. G. & Moore, B. S. (1994) *J. Am. Chem. Soc.* 116, 11600–11601.
11. Paiva, N. L. & Demain, A. L. (1991) *J. Nat. Prod.* 54, 167–177.
12. McAlpine, J. B., Swanson, S. J., Jackson, M. & Whittern, D. N. (1991) *J. Antibiot.* 44, 688–690.
13. Vezina, C., Kudelski, A. & Sehgal, S. N. (1975) *J. Antibiot.* 28, 721–726.
14. Durous, J. & Suffness, M. (1981) *Cancer Treat. Rev.* 8, 63–69.
15. Calne, R. Y., Collier, D. S., Lim, S., Polard, S. G., Samaan, A., White, D. J. G. & Thiru, S. (1989) *Lancet* ii, 227.
16. Baeder, W. L. (1992) *Clin. Exp. Immunol.* 89, 174–178.
17. Collier, D. S., Calne, R. S., Pollard, S. G., Friend, P. J. & Thiru, S. (1991) *Transplant. Proc.* 23, 246–247.
18. Brown, E. J., Albers, M. W., Shin, T. B., Ichikawa, K., Keith, C. T., Lane, W. S. & Schreiber, S. L. (1994) *Nature (London)* 369, 756–758.
19. Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab. Press, Plainview, NY), 2nd Ed.
20. Hanahan, D. (1985) in *DNA Cloning: A Practical Approach*, ed. Glover, D. M. (IRL, Oxford), Vol. 1, pp. 109–132.
21. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* 74, 5463–5467.
22. Staden, R. (1984) *Nucleic Acids Res.* 12, 521–528.
23. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990) *J. Mol. Biol.* 215, 403–410.
24. Pearson, W. R. (1990) *Methods Enzymol.* 183, 63–98.
25. Devereux, J., Haeberli, P. & Smithies, O. (1984) *Nucleic Acids Res.* 12, 387–395.
26. Tabor, S. & Richardson, C. C. (1985) *Proc. Natl. Acad. Sci. USA* 82, 1074–1078.
27. Lee, S. & Lipmann, F. (1975) *Methods Enzymol.* 43, 585–602.
28. MacNeil, D. J., Occi, J. L., Gewain, K. M., MacNeil, T., Gibbons, P. H., Ruby, C. L. & Danis, S. J. (1992) *Gene* 115, 119–125.
29. Albertini, A. M., Caramori, T., Scoffone, F., Scotti, C. & Galizzi, A. (1995) *Microbiology (Reading, U.K.)* 141, 299–309.
30. Marahiel, M. A. (1992) *FEBS Lett.* 307, 40–43.
31. Nielsen, J. B., Hsu, M.-J., Byrne, K. M. & Kaplan, L. (1991) *Biochemistry* 30, 5789–5796.
32. Clarke, L. & Carbon, J. (1976) *Cell* 9, 91–99.
33. Haydock, S. F., Dowson, J. A., Dhillon, N., Roberts, G. A., Cortés, J. & Leadlay, P. F. (1991) *Mol. Gen. Genet.* 230, 120–128.
34. Shafiee, A., Motamedi, H. & Chen, T. (1994) *Eur. J. Biochem.* 225, 755–764.
35. Horii, M., Ishizaki, T., Paik, S.-Y., Manome, T. & Murooka, Y. (1990) *J. Bacteriol.* 172, 3644–3653.
36. Sans, N., Schindler, U. & Schröder, J. (1988) *Eur. J. Biochem.* 173, 123–130.
37. Aparicio, J. F., Caffrey, P., Marsden, A. F. A., Staunton, J. & Leadlay, P. F. (1994) *J. Biol. Chem.* 269, 8524–8528.

[||]Motamedi, H., Cai, S.J., Streicher, S.S. & Shafiee, A., Seventh International Symposium of the Genetics of Industrial Microorganisms, June 26–July 1, 1994, Montreal.