Rapamycin biosynthesis: elucidation of gene product function†

Matthew A. Gregory,[a] Hui Hong,[b] Rachel E. Lill,[a] Sabine Gaisser,[a] Hrvoje Petkovic,[a] Lindsey Low,[a] Lesley S. Sheehan,[a] Isabelle Carletti,[a] Sarah J. Ready,[a] Michael J. Ward,[a] Andrew L. Kaja,[a] Alison J. Weston,[a] Iain R. Challis,[a] Peter F. Leadlay,[c] Christine J. Martin,[a] Barrie Wilkinson[a] and Rose M. Sheridan*[a]

Received 21st June 2006, Accepted 11th August 2006
First published as an Advance Article on the web 22nd August 2006
DOI: 10.1039/b608813a

The function of gene products involved in the biosynthesis of the clinically important polyketide rapamycin were elucidated by biotransformation and gene complementation

Interest in rapamycin **1** and its derivatives has surged over the past few years following appreciation of their immense pharmacological potential. These compounds are currently the most selective kinase inhibitors known and the only inhibitors of mTOR (mammalian target of rapamycin) associated kinase activity.[1] Rapamycins bind the cyclophilin FKBP12, and this complex binds and inhibits the function of mTOR. mTOR, a serine–threonine kinase, appears to act as a central controller sensing cellular environment and regulating translation initiation through the eukaryotic initiation factor 4E and ribosomal p70 S6 kinase pathways. Rapamycin (Sirolimus, Rapamune) has been licensed for use as an immunosuppressant after organ transplantation, and has shown potential therapeutic use in the treatment of cardiovascular, autoimmune and neurodegenerative diseases.[2] In addition, rapamycin and its derivatives represent exciting candidates for anticancer therapeutic development, with three rapamycin derivatives, CCI-779, RAD001 and AP23573, presently in clinical trials as anticancer agents.[3]

Rapamycin is a lipophilic macrolide produced by *Streptomyces hygroscopicus* NRRL 5491.[4] The biosynthetic gene cluster has been shown to encode a modular polyketide synthase (PKS), a non-ribosomal peptide synthetase-like protein, and other biosynthetic genes.[5] Biosynthesis is initiated by incorporation of a shikimate derived 4,5-dihydroxycyclohex-1-ene carboxylic acid (DHCHC) starter unit,[6] the biosynthesis of which is dependent on RapK.[7] Macrocycle biosynthesis is completed by incorporation of pipecolic acid (an L-lysine derived amino acid formed by the L-lysine cyclodeaminase RapL[8]) and ring closure. The first enzyme free intermediate, pre-rapamycin **2**,[9] is thought to be modified by three methyltransferases (*rapI*, *rapM* and *rapQ*) and two cytochrome P450 monooxygenases (*rapJ*, *rapN*), with the support of the associated putative ferridoxin *rapO*, to yield **1** (Fig. 1).

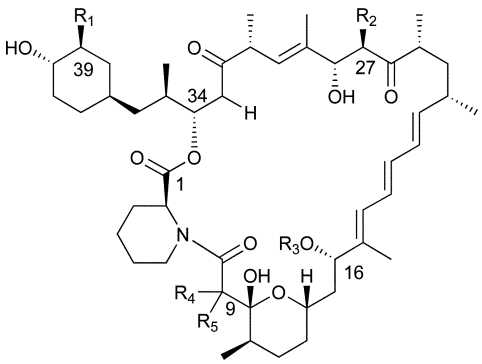

Fig. 1 Structure of rapamycin and possible post-PKS variations; R_1 = OH or OCH_3; R_2 = H, OH or OCH_3; R_3 = H or CH_3; R_4/R_5 = H/H or O. For rapamycin (**1**), R_1, R_2 = OCH_3; R_3 = CH_3; R_4/R_5 = O; for pre-rapamycin (**2**), R_1 = OH, R_2 = H, R_3 = H, R_4/R_5 = H/H.

In the present study, complete rapamycin genes associated with the modification of **2** were inserted in isolation or in combination into the chromosome of *S. hygroscopicus* MG2–10, comprising the deletion of genes *rapKIJMNOQL*,[9] and *Streptomyces lividans* TK24, using integrative expression plasmids based on pSET152.[10] In *S. hygroscopicus*, *rapK* and *rapL* were always included to ensure precursor supply except where indicated. Gene expression was driven by the promoter P*actI* with its cognate activator *actII*-orf4 from *Streptomyces coelicolor*,[10] which was found to be an effective promoter in both organisms. Thus, through a combination of complementation into the deletion mutant *S. hygroscopicus* MG2–10, and heterologous biotransformation with *S. lividans* TK24, the gene product functions were elucidated and differentiated, and rapamycin analogues were isolated (see ESI† for methods and analysis).

To confirm the function and differentiate the site of action of RapI and RapM, *S. hygroscopicus* MG2–10 was complemented with *rapI* (BIOT-2034), *rapM* (BIOT-2209) or *rapIM* (BIOT-2210). Compounds **3**, **4** and **5** respectively were isolated in good yields and to a high degree of purity using preparative chromatography. Structural elucidation was achieved by a combination of LCMS/MS and 2D NMR analysis. Full detail is provided in the supplementary information but, briefly, LCMS/MS isolated the structural changes to the anticipated fragments of the molecule as shown in Table 1. These fragmentation pathways have been verified for **1** and **2** using accurate mass FT-ICR-MS/MS methods and are directly applicable to our intermediate structures (Fig. 2).[7,9] NMR data were fully consistent with the anticipated structures.

All compounds were derivatives of **2**. **3** was *O*-methylated only at the C39 position, and **4** was *O*-methylated only at the C16 position. Where the deletion strain was complemented with *rapIM*, the

[a]*Biotica, Chesterford Research Park, Little Chesterford, UK CB10 1XL. E-mail: rose.sheridan@biotica.com; Fax: +44 (0)1799 532920; Tel: +44 (0)1799 532925*
[b]*Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge, UK CB2 1EW*
[c]*Department of Biochemistry, University of Cambridge, 80 Tennis Court Road, Cambridge, UK CB2 1GA*

† Electronic supplementary information (ESI) available: Methods and data for strain and plasmid construction, fermentation, isolation, spectroscopic analysis and structural determination. See DOI: 10.1039/b608813a

Table 1 Isolated rapamycin analogues and their associated mass spectrometry fragmentation patterns

	R_1	R_2	R_3	R_4/R_5	A	B	C	D	E
1	OMe	OMe	Me	=O	936.4	807.2	642.2	614.2	320.1
2	OH	H	H	H,H	864.3	735.3	614.2	556.2	306.0
3	OMe	H	H	H,H	878.4	749.3	628.3	556.2	306.0
4	OH	H	Me	H,H	878.4	749.4	614.2	570.2	306.0
5	OMe	H	Me	H,H	892.7	763.6	628.4	570.3	306.0
6	OMe	OH	H	H,H	894.7	765.6	628.4	572.3	306.0
7	OH	OH	Me	H,H	894.4	765.4	614.3	586.2	306.0
8	H	OMe	H	=O	892.5	763.5	612.3	600.4	320.1
9	OH	H	H	=O	878.4	749.3	628.2	570.2	320.1

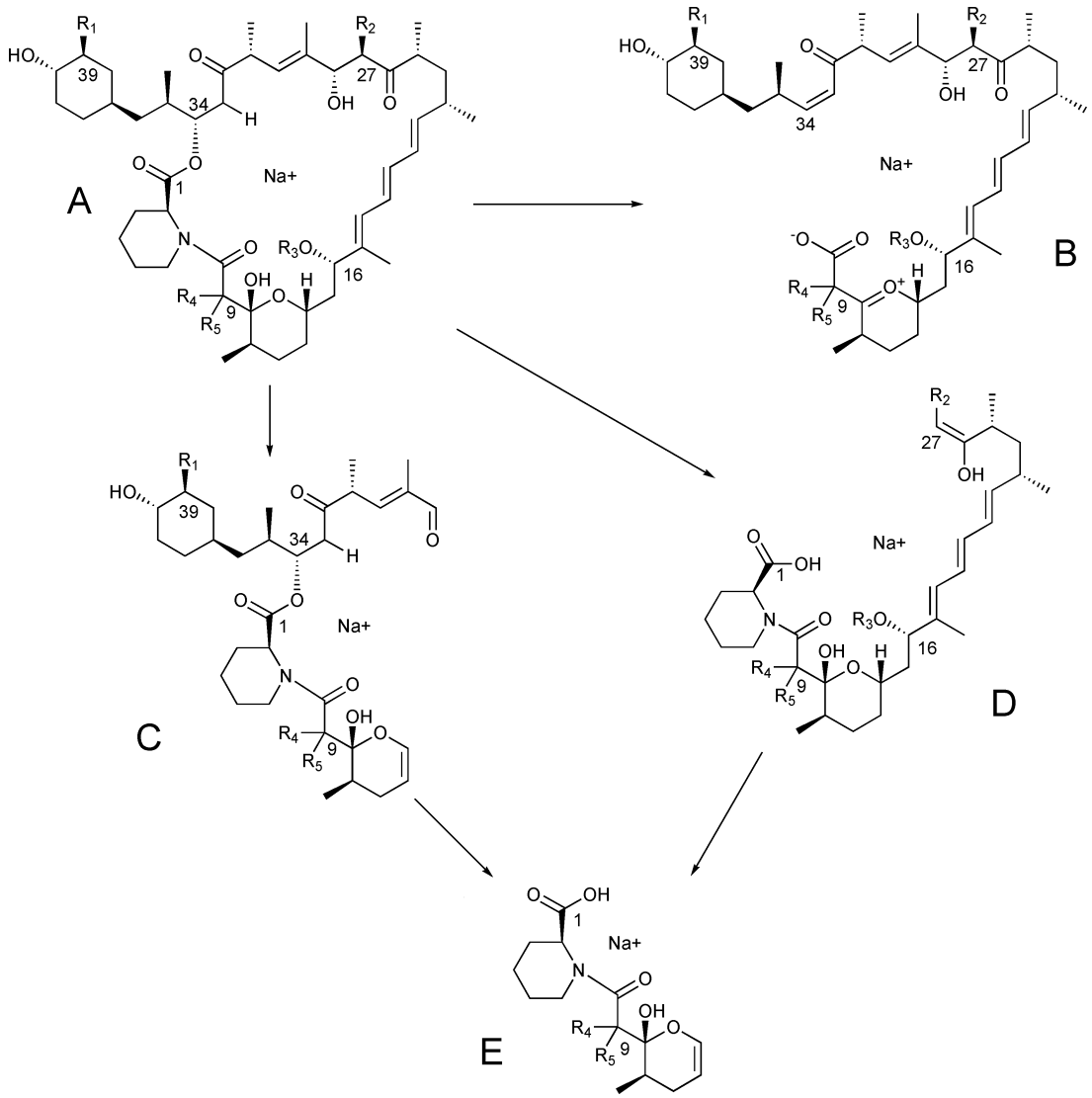

Fig. 2 Fragmentation pattern of rapamycin and structural definition of rapamycin mass spectrometry fragments A–E.

isolated compound **5** was *O*-methylated both at C39 and C16. As further confirmation of the function and site of action of these enzymes, *rapI* and *rapM* were heterologously expressed separately in *S. lividans* TK24. On feeding **2** to these strains, compounds consistent with **3** and **4** respectively were observed by chromatographic methods and MS/MS for which the masses of fragments C and D are characteristic for these changes (Table 1).

A third *O*-methylation at C27 is also observed during the biosynthesis of **1**. Prior to this, hydroxylation at C27 of **2** is required and RapN had been postulated as the monooxygenase

responsible for insertion of oxygen into the C–H bond at this position. Due to the juxtaposition of *rapO* to *rapN* in the biosynthetic gene cluster, where they are present as an operon of translationally fused open reading frames under a single promoter, *rapO* (encoding the ferridoxin RapO) was retained with *rapN* with the assumption that this ferridoxin supports electron transfer to RapN. From the *S. hygroscopicus* MG2–10 strains complemented with *rapINO* and *rapMNO* respectively (BIOT-2212 and BIOT-1815) the compounds **6** and **7** were isolated. After analysis by LCMS/MS and 2D NMR it was clear that both compounds were hydroxylated at C27. Moreover, no C27-methoxy forms of **6** or **7** were observed in the culture broths, and it would appear that the *O*-methyltransferases RapI and RapM are unable to *O*-methylate this position. As expected, compounds **6** and **7** were *O*-methylated at the C39 and C16 groups respectively, confirming the absolute regio-specificity of the *O*-methyltransferases RapI for the C39 position and RapM for the C16 position.

The third candidate *O*-methyltransferase, RapQ, was investigated in the *S. hygroscopicus* MG2–10 strain BIOT-3373, complemented with only *rapJNOQL* and fed cyclohexanecarboxylic acid (CHCA), an analogue of DHCHC which is not produced in this strain due to the absence of RapK. This starter unit is hydroxylated at the C4 position prior to its incorporation by the rapamycin PKS.[7] A compound **8** was isolated in good yield and to high purity following fermentation. Based on 2D NMR analysis a methoxy group was clearly shown to be present at C27, demonstrating the regio-specificity of RapQ for this position. When *rapQ* alone was heterologously expressed in *S. lividans* TK24, no modification of **2** (which lacks the C27 hydroxyl group) was detected further supporting the fact that RapQ does not *O*-methylate other positions.

Confirmation and differentiation of the function and specificity of the two cytochrome P450 monooxygenases was performed using the same approach of complementation and heterologous biotransformation. As described above, the inclusion of *rapN* in a complemented strain resulted in hydroxylation at C27 alone confirming the monooxygenase function of RapN and demonstrating its regio-specificity for this position.

Strain BIOT-3373 producing **8** described above is complemented with both monooxygenase genes *rapN* and *rapJ.* As seen for **6** and **7**, hydroxlation by RapN at C27 had occurred allowing *O*-methylation of this group by RapQ. Additionally, a keto group was incorporated at C9 in **8**, something which had not been observed in the absence of *rapJ* in any of the strains tested above. This suggested that RapJ had the suprising ability of performing two rounds of oxygen insertion to form the C9-keto group. To further investigate the function of RapJ, *S. hygroscopicus* MG2–10 was complemented with *rapJ* (BIOT-1811). The resulting compound **9** was isolated in good yield and to high purity. However, due to the presence of multiple rotamers of **9** it was difficult to fully assign the structure of this compound. Fortunately, it was clear that no new methoxy signals were present in the NMR spectra and the connectivity around the C9-keto group could be determined for the major rotamer, verifying the site of action of RapJ and its function as a monooxygenase. Additional analysis by LCMS/MS was consistent with this modification as the increase in mass by 14 units was clearly not due to *O*-methylation and was located within fragment E (Table 1). As further confirmation of the function of RapJ, *S. lividans* TK24 complemented with *rapJ*, *rapK* or *rapKJ* respectively were fed with **2** and the culture extracts analysed by LCMS/MS. The strains expressing *rapJ* alone or in combination with *rapK* converted **2** to a compound with an LCMS/MS profile consistent with **9**. There is significant precedent for a cytochrome P450 monooxygenase to catalyse two rounds of oxygen insertion. The enzyme BorI involved in borrelidin biosynthesis is believed to convert the C12-methyl carbon of the borrelidin precursor to a formyl species prior to the biosynthesis of a novel nitrile moiety.[11] This activity has been confirmed *in vitro* (unpublished work).[12] Similarly, the cytochrome P450 monooxygenase TylI is involved in the formation of a C18-formyl group during the biosynthesis of the macrolide tylosin.[13] A BLAST protein homology search of RapJ revealed that amongst its closest relatives are PimG, CanC, AmphN and NysN. It has been proposed that these enzymes might catalyse the oxidation of a methyl group to a carboxylate, and based on the analysis of genetic mutation AmphN does appear to introduce two oxygen atoms into an unactivated methyl group.[14]

Finally, complete complementation of *S. hygroscopicus* MG2–10 with *rapIJMNOQ* (BIOT-3206) resulted in the production of **1** at wild-type levels, confirmed by LCMS/MS and 2D NMR and was consistent with a pure standard of **1**.

It is striking that complementation of the deletion strain, *S. hygroscopicus* MG2–10, produces such a variety of compounds. Indeed all twenty-four possible intermediate compounds between **2** and **1** have been observed after the complementation of this strain with all possible combinations of the deleted genes (data not shown). However, not all are produced in good yield, indicating that although there are clear parallel pathways for complete rapamycin biosynthesis, there is a preferred route or routes. On complementation with *rapI*, *rapM* or *rapJ*, the expected compounds **3**, **4**, **9**, were efficiently produced through fermentation. *O*-Methylation at C27 is dependent first upon the action of RapN, forming the C27 hydroxyl group. However, yields of **6** and **7** from strains BIOT-2212 and BIOT-1815 were low with **4** accumulating as an intermediate in BIOT-1815 fermentations suggesting **4** is not the ideal substrate for RapN. The *O*-methylation at C27 by RapQ was consistently found to be the most difficult modification to achieve, but appeared to occur more readily after introduction of the C9-keto group. In addition, this *O*-methylation appeared to occur late in the fermentation process with the intermediate 27-*O*-desmethylrapamycin accumulating and clearly being converted to the final product later on during the fermentation of the wild-type rapamycin producer. We propose that a preferred complete rapamycin biosynthetic route occurs *via O*-methylation at C39 (RapI), introduction of a keto group at C9 (RapJ), *O*-methylation at C16 (RapM), hydroxylation at C27 (RapN) with the final step being *O*-methylation at C27 (RapQ).

Utilising the versatility of this system has allowed the production of a wide range of rapamycin compounds[7,8] in order to explore the structure activity relationships (SAR) of rapamycin analogues; a more comprehensive and detailed description of their properties will be provided elsewhere. Simple changes in the level of *O*-methylation have resulted in the production of more soluble compounds with altered metabolism profiles, for example compound **6** is three times more soluble than rapamycin, has an improved Caco-2 permeabilty and a 20 fold reduction in Caco-2 B–A/A–B flux ratio (efflux) relative to **1**. Key sites of metabolism, such as the C39 methoxy group, which is the major target of

Cyp3A4 for rapamycin metabolism, can be removed by feeding CHCA to strains lacking RapK. Despite these changes to the molecule, potency can be retained, with **6** comparing favourably with rapamycin in the NCI 60 cancer cell line panel giving an average log GI_{50} of −6.22 *versus* −6.8 for rapamycin. Furthermore, in a COMPARE analysis[15] *versus* rapamycin, **6** scores 0.5 suggesting that although the two compounds have a similar average potency there are significant differences in the profile of activities against the cell lines.

By controlling the post-PKS processing of **2** and combining this with alternative starter units, pipecolate analogues, and even PKS level changes, the combinatorial biosynthesis of a wide range of rapamycin analogues is now possible. These offer great potential as new agents with enhanced therapeutic properties and as chemical genetic probes of the molecular pathways involved in rapamycin action.

Acknowledgements

We thank Dr Paul Gates for help with FT-ICR-MS analysis. We also thank the staff of the DNA sequencing facility, Department of Biochemistry. Work done in PFL's laboratory was supported by a project grant from the BBSRC (UK). We also thank Inpharmatica Ltd., Cambridge, UK for the solubility and Caco-2 permeation assays and the National Institute of Cancer, Bethesda, USA for the *in vitro* 60 cell line anticancer assays.

References

1 S. Huang and P. J. Houghton, *Drug Resist. Updates*, 2002, **4**, 378–391.
2 S. Huang, M. A. Bjornsti and P. J. Houghton, *Cancer Biol. Ther.*, 2003, **2**, 222–232; E. C. Perin, *Rev. Cardiovasc. Med.*, 2005, **6**, S13–S21; R. P. Carlson, D. A. Hartman, L. A. Tomchek, T. L. Walter, J. R. Lugay, W. Calhoun, S. N. Sehgal and J. Y. Chang, *J. Pharmacol. Exp. Ther.*, 1993, **266**, 1125–1138; Z. Berger, B. Ravikumar, F. M. Mendez, L. G. Oroz, B. R. Underwood, M. N. Pangalos, I. Schmitt, U. Wullner, B. O. Evert, C. J. O'Kane and D. C. Rubinsztein, *Hum. Mol. Genet.*, 2006, **15**, 433–442.
3 S. Vignot, S. Faivre, D. Aguire and E. Raymond, *Ann. Oncol.*, 2005, **16**, 525–537.
4 S. N. Sehgal, H. Baker and C. Vézina, *J. Antibiot.*, 1975, **2**, 727–733.
5 T. Schwecke, J. F. Aparicio, I. Molnár, A. König, L. E. Khaw, S. F. Haydock, M. Oliynyk, P. Caffrey, J. Cortés, J. B. Lester, G. A. Böhm, J. Staunton and P. F. Leadlay, *Proc. Natl. Acad. Sci. U. S. A.*, 1995, **92**, 7839–7843; N. L. Paiva, A. L. Demain and M. F. Roberts, *J. Nat. Prod.*, 1991, **54**, 167–177; I. Molnár, J. F. Aparicio, S. F. Haydock, L. E. Khaw, T. Schwecke, A. König, J. Staunton and P. F. Leadlay, *Gene*, 1996, **16**, 1–7; J. F. Aparicio, I. Molnár, T. Schwecke, A. König, S. F. Haydock, L. E. Khaw, J. Staunton and P. F. Leadlay, *Gene*, 1996, **16**, 9–16; N. L. Paiva, A. L. Demain and M. F. Roberts, *Enzyme Microb. Technol.*, 1993, **15**, 581–585.
6 P. A. S. Lowden, B. Wilkinson, G. A. Böhm, S. Handa, H. G. Floss, P. F. Leadlay and J. Staunton, *Angew. Chem., Int. Ed.*, 2001, **40**, 777–779.
7 M. A. Gregory, H. Petkovic, R. E. Lill, S. J. Moss, B. Wilkinson, S. Gaisser, P. F. Leadlay and R. M. Sheridan, *Angew. Chem., Int. Ed.*, 2005, **44**, 4757–4760.
8 L. E. Khaw, G. A. Böhm, S. Metcalfe, J. Staunton and P. F. Leadlay, *J. Bacteriol.*, 1998, **180**, 809–814; G. J. Gatto, Jr., M. T. Boyne, N. L. Kelleher and C. T. Walsh, *J. Am. Chem. Soc.*, 2006, **128**, 3838–3847.
9 M. A. Gregory, S. Gaisser, R. E. Lill, H. Hong, R. M. Sheridan, B. Wilkinson, H. Petkovic, A. J. Weston, I. Carletti, H. L. Lee, J. Staunton and P. F. Leadlay, *Angew. Chem., Int. Ed.*, 2004, **43**, 2551–2553.
10 T. Kieser, M. J. Bibb, M. J. Buttner, K. F. Chater and D. A. Hopwood, in *Practical Streptomycete Genetics*, The John Innes Foundation, Norwich, UK, 2000.
11 C. Orlanos, S. J. Moss, A. F. Braña, R. M. Sheridan, V. Math, A. J. Weston, C. Méndez, P. F. Leadley, B. Wilkinson and J. A. Salas, *Mol. Microbiol.*, 2004, **52**, 1745–1756.
12 P. F. Leadlay, unpublished data.
13 R. Fouces, E. Mellado, B. Diez and J. L. Barredo, *Microbiology*, 1999, **145**, 855–868.
14 J. Deveraux, P. Haeberli and O. Smithies, *Nucleic Acids Res.*, 1984, **12**, 387–395; J. F. Aparicio, R. Fouces, M. V. Mendez, N. Olivera and J. F. Martín, *Chem. Biol.*, 2000, **7**, 895–905; M. Carmody, B. Byrne, B. Murphy, C. Breen, S. Lynch, E. Flood, S. Finnan and P. Caffrey, *Gene*, 2004, **343**, 107–115; P. Caffrey, S. Lynch, E. Flood, S. Finnan and M. Oliynyk, *Chem. Biol.*, 2001, **8**, 713–723; T. Brautaset, O. N. Sekurova, H. Sletta, T. E. Ellingsen, A. R. Strom, S. Valla and S. B. Zotchev, *Chem. Biol.*, 2000, **7**, 395–403.
15 K. D. Paull, R. H. Shoemaker, L. Hodes, A. Monks, D. A. Scuderio, L. Rubinstein, J. Plowman and M. R. Boyd, *J. Natl. Cancer Inst.*, 1989, **81**, 1088–1092.

