

Primary and secondary metabolism, and post-translational protein modifications, as portrayed by proteomic analysis of *Streptomyces coelicolor*

A. R. Hesketh,¹ G. Chandra,¹ A. D. Shaw,^{2,3}
J. J. Rowland,³ D. B. Kell,² M. J. Bibb¹ and
K. F. Chater^{1*}

¹Department of Molecular Microbiology, John Innes Centre, Norwich Research Park, Colney, Norwich NR4 7UH, UK.

²Institute of Biological Sciences, and ³Department of Computer Science, University of Wales, Aberystwyth SY23 3DD, UK.

Summary

The newly sequenced genome of *Streptomyces coelicolor* is estimated to encode 7825 theoretical proteins. We have mapped approximately 10% of the theoretical proteome experimentally using two-dimensional gel electrophoresis and matrix-assisted laser desorption ionization time-of-flight (MALDI-TOF) mass spectrometry. Products from 770 different genes were identified, and the types of proteins represented are discussed in terms of their annotated functional classes. An average of 1.2 proteins per gene was observed, indicating extensive post-translational regulation. Examples of modification by N-acetylation, adenylation and proteolytic processing were characterized using mass spectrometry. Proteins from both primary and certain secondary metabolic pathways are strongly represented on the map, and a number of these enzymes were identified at more than one two-dimensional gel location. Post-translational modification mechanisms may therefore play a significant role in the regulation of these pathways. Unexpectedly, one of the enzymes for synthesis of the actinorhodin polyketide antibiotic appears to be located outside the cytoplasmic compartment, within the cell wall matrix. Of 20 gene clusters encoding enzymes characteristic of secondary metabolism, eight are represented on the proteome map, including three that specify the production of novel metabolites.

This information will be valuable in the characterization of the new metabolites.

Introduction

The genome of a plasmid-free derivative, strain M145, of the Gram-positive bacterium *Streptomyces coelicolor* A3(2) has recently been completely sequenced and annotated (Bentley *et al.*, 2002). At 8.67 Mb and 7825 annotated genes, it is nearly twice the size of the *Escherichia coli* (Blattner *et al.*, 1997) and *Bacillus subtilis* (Kunst *et al.*, 1997) genomes, presumably reflecting the lifestyle of the organism. Streptomycetes are mycelial, saprophytic soil bacteria that undergo complex morphological differentiation and produce a range of diverse secondary metabolites with important applications in human medicine and agriculture (for a review, see Champness, 2000). Spore germination is followed by the development of branched hyphae that grow on and into appropriate substrates. In response to complex but still poorly defined signals, the substrate mycelium produces aerial hyphae that eventually undergo septation to yield chains of unigenomic spores. As the aerial branches grow, the substrate mycelium typically begins to produce the various antibiotics. Before the genome sequencing project began, *S. coelicolor* was known to produce four antibiotics (one of them plasmid determined), two of which, actinorhodin (Act) and undecylprodigiosin (Red), are pigmented, and a polyketide spore pigment. It is now apparent that there are about 20 gene clusters that are likely to direct the biosynthesis of what may broadly be considered as secondary metabolites (Bentley *et al.*, 2002). In addition, Bentley *et al.* (2002) predicted more than 800 secreted proteins, 65 sigma factors, 37 Ser/Thr kinase homologues, 85 two-component sensor histidine kinases and a host of transcriptional regulators.

The complete sequencing of an organism's genome immediately allows study of overall gene expression at the levels of mRNA abundance, typically using DNA microarrays (the transcriptome) (Schena, 2000; Lucchini *et al.*, 2001), and protein profiling, by two-dimensional (2D) PAGE coupled with high-throughput matrix-assisted laser desorption ionization time-of-flight (MALDI-TOF) mass

Accepted 15 August, 2002. *For correspondence. E-mail keith.chater@bbsrc.ac.uk; Tel. (+44) 1603 450 297; Fax (+44) 1603 450 778.

spectrometry (the proteome) (Blackstock and Weir, 1999; Mann *et al.*, 2001). Analysis of the proteome is more complicated than transcriptome analysis because of the diverse physical and chemical properties of proteins, and because of the need for any given protein to be extracted in sufficient quantity not only to be detected, but also to be identified. Thompson and co-workers have used 2D-PAGE extensively to analyse changes in the pattern of radiolabelled protein profiles in pulse-chase experiments in *S. coelicolor* during growth and in response to stress, but did not have the benefit of the complete genome sequence to use mass spectrometry to identify proteins of interest (Puglia *et al.*, 1995; Vohradsky *et al.*, 1997; 2000). In this study, we have mapped a substantial part of the *S. coelicolor* proteome using 2D-PAGE and MALDI-TOF peptide mass fingerprint analysis. Particular attention is paid to the representation of proteins involved in secondary metabolic pathways in comparison with those from primary metabolism, and to the extent of regulation by post-translational modification mechanisms.

Results

Analysis of the proteome using 2D-PAGE

Streptomyces coelicolor M145 produces the pigmented antibiotics actinorhodin (Act) and the prodigiosin complex (Red) in a growth phase-dependent manner (Takano *et al.*, 1992; Gramajo *et al.*, 1993). Thus, in liquid minimal medium supplemented with casamino acids (SMM), pigment production begins at the transition between exponential growth and stationary phase. In order to observe proteins from both primary and secondary metabolic pathways, pigmented mycelium from transition phase cultures was harvested and disrupted directly into the strongly denaturing isoelectric focusing buffer UTCHAPS (see *Experimental procedures*). Protein extracts were sepa-

rated by 2D-PAGE using several different immobilized pH gradient (IPG) strips for the first-dimension separation (pH 4–7, pH 6–11, pH 4.5–5.5 and pH 5.5–6.7), 12.5% SDS-PAGE for the second and silver nitrate staining for protein detection (Fig. 1). The proteins visible on each gel were counted after spot detection of scanned gel images using PHORETIX 2D image analysis software, and compared with the total number expected assuming that every potential gene was expressed (Table 1). To calculate the number of spots expected, proteins in the theoretical proteome with predicted isoelectric point values within the pH range of the IPG strip used, and with molecular weights between 8 kDa and 140 kDa (the detectable range using 12.5% PAGE) were counted and multiplied by a factor of 1.2, representing the observed average number of protein spots per gene (see Table 1). Separation using the isoelectric point (pI) range pH 4–7 produced 1051 detectable spots, representing 19% of the total number of proteins predicted for this region. Excluding secreted and membrane proteins, which would not generally be expected in the protein extract, the proportion was 25%. In the pI range pH 6–11, these numbers were 8% for the theoretical total proteome and 12% for the theoretical extractable proteome. Using the narrow-range IPG strips pH 4.5–5.5 and pH 5.5–6.7, 1555 and 913 spots were detected, respectively, corresponding to 82% and 41% of the total number of predicted extractable proteins. Thus, the higher resolution and higher loading possible on these 'zoom gels' approximately doubled the number of proteins detectable.

Identification of separated proteins: towards a proteome map and database

Protein spot identification was performed exclusively on gels stained with colloidal Coomassie G250 and mapped

Table 1. Estimate of the proportion of the *S. coelicolor* proteome that can be detected by silver staining of 2D-PAGE separations of transition phase protein extracts and MALDI-TOF identification of protein spots illustrating the extent of post-translational modifications.

	Isoelectric point range (pH)			
	4–7	4.5–5.5	5.5–6.7	6–11
Protein spots detected ^a	1051	1555	913	379
Percentage of expected protein spots detected ^b	19%	63%	37%	8%
Percentage of expected extractable protein spots detected ^c	25%	82%	41%	12%
Protein spots identified ^d	443	459	193	210
Different gene products represented	374	360	170	179
Ratio of protein spots per gene product	1.18	1.27	1.13	1.17

a. On silver-stained gels.

b. The total number of gene products with theoretical isoelectric point values and molecular weights within the range of the gel separation (8–140 kDa) was multiplied by 1.2, the observed average number of protein spots per gene as a result of post-translational modifications.

c. Same as a, but excluding proteins annotated as being membrane or secreted.

d. From colloidal Coomassie-stained gels.

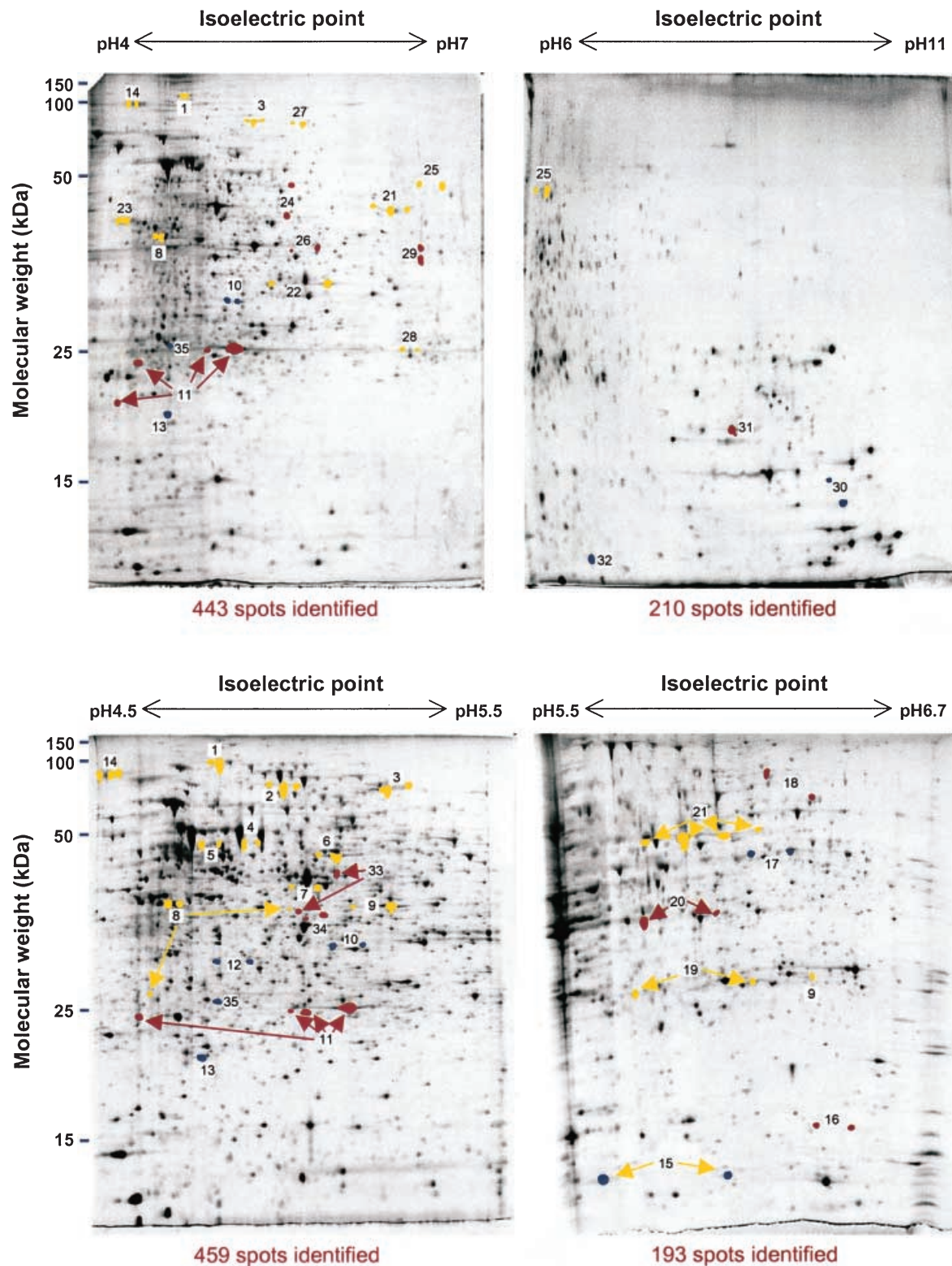


Fig. 1. A proteome map of *S. coelicolor* showing 2D-PAGE separations using IPG strips pH 4–7, pH 6–11, pH 4.5–5.5 and pH 5.5–6.7. The number of spots identified is given at the foot of each reference gel. Protein spots from metabolic pathways that were identified at positions inconsistent with their predicted amino acid sequences, and suggestive of post-translational modification, are marked in yellow (primary metabolism) or red (secondary metabolism). Blue spots correspond to other examples of post-translational modifications that are mentioned in the text and in Table 3. Separations have been performed on numerous occasions and generally show high spatial reproducibility.

Table 2. Types of proteins identified on the proteome map.

Protein class ^a	Number in genome ^a	Number identified ^b
Hypothetical	1049	68 (6%)
Conserved hypothetical	1322	114 (8%)
Secreted	530	30 (5%)
ABC transport ATP binding	109	16 (14%)
Lipoproteins	124	10 (8%)
Sigma factors	65	0 (0%)
Antisigma factors	7	2 (28%)
Antisigma factor antagonists	15	1 (6%)
Two-component system regulators	80	11 (13%)
Other regulators	592	36 (6%)
Two-component system kinases	85	0 (0%)
Ser/Thr kinases	37	1 (2%)
Other	3923	482 (12%)
All	7816	770 (10%)

a. Based on annotation of the *S. coelicolor* genome sequence (http://www.sanger.ac.uk/Projects/S_coelicolor/).

b. Percentage identified is included in parentheses.

back onto the analytical silver-stained gels in Fig. 1. Although silver staining is 5–10 times more sensitive than colloidal Coomassie staining (and about 100 times more sensitive than conventional Coomassie staining), further mass spectrometry analysis of protein spots is much more efficient after staining with colloidal Coomassie than with silver. A total of 1305 protein spots were unambiguously assigned to sequenced genes using MALDI-TOF peptide mass fingerprint analysis, at a success rate of ~90% (Table 1). There is some redundancy in these figures, with several spots being represented in more than one pI range. On each gel, some proteins were identified in more than one position, indicative of post-translational modifications. The average number of protein spots per gene product was 1.2. The products from 770 different genes were assigned to the proteome map, representing ~10% of the genome and corresponding to many different classes of proteins (Table 2). The master gel reference maps are the basis for the production of a publicly accessible proteome database (http://qbab.aber.ac.uk/S_coeli/ReferenceGel/refgel.html, see last Results section).

Analysis of co- and post-translational modifications

The results of MALDI-TOF identification of protein spots summarized in Table 1 indicated a significant amount of regulation at the post-transcriptional level in *S. coelicolor*, as in most other bacteria. Table 3 lists those gene products that were identified at pI and/or molecular weight coordinates significantly different from predicted values. In general, these fell into two categories: one or more spots with lower than expected molecular weights and often different pI values, indicating modification by proteolytic processing (Table 3A), and multiple spots with the same apparent molecular weights but different pI values, indi-

catating modification by covalent addition of a small adduct (Table 3B). In the former case, the full-length protein was often, but not always, detected. A similar fraction of proteins of unknown function and annotated as 'hypothetical' or 'conserved hypothetical' was subject to post-translational modification, but these will be described in a later publication and are not considered in Table 3. Evidence in the peptide mass fingerprint data for the specific nature of the modification was obtained in 10 out of the 88 proteins listed (Table 4). Figure 2 illustrates the characterization of an example of each type of modification by interpretation of the MALDI-TOF data.

N-acetylation. The putative cellulose-binding protein SCO5396 was identified in two positions on the proteome map (spots labelled 10 in Fig. 1) with the same apparent molecular weight but significantly different pI values. The more basic spot gave a tryptic peptide at 1370.76 Da corresponding to the N-terminal peptide 2–13, indicating that the initiator Met residue had been co-translationally removed (Fig. 2A, top). In the more acidic spot (i.e. on the left in Fig. 1), this peptide had increased in mass by 42.0 Da, diagnostic of the addition of an acetyl group, presumably on the N-terminal Ser residue (Fig. 2A, bottom). A similar modification was identified for the 20S proteasome alpha subunit (spots 12 in Fig. 1) and for the putative ABC transport protein BldK-ORFD (spots 17 in Fig. 1), although the presumptive N-terminal acetylated residue is Thr in the latter (data not shown).

Adenylylation. By analogy with the nitrogen regulatory system in *E. coli*, glutamine synthetase I (GSI) was already believed to be modified by adenylylation on a conserved Tyr-397 residue, and biochemical evidence for this has been reported previously (Fink *et al.*, 1999). The MALDI-TOF data showed the tryptic peptide 395–420 at the expected mass of 2807.4 Da for the larger, more basic spot 6 in Fig. 1, but this increased by 329.0 Da, corresponding to the addition of an adenylyl group, in the smaller spot on the left (Fig. 2B). Adenylylation of the nitrogen regulatory protein GlnK (spot 15 in Fig. 1) was identified similarly, but this is described in detail elsewhere (Hesketh *et al.*, 2002).

Proteolytic processing. The 50S ribosomal protein L9 was identified as two spots that differed in both pI and molecular weight (spots 30 in Fig. 1), suggesting modification by proteolytic processing. The spot with the higher molecular weight showed the tryptic peptide corresponding to amino acids 3–22 at the expected mass of 1993.10 Da, in agreement with this being the unprocessed form of the protein (Fig. 2C, top). In the lower of the two spots, the peptide decreased in mass by 113.12 Da (Fig. 2C, bottom), corresponding to the loss of

Table 3. Proteins identified at pI and molecular weight co-ordinates on the 2D-PAGE maps that are inconsistent with their predicted amino acid sequences, indicating possible post-translational modification either by proteolytic processing (A) or by covalent addition of a low-molecular-weight adduct (B).

ORF no.	Gene product	Spot no. in Fig. 1
A		
SCO0498	Probable peptide monooxygenase	33
SCO1230	Putative tripeptidylaminopeptidase	
SCO1644	20S proteasome beta subunit precursor	35
SCO1815	3-Oxoacyl-[acyl carrier protein] reductase	
SCO1860	Putative secreted protein	
SCO1947	Glyceraldehyde-3-phosphate dehydrogenase	9
SCO1998	30S ribosomal protein S1	
SCO2619	ATP-dependent Clp protease subunit	13
SCO2920	Probable secreted proteinase	
SCO3236	Probable oxygenase	34
SCO3671	Heat shock protein DNAK	
SCO3909	50S ribosomal protein L9	30
SCO4296	Chaperonin GroEL2	
SCO4439	D-Alanyl-D-alanine carboxypeptidase	
SCO4643	UDP-N-acetylanoylpyruvoylglucosamine reductase	
SCO4662	Elongation factor Tu1	
SCO4762	Chaperonin GroEL1	
SCO4808	Succinyl-CoA synthetase beta chain	8
SCO4824	Bifunctional protein methylene tetra-hydrofolate dehydrogenase/cyclohydrolase	
SCO5073	Putative oxidoreductase ActVI-ORF2	29
SCO5074	Putative dehydratase ActVI-ORF3	31
SCO5088	Ketoacyl synthase beta subunit ActI-ORF2	24
SCO5113	Oligopeptide ABC transport protein BldK-ORFB	
SCO5254	Superoxide dismutase precursor	32
SCO5371	ATP synthase alpha chain	
SCO5477	Probable oligopeptide-binding lipoprotein	
SCO5624	30S ribosomal protein S2	
SCO5737	Guanosine pentaphosphate synthetase	
SCO5769	Recombinase RecA	
SCO5892	Polyketide synthase RedL	18
SCO6282	3-Oxoacyl-[acyl carrier protein] reductase	11
B		
SCO1074	Putative peptidase	
SCO1476	S-adenosylmethionine synthetase	
SCO1594	Phenyl-tRNA synthetase beta chain	
SCO1643	20S proteasome alpha subunit	12
SCO1814	Enoyl-[acyl carrier protein] reductase	
SCO1935	Transketolase	3
SCO1947	Glyceraldehyde-3-phosphate dehydrogenase	9
SCO2180	Dihydrolipoamide dehydrogenase	21
SCO2181	Probable dihydrolipoamide S-succinyltransferase	14
SCO2198	Glutamine synthetase I	6
SCO2390	3-Oxoacyl-[acyl carrier protein] synthase II	
SCO2478	Putative reductase	
SCO2618	ATP-dependent Clp protease subunit 2	
SCO2822	Putative decarboxylase	
SCO3096	Enolase	23
SCO3229	4-Hydroxyphenylpyruvic acid dioxygenase	
SCO3337	Probable pyrroline-5-carboxylate reductase	28
SCO3409	Inorganic pyrophosphatase	
SCO3549	Antisigma factor antagonist BldG	
SCO3563	Acetyl-CoA synthase	
SCO3671	Heat shock protein DNAK	
SCO3878	DNA polymerase III beta chain	
SCO3890	Thioredoxin reductase	
SCO3906	30S ribosomal protein S6	
SCO3970	XAA-Pro aminopeptidase I	
SCO4164	Putative thiosulphate sulphurtransferase	
SCO4209	Phosphoglycerate mutase	19
SCO4240	ABC transport ATP binding protein MsiK	
SCO4296	Chaperonin GroEL2	
SCO4645	Aspartate aminotransferase	7
SCO4652	50S ribosomal protein L10	

Table 3. cont.

ORF no.	Gene product	Spot no. in Fig. 1
SCO4661	Elongation factor G	
SCO4662	Elongation factor Tu1	
SCO4677	Putative regulatory protein	
SCO4729	DNA-directed RNA polymerase alpha chain	
SCO4762	Chaperonin GroEL1	
SCO4770	Inosine-5'-monophosphate dehydrogenase	25
SCO4792	LuxR family two-component regulator	
SCO4808	Succinyl-CoA synthetase beta chain	8
SCO4809	Succinyl-CoA synthetase alpha chain	22
SCO4921	Probable acyl-CoA carboxylase complex A chain	
SCO5071	Hydroxylacyl-CoA dehydrogenase ActVI-ORFA	16
SCO5079	Hypothetical protein ActVA-ORF4	20
SCO5080	Possible hydrolase ActVA-ORF5	26
SCO5115	Probable ABC transporter intracellular ATPase chain BldK-ORFD	17
SCO5371	ATP synthase alpha chain	
SCO5396	Putative cellulose-binding protein	10
SCO5419	Putative thioredoxin	
SCO5515	Probable D-3-phosphoglycerate dehydrogenase	4
SCO5535	Putative carboxyl transferase	
SCO5584	Nitrogen regulatory protein GlnK	15
SCO5625	Elongation factor Ts	
SCO5699	Prolyl-tRNA synthetase	
SCO5737	Guanosine pentaphosphate synthetase	
SCO5777	Glutamate uptake system ATP-binding protein	
SCO5803	SOS regulatory protein LexA	
SCO5999	Aconitase	1
SCO6198	Probable secreted protein	
SCO6282	3-Oxoacyl-[acyl carrier protein] reductase	11
SCO6636	Bacteriophage ϕ -C31 resistance protein PglZ	
SCO6663	Transketolase	27
SCO7000	Isocitrate dehydrogenase	2
SCO7036	Argininosuccinate synthase	5
SCO7510	Peptidyl-prolyl <i>cis-trans</i> isomerase	

Some proteins appear to be subject to both types of modification and are present in both parts. Only proteins from primary or secondary metabolic pathways, or with identified modifications, are assigned to spots in Fig. 1.

the isoleucine residue at its N-terminus (expected mass loss 113.08 Da). This indicates removal of the N-terminal MKI residues from ribosomal protein L9. The theoretical isoelectric point (9.97) and molecular weight (15.6 kDa) of L9 protein N-terminally truncated in this way are in good agreement with the position of the lower spot in Fig. 1 (predicted values for the unprocessed protein are pH 9.60 and 15.9 kDa). MALDI-TOF data indicating proteolytic

processing of the superoxide dismutase precursor protein (SCO5254; spot 32 in Fig. 1) and the 20S proteasome beta subunit precursor (SCO1644; spot 35 in Fig. 1) agreed with previous reports (Kim *et al.*, 1998; Nagy *et al.*, 1998). Evidence for the proteolysis of *S. coelicolor* ClpP1 (SCO2619; spot 13 in Fig. 1) has also been reported, although the processing site was not defined (de Crecy-Lagard *et al.*, 1999).

Table 4. Co- and post-translational modification of proteins suggested by MALDI-TOF peptide mass fingerprint analysis.

Protein	Modification
SCO1643	20S proteasome alpha subunit
SCO5115	ABC transporter intracellular ATPase BldK-ORFD
SCO5396	Putative cellulose-binding protein
SCO2198	Glutamine synthetase I
SCO5584	Nitrogen regulatory protein GlnK
SCO1644	20S proteasome β -subunit precursor
SCO2619	ATP-dependent Clp protease proteolytic subunit I
SCO3909	50S ribosomal protein L9
SCO5074	Putative dehydratase ActVI-ORF3
SCO5254	Superoxide dismutase precursor
	Removal of initiator Met residue and acetylation of N-terminal Ser
	Removal of initiator Met residue and acetylation of N-terminal Thr
	Removal of initiator Met residue and acetylation of N-terminal Ser
	Adenylation of peptide 395–420
	Adenylation of peptide 48–58
	Proteolytic processing between amino acid residues 53/54
	Proteolytic processing between amino acid residues 22/23
	Proteolytic processing between amino acid residues 3/4
	Proteolytic processing between amino acid residues 31/32 and 34/35
	Proteolytic processing between amino acid residues 14/15

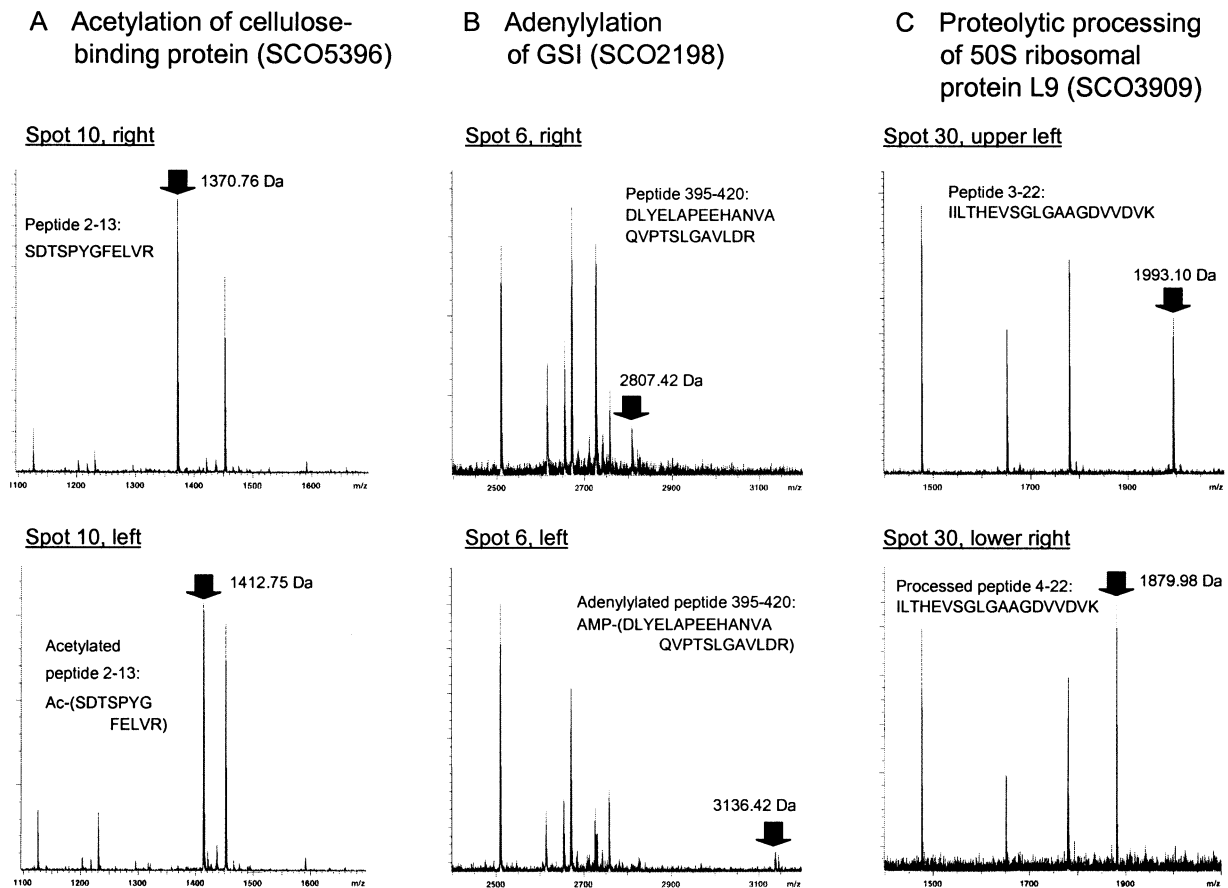


Fig. 2. Examples of MALDI-TOF peptide mass fingerprint data identifying post-translational modification of proteins by (A) acetylation (predicted mass increase of 42.0 Da), (B) adenylylation (predicted mass increase of 329.0 Da) and (C) proteolytic processing (predicted mass loss for Ile of 113.08 Da). In each fingerprint, the x-axis is mass/charge ratio (m/z), and the y-axis is absolute intensity. The observed mass accuracies in the MALDI-TOF fingerprints are: (A) 60 p.p.m.; (B) 20 p.p.m.; and (C) 20 p.p.m.

Co-translational modification by removal of the N-terminal initiator Met occurred in 149 out of 190 (78%) cases in which the N-terminal peptide was detected. Processing by the N-methionylaminopeptidase was observed if the second amino acid was Ser (45 out of 149 examples), Thr (40), Ala (32), Pro (27), Gly (4) or Val (1). On only one occasion was any of these six amino acids found as the second residue if the N-terminal methionine was retained. This is in broad agreement with observations in *B. subtilis* in which two-thirds of proteins were reported to be similarly processed (Buttner *et al.*, 2001)

Representation of primary and secondary metabolic pathways

Streptomycetes produce $\approx 60\%$ of all commercially useful antibiotics (Berdy, 1984; Miyodah, 1993). Most antibiotics are the product of complex biosynthetic pathways, often encoded by 20–30 clustered genes. Production of the pigmented antibiotics actinorhodin (Act) and undecyl-

prodigiosin (Red) has been studied extensively in *S. coelicolor* (for reviews, see Chater and Bibb, 1997; Champness, 2000) and involves clusters of approximately 22 genes each. A third cluster producing the calcium-dependent antibiotic (CDA) and consisting of 40 genes has also been studied (Chong *et al.*, 1998; Huang *et al.*, 2001). In the extensive spot identification studies for the preparation of a proteome map of *S. coelicolor* (see above), more than one-third of the proteins from each of these pathways was identified (Table 5). Proteins from three clusters that specify metabolic products that have yet to be identified experimentally (Bentley *et al.*, 2002) were also extensively represented: a type I polyketide synthase cluster (SCO6273–6288), a deoxysugar/glycosyltransferase cluster (SCO0381–0401) and a non-ribosomal peptide synthetase cluster predicted to encode a novel siderophore, coelichelin (SCO0489–0499). Several clusters were unrepresented.

The primary metabolic pathways were somewhat more highly represented, with $\approx 60\%$ or more of proteins anno-

Table 5. Summary of proteins identified on the proteome map from secondary metabolic pathways during growth in SMM.

Biosynthetic pathway/gene cluster	Pathway genes annotated	Pathway gene products identified	Percentage pathway detected
Calcium-dependent antibiotic CDA, SCO3210–3249	40	16	40
Actinorhodin, SCO5071–5092	22	15	68
Undecylprodigiosin, SCO5877–5898	22	8	36
Deoxysugar synthase/glycosyltransferase cluster, SCO0381–0401	21	7	33
Type I polyketide synthase, SCO6273–6288	16	6	38
Hopanoids, SCO6759–6771	13	1	7
Siderophore (coelichelin), SCO0489–0499	11	5	45
Siderophore (desferrioxamine), SCO2783–2785	4	4	100

Chromosomal locations and extent of gene clusters are taken from Bentley *et al.* (2002). No proteins were detected from the remaining predicted secondary metabolic clusters, i.e. siderophore (coelibactin), SCO7681–7691 (11 ORFs); siderophore, SCO5799–5801 (three ORFs); non-ribosomal peptide synthetases, SCO6429–6438 (10 ORFs); type II fatty acid synthetases, SCO1265–1273 (nine ORFs); WhiE spore pigment cluster, SCO5314–5320 (seven ORFs); isorenieratone, SCO0185–0191 (seven ORFs); eicosapentaenoic acid, SCO0124–0129 (six ORFs); chalcone synthetases, SCO1206–1208 (three ORFs); chalcone synthetases, SCO7669–7671 and SCO7222 (three ORFs); sesquiterpene cyclase, SCO5222–5223 (two ORFs); type I polyketide synthetases, SCO6826–6827 (two ORFs); geosmin, SCO6703 (one ORF).

tated as enzymes of glycolysis, the TCA cycle, the pentose phosphate pathway, purine ribonucleotide biosynthesis and pyrimidine nucleotide biosynthesis being mapped (Table 6). About 50% of amino acid biosynthetic proteins were identified. Lists of individual members of all pathways considered here, including those that were and were not identified during proteome mapping, can be found in the *Supplementary material*.

Several proteins from both primary and secondary metabolic pathways were identified at more than one position on the proteome map, and may therefore be subject to regulation by post-translational modification (see Table 3 and Fig. 1). Primary metabolic proteins of this type (yellow spots on Fig. 1) were usually detected as chains of two or more spots with the same apparent molecular weights but different pI values, suggesting covalent modification by low-molecular-weight adducts. Apart from adenylation of GSI (see above), the nature of the modification of these proteins was not revealed by MALDI-TOF peptide mass fingerprint data. Both covalent modification and proteolytic processing were also found in secondary metabolic proteins (red spots in Fig. 1). For example, the Act pathway members SCO5071 (spot 16) and SCO5075 (spot 20)

and the putative 3-oxoacyl[acyl carrier protein] reductase from the type I polyketide cluster (SCO6282; spot 11) appear to be covalently modified, whereas proteolytic processing was detected for RedL (SCO5892; spot 18) from the undecylprodigiosin cluster, the putative dehydratase ActVI-ORF3 (SCO5074; spot 31) from the Act cluster, the putative peptide monooxygenase from coelichelin biosynthesis (SCO0498; spot 34) and the putative oxygenase from the CDA pathway (SCO3236; spot 33). Two Act biosynthetic proteins, SCO5088 (spot 24) and SCO5073 (spot 29), were detected both at their predicted pI and molecular weight co-ordinates and at positions with similar pI values but significantly higher molecular weights. This may indicate that these proteins form stable multimers that are not completely denatured under the conditions used. Only the modification of ActVI-ORF3 (SCO5074; spot 31) could be characterized further from the mass spectrometry data, which showed that it involves proteolytic processing at two sites between amino acid residues 31/32 and 34/35 (data not shown). Analysis of a sample obtained by externally extracting proteins from intact washed mycelium with a solution containing 2% SDS and 50 mM dithiothreitol (DTT) revealed that, sur-

Table 6. Summary of proteins identified on the proteome map from primary metabolic pathways during growth in SMM.

Biosynthetic pathway	Pathway genes annotated ^a	Pathway gene products identified	Percentage pathway detected
Glycolysis ^b	24	13	55
Pentose phosphate pathway ^b	14	10	71
TCA cycle ^b	21	12	57
Amino acid biosynthesis ^c	115	55	48
Purine ribonucleotide biosynthesis ^b	18	12	67
Pyrimidine ribonucleotide biosynthesis ^b	8	5	62

a. Derived from interpretation of the annotated *S. coelicolor* genome sequence at http://www.sanger.ac.uk/Projects/S_coelicolor/.

b. See *Supplementary material* for details of pathway members.

c. See *Supplementary material* for details of individual pathways.

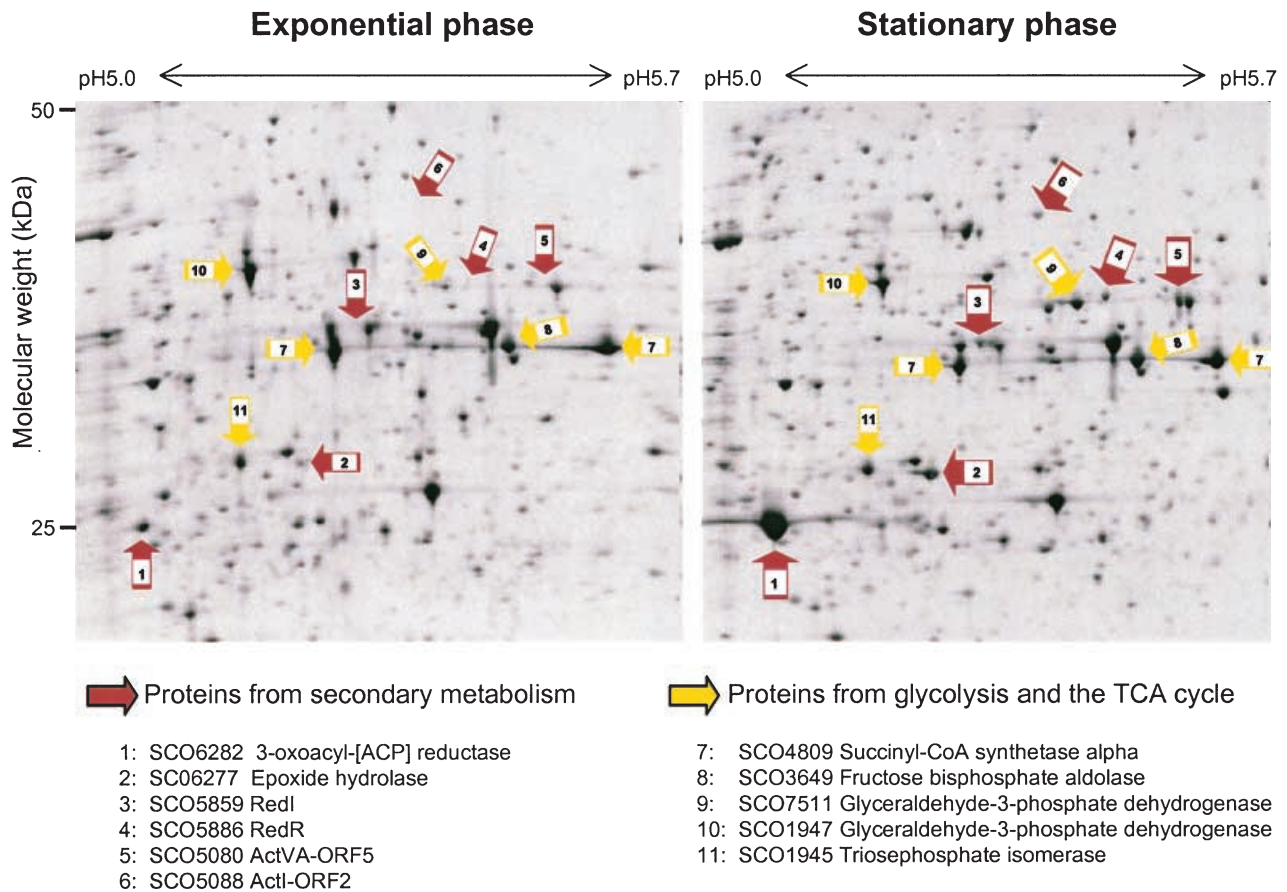


Fig. 3. A comparison of the expression of certain primary and secondary metabolic proteins between exponentially growing and stationary phase cultures of *S. coelicolor* M145. Larger arrows represent higher protein level in one gel than in the other. The experiment was performed on duplicate biological samples, and both showed similar protein expression profiles.

prisingly, the truncated ActVI-ORF3 is located extracellularly, presumably non-covalently associated with the cell wall matrix (Fig. 4).

Expression analysis of certain primary and secondary metabolic proteins

Production of the pigmented antibiotics Act and Red in liquid minimal medium supplemented with casamino acids (SMM) begins in the transition phase between exponential growth and stationary phase. Figure 3 shows a comparison of the expression of certain primary and secondary metabolic proteins between exponentially growing and stationary phase cultures. Two proteins from the Act cluster (ActVA-ORF5 and ActI-ORF2) and two from the Red (RedI and RedR) were present in the stationary phase cultures but were not detected during exponential growth. Similarly, the putative 3-oxoacyl-[acyl carrier protein] reductase (SCO6282) and epoxide hydrolase (SCO6277) enzymes from a type I polyketide secondary metabolic cluster showed a large increase in abundance

between the exponential and stationary phase samples. Triose phosphate isomerase (SCO1945), fructose bisphosphate aldolase (SCO3649) and succinyl-CoA synthetase alpha chain (SCO4809), all enzymes from central carbon metabolism, were present in approximately constant amounts in both samples. Interestingly, two isoforms (SCO7511 and SCO1947) of glyceraldehyde-3-phosphate dehydrogenase from the glycolytic pathway showed different patterns of expression. SCO1947 levels were the same in both exponential and stationary phase, but SCO7511, barely detectable in exponentially growing cultures, increased markedly in stationary phase.

The Streptomyces coelicolor online proteome map

The 2D gel images in Fig. 1, together with the spot identification results, are being made available over the World Wide Web to provide a reference map of the *S. coelicolor* proteome. At the time of writing, the reference gel for the pH 4–7 isoelectric point range was fully accessible, whereas those for pH 4.5–5.5, pH 5.5–6.7 and pH 6–11

were under construction (http://qbab.aber.ac.uk/S_coeli/ReferenceGel/refgel.html). A tool for comparing two or more gel images with each other is also accessible from the same website at http://qbab.aber.ac.uk/S_coeli/flicker.html.

Discussion

Silver staining is one of the most sensitive methods for detecting proteins on polyacrylamide gels, allowing as little as 0.5–2 ng of an individual protein to be detected (Berggren *et al.*, 2000; Gorg *et al.*, 2000). The proportion of the *S. coelicolor* proteome that can be analysed using silver staining of 2D-PAGE-separated proteins depends, at least in part, on the pH range of the IPG strip used for the first-dimension separation (see Table 1). Using the intermediate-range IPG strip covering pH 4–7, about 20–25% of the proteins in the theoretical proteome predicted to be observable in the resulting 2D-PAGE gel were actually detected. This figure was reduced to 10–12% when analysing the more basic pI range using the pH 6–11 IPG strip. Attempts to increase the percentage of the proteome detectable using the pH 4–7 and pH 6–11 IPG strips by loading more protein sample resulted in poorer separations (data not shown). This was especially true of the pH 6–11 strips, which were generally the most difficult to use successfully. However, good separation of higher protein loads was achieved using the narrow-range IPG strips pH 4.5–5.5 and 5.5–6.7, significantly increasing the proportion of theoretical proteins detected to over 40%. Clearly, and more pertinently, the limited conditions used for culturing *S. coelicolor* in this study will also have restricted the number of proteins detected, as not all the proteome will be expressed. In particular, *S. coelicolor* does not sporulate in shaken liquid-grown cultures, so proteins involved in this differentiation process are unlikely to be present. Nevertheless, it is striking that the combined data for the pH 4.5–5.5 and 5.5–6.7 isoelectric point ranges indicate that 2468 out of a total of predicted 4932 protein spots can be detected (see Table 1), suggesting that transition phase mycelium in this supplemented minimal medium is expressing 50% of its genes (i.e. by extrapolation, some 3912 of the 7825 in the genome).

Analysis of post-translational regulation

Although the regulation of cellular processes and adaptive responses in bacteria has been studied extensively at the transcriptional level, post-translational regulatory mechanisms are much less understood. Proteomic analysis using 2D-PAGE coupled with mass spectrometry offers the opportunity to observe and identify modification of proteins at a general cellular level. The functions of proteins can be altered by specific covalent attachment of

adducts (e.g. phosphorylation, glycosylation, nucleotidylation, acetylation) or by proteolytic processing. Both events alter the pI and/or mass of the protein and so can usually be observed on 2D-PAGE. In the *S. coelicolor* proteome map, ≈ 110 out of a total of 770 gene products were identified in more than one position, or at pI and apparent mass co-ordinates significantly different from predicted values, indicating extensive regulation at the post-translational level (88 are listed in Table 3): a proportion similar to that reported for other prokaryotes (Buttner *et al.*, 2001; Tonella *et al.*, 2001). The initial peptide mass fingerprint analysis identified the modification of only 10 of these proteins, probably partly because fingerprint data typically cover ≈ 30 –60% of a protein's sequence (data not shown), so modification at positions not represented will be missed. Protein digestion with endopeptidases other than trypsin before mass spectrometry may help in this situation, but some covalently attached groups are labile and cannot be detected routinely. Phosphorylation of His occurs on nitrogen, producing a high-energy phosphoramidate bond that is easily hydrolysed in acidic conditions, whereas phospho-Asp is the most labile phosphorylated residue, with a half-life of a few hours under physiological conditions (reviewed by Robinson *et al.*, 2000). These modifications, the basis of bacterial two-component regulatory systems, are therefore unlikely to survive staining of the 2D gels, which involved the use of a fixing solution containing 10% acetic acid, and mass spectrometry. Identification of the more stable phosphoester phosphorylation that occurs on Ser, Thr or Tyr residues using PAGE followed by mass spectrometry has been reported (Godovac-Zimmermann *et al.*, 1999; Buttner *et al.*, 2001), but no examples were detected in this work. Of the modifications identified in this study, three were by acetylation at the N-terminus after removal of the initiator Met residue, two involved adenylation, and five proteins had been N-terminally truncated. Modification of the 20S proteasome by N-acetylation in yeast involves removal of the Met residue by N-methionylaminopeptidase, followed by acetylation with an N^α-acetyltransferase (Kimura *et al.*, 2000). Homologues of these enzymes are predicted from the *S. coelicolor* genome and are presumably responsible for the observed modification of the 20S proteasome alpha subunit, and of BldK-ORFD and the putative cellulose-binding protein. The functional significance of N^α-acetylation of these proteins is unclear.

Analysis of the biosynthesis of antibiotics and other secondary metabolites

Post-translational modification. The proteome mapping results indicate that at least six enzymes in the Act cluster are subject to post-translational modifications, and one

each from the undecylprodigiosin (SCO5877–5898), ‘coelichelin’ (SCO0489–0499) and type I polyketide synthase (SCO6273–6288) clusters (see Table 3). These metabolic pathways may therefore be significantly regulated at the post-translational level. The *actVI-ORF3* gene (SCO5074) encodes a putative dehydratase enzyme that is part of the biosynthetic pathway for Act, a blue member of the benzoisochromanequinone (BIQ) family of pigmented antibiotics. Mutation in *actVI-ORF3* causes a 50% decrease in the accumulation of Act (and a reddish brown phenotype), and it is believed to be responsible for pyran ring closure, an intermediate step in Act biosynthesis (that may also occur spontaneously), leading to the formation of the BIQ chromophore (Fernandez-Moreno *et al.*, 1994; Ichinose *et al.*, 1998a; Taguchi *et al.*, 2000). In this study, ActVI-ORF3 protein was detected in two forms: a major spot corresponding to ActVI-ORF3 truncated by removal of the N-terminal 31 amino acid residues, and a minor one in which the first 34 amino acids were missing (spots 31 in Fig. 1). The full-length protein was not detected. Analysis of the protein sequence using the SIGNALP server at <http://www.cbs.dtu.dk/services/SignalP/> predicted a signal peptide cleavage site between amino acids 31 and 32, in agreement with the observed results.

The protein extract analysed was prepared from mycelium that had been collected by centrifugation and subjected to two washes, indicating that, if ActVI-ORF3 is exported, it is effectively retained within the cell wall matrix. This was confirmed by analysis of a preparation obtained by externally extracting proteins from intact washed mycelium (Fig. 4). Export of an enzyme required for the efficient biosynthesis of an antibiotic that is thought to be produced intracellularly appears to be anomalous, particularly as it is involved in an intermediate biosynthetic

event. Database analysis of the other members of the Act cluster revealed only one other protein with a predicted signal peptide sequence, the putative integral membrane transport-related protein ActII-ORF3, and it seems unlikely that all the Act biosynthetic steps following BIQ chromophore formation are extracellular. It is possible that full-length ActVI-ORF3 catalyses the intermediate biosynthetic event intracellularly, and the exported processed protein is responsible for a different extracellular activity, perhaps later in the pathway. However, no full-length protein was detected, so the exported enzyme may perform the pyran ring closure on a secreted metabolite that is then reabsorbed by the cell for completion of the subsequent biosynthetic steps. Only two homologues of *actVI-ORF3* are known, and both are also in BIQ antibiotic biosynthetic clusters: *gra-ORF18* from the granaticin cluster (Ichinose *et al.*, 1998b), and *med-ORF5* from medermycin biosynthesis (K. Ichinose, personal communication). Interestingly, the encoded proteins of both sequences are also predicted to contain N-terminal signal peptides. The observation that ActVI-ORF3 is exported to the cell wall matrix may therefore be of general significance to the BIQ family of antibiotics, and provide a new insight into their biosynthetic pathways.

Pathway representation. Of the 20 putative secondary metabolite gene clusters identified in the *S. coelicolor* genome sequence (Bentley *et al.*, 2002), eight had at least one protein each identified on the proteome map (see Table 5). The biosynthetic clusters for the antibiotics Act, Red and CDA had 68%, 36% and 40% of their proteins represented. There are several possible explanations for failure to detect all the proteins for each pathway. Some are putative integral membrane proteins, so cannot readily be detected using conventional 2D-PAGE. Some Red and CDA proteins have molecular weights >200 kDa, outside the range of the 12.5% gels used (for cluster members, see *Supplementary material*). The pathway-specific regulatory proteins for Act and Red biosynthesis were not detected despite a detailed search of their predicted gel locations, even though the relevant genes are known to be transcribed in equivalent cultures of the same strain (Hesketh *et al.*, 2001). The proteins are presumably present but at levels below the limit of detection. However, the transcriptional regulator for CDA production, CdaR, was identified, as well as the two-component system response regulator AbsA2, which is also encoded by a gene in the CDA cluster (Anderson *et al.*, 2001). Interestingly, proteins from three putative secondary metabolite clusters producing chemicals of as yet unknown structure were also identified on the map, corresponding to the deoxysugar synthase/glycosyltransferase cluster SCO0381–0401, the ‘coelichelin’ cluster for a predicted (but not yet confirmed) oligopeptide siderophore

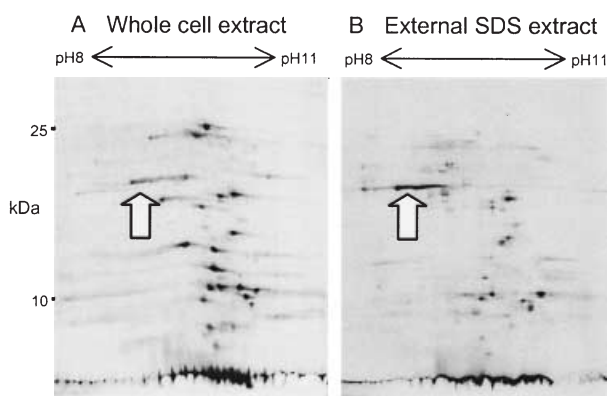


Fig. 4. Processed ActVI-ORF3 is present in the cell wall matrix. 2D-gel separations of protein extracts prepared (A) from whole cells or (B) by externally washing cells in a 2% SDS extraction buffer. The arrow indicates ActVI-ORF3 protein. The experiment was performed on duplicate biological samples, and both showed similar protein expression profiles.

(SCO0489–0499) and the type I polyketide cluster SCO6273–6288. The 3-oxoacyl-[acyl carrier protein] reductase (SCO6282) from the latter pathway is expressed in a growth phase-dependent manner (Fig. 3) and, in early stationary phase cultures, is perhaps the most abundant protein spot in the entire detectable proteome. Presumably the metabolites produced by these three pathways are being synthesized under the conditions used. In addition to coelichelin, three other siderophores are postulated for *S. coelicolor*. All the proteins from one of these, desferrioxamine (SCO2782–2785) were identified, but none was seen from either of the other two pathways. The WhiE polyketide spore pigment is associated with sporulation of surface-grown cultures (Davis and Chater, 1990), and the expression of the *whiE* genes depends on a series of sporulation-specific regulatory genes (Kelemen *et al.*, 1998), no corresponding proteins of which were detected here. It is therefore not surprising that WhiE proteins were not detected in samples prepared from liquid-grown vegetative mycelium.

Analysis of primary metabolism

Proteins annotated as functioning in primary metabolic pathways were highly represented on the proteome map with $\approx 60\%$ or more of the enzymes assigned to glycolysis, the TCA cycle, the pentose phosphate pathway and purine and pyrimidine ribonucleotide biosynthesis being identified (see Table 6). This figure may be deceptively low because these pathways in *S. coelicolor* often appear to have more than one enzyme isoform for some of the metabolic steps, potentially offering additional levels of control or developmental responsiveness at these points. For example, only two steps in glycolysis have just a single gene product assigned to them; the remainder typically have two or three each (Fig. 5). At least one isoform was present for each step, so the pathway is in theory 100% functionally represented. Understanding the regulation of these isoenzymes under different nutritional conditions or different developmental stages. Figure 5 shows that, of the three predicted glyceraldehyde-3-phosphate dehydrogenase isoforms from glycolysis, only two were detected in this study (see also *Supplementary material*). Of these, SCO1947 appears to be constitutive, whereas SCO7511 is growth phase dependent (see Fig. 3). Thus, even the limited conditions in this study yielded useful information about isoenzyme expression patterns (see *Supplementary material*). In *Streptomyces aureofaciens*, two genes encoding glyceraldehyde-3-phosphate dehydrogenase activity also appear to be differently regulated: transcription of one, highly homologous to SCO7511, is activated by the regulator GapR, whereas the other is GapR independent (Spusansky *et al.*, 2001).

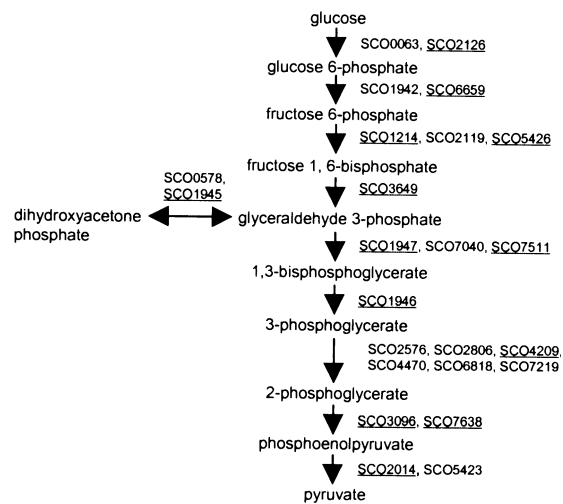


Fig. 5. The glycolytic pathway in *S. coelicolor* illustrating the assignment of multiple gene products to individual metabolic steps, based on the annotation of the genome sequence at http://www.sanger.ac.uk/Projects/S_coelicolor/. Gene products identified in the proteome map are underlined.

Hood *et al.* (1992) reported that most amino acid biosynthetic enzymes are expressed at very low constitutive levels in *S. coelicolor*, so it is reassuring to note that $\approx 50\%$ could be detected and identified using proteomics. These include TrpB and TrpC1 from the tryptophan biosynthetic pathway, the corresponding transcripts of which could barely be detected in previously reported S1 nuclease protection experiments (Hu *et al.*, 1999).

Perspectives

No attempt was made in the work reported here to simplify the protein separations by subdividing the sample into soluble and membrane fractions, and this is reflected in the 10 lipoproteins and 16 ABC transport system ATP-binding subunits identified on the proteome map (see Table 2). No integral membrane proteins were detected and, even when a membrane protein sample was prepared and analysed, no proteins identified possessed more than three predicted transmembrane helices (data not shown). This is consistent with previous studies in other bacteria (Molloy *et al.*, 1999; Nilsson and Davidsson, 2000). Another group of proteins also strikingly absent from the list of those identified is the sigma factors, which determine promoter selection by RNA polymerase. There are an extraordinarily high number (65) of sigma factors encoded in the *S. coelicolor* genome, as well as at least seven antisigma factors and 15 antisigma factor antagonists (some of each of which were detected). Sigma factor heterogeneity in bacteria is an important way of regulating expression of specific sets of genes under

different conditions and in different cell types (for a review, see Gross *et al.*, 1998). Although no sigma factors were identified, about half (31 out of 65) have predicted pI values >pH 7 and so can only be detected using the pH 6–11 IPG strips that give the poorest results (see Table 1). However, the principal vegetative sigma factor, HrdB, which is expected to be moderately abundant, has a predicted pI of 6.13, and its absence from the pH 5.5–6.7 proteome map may indicate that the sample preparation method, which does not remove DNA, is inappropriate. Among other regulatory proteins, 11 out of a possible 80 two-component system response regulators plus a further 36 proteins annotated as 'regulators' were detected. Only one Ser/Thr kinase and no two-component system sensor kinases were detected, possibly reflecting the difficulty in analysing proteins with several transmembrane domains. Interestingly, 30 secreted proteins were identified on the proteome map, at least 10 of which appear, from comparison of their observed pI and molecular weight positions with predicted values, to have been processed already and therefore exported. Presumably, these proteins do not end up exclusively in the culture medium.

Clearly, this approach to proteomic analysis of *S. coelicolor* has some important deficiencies, not least the absence of certain groups of proteins and the relatively poor initial success rate for specifically characterizing post-translational modifications. Nevertheless, it offers unique insights into the molecular biology of the organism: the expression of many proteins from both primary and secondary metabolic pathways can be analysed; post-translational regulation of proteins can be investigated on a general cellular level; and the physical locations for some proteins can be determined. Approximately 35% of the *S. coelicolor* genome is annotated as encoding 'hypothetical' proteins of unknown function, and 182 proteins of this type were identified on the proteome map. No examples were found of overt disagreements between predicted and observed proteins, providing experimental support for the general accuracy of the *S. coelicolor* genome sequence and its annotation. Information about the expression, post-translational modification and physical location of these (no longer!) hypothetical proteins from proteomic analyses is going to be important for assigning future functions. As many of the hypothetical proteins have apparent orthologues in pathogenic mycobacteria, some of the information may have an immediate application in the context of pathogenesis.

Experimental procedures

Materials

Immobilized pH gradient (IPG) strips and IPG buffers were purchased from Amersham Biosciences. Acrylamide,

duracryl and chemicals for silver staining were obtained from Genomic Solutions. All other chemicals were from Sigma unless stated in the text.

Strain and growth conditions

Streptomyces coelicolor M145 was cultivated with vigorous agitation at 30°C in minimal medium supplemented with 0.2% casamino acids (SMM) as described previously (Kieser *et al.*, 2000). Briefly, a high-density spore preparation [about 10¹⁰ colony-forming units (cfu) ml⁻¹] was pregerminated in 2× YT media for 7 h at 30°C. Germlings were harvested by centrifugation (5 min at 4000 g), resuspended in SMM and sonicated briefly to disperse any aggregates before inoculation into 50 ml of SMM in 250 ml siliconized flasks containing coiled stainless steel springs. Each flask received the equivalent of 5 × 10⁷ spore cfu.

Protein extraction and 2D-gel electrophoresis

The preparation of the protein sample to be separated was kept as simple as possible, disrupting harvested mycelia directly into denaturing isoelectric focusing (IEF) buffer containing Pefabloc SC (Roche) protease inhibitor. This method minimized unwanted protein degradation that was observed in initial studies using a Tris buffer containing Pefabloc SC for cell lysis (data not shown). Removal of DNA, RNA, salts and polysaccharides from the sample was therefore not performed (other than the shearing of viscous DNA and RNA by sonication), and the protein extracts were sometimes difficult to focus when the sample was loaded by in-gel rehydration. Loading the sample, soaked in paper strips, at the anodic end of IPG strips that had already been rehydrated improved the reproducibility of separations significantly.

Mycelium for protein extraction was harvested from cultures by brief centrifugation (30 s at 4000 g) at room temperature and immediately frozen in liquid nitrogen. Typically, mycelium from 10 ml culture aliquots was collected, and the transfer time from culture flask to frozen sample was 1.5 min. Frozen cells were thawed on ice in 5 ml of washing buffer (40 mM Tris, pH 9.0, 1 mM EGTA, 1 mM EDTA), then pelleted by centrifugation (5 min at 3000 g) at 4°C. Washed cells were resuspended in 400 µl of denaturing IEF buffer UTCHAPS [7 M urea, 2 M thiourea, 4% (w/v) CHAPS, 40 mM Tris, pH 9.0, 1 mM EDTA, 50 mM DTT, 4 mM Pefabloc SC protease inhibitor (Roche)] and disrupted by sonication (Sanyo Soniprep 150; 10 × 2 s bursts at amplitude 7.5 µm) while cooling in an ethanol-ice bath. Cell debris was removed by centrifugation (15 min, 10 000 g, 4°C), and the protein extract was stored frozen in aliquots at -80°C until use. For the preparation of extracellular protein extracts, washed cells were resuspended in extraction buffer consisting of 40 mM Tris, pH 9.0, 1 mM EDTA, 50 mM DTT, 2% SDS and 4 mM Pefabloc SC protease inhibitor. Cells were vortexed for 2 min before being pelleted by centrifugation (5 min, 10 000 g, 4°C). The supernatant was precipitated using the Amersham 2D Clean-Up kit according to the manufacturer's instructions, and the precipitated proteins were redissolved in UTCHAPS and stored at -80°C until use.

For first-dimension isoelectric focusing, 18 cm IPG strips pH 4–7, pH 6–11, pH 4.5–5.5 or pH 5.5–6.7 (Amersham Biosciences) were rehydrated overnight in IEF buffer containing 1% ampholytes according to the manufacturer's instructions using a Phasor isoelectric focusing unit (Genomic Solutions) set at 20 V. Protein samples to be separated were applied to the rehydrated strips at the anodic end, soaked in a 5–10 mm section of IEF electrode strip (Amersham Biosciences). Separation was performed for 120 000 V-h with a maximum voltage of 5000 V. After isoelectric focusing, IPG strips were equilibrated for the second dimension for 15 min in IPG equilibration buffer (50 mM Tris, pH 6.8, 6 M urea, 30% glycerol, 1% SDS and 0.01% bromophenol blue) plus 80 mM DTT, then for 10 min in IPG equilibration buffer plus 135 mM iodoacetamide. Approximately 1 cm was removed from the anodic end of each equilibrated strip, before applying the strip to the top of a vertical 12.5% SDS-PAGE gel for second-dimension separation using the Investigator 5000 system from Genomic Solutions.

Large-format gels (24 cm wide × 22 cm high × 1 mm thick) for the second-dimension separation were cast in-house using the system supplied by Genomic Solutions. 12.5% SDS-PAGE gels, prepared using 0.65% N,N-methylenebisacrylamide as cross-linking agent, were used for pH 4–7 and pH 4.5–5.5 IPG strips, whereas pH 5.5–6.7 and pH 6–11 strips were applied to similar gels made with 0.8% cross-linker. Electrophoresis was performed while cooling using the maximum setting (≈ 4 – 10°C) at 20 000 mW per gel constant power and a maximum voltage of 500 V. Gels were then stained with either colloidal Coomassie G-250 (Neuhoff *et al.*, 1988) or silver nitrate (Rabilloud, 1992) and scanned in a ProXPRESS proteomic imaging system (Perkin-Elmer). Image analysis was performed using PHORETIX 2D version 5.1 (NonLinear Dynamics): spot detection was optimized automatically using the 'Spot Detection Wizard' and then edited manually; background subtraction was performed automatically using the 'Mode of Non-Spot' setting; images were then normalized to the total spot volume for quantification. Spot filtering was not used.

Protein identification using mass spectrometry

Protein spots of interest were excised from colloidal Coomassie-stained gels, in-gel digested with trypsin and identified by MALDI-TOF peptide mass fingerprint analysis. Excised gel pieces (≈ 1 -mm-diameter circles cut from 1-mm-thick gels) were washed twice with 100 μl of 50 mM ammonium bicarbonate for 15 min, once with 100 μl of 20% acetonitrile–40 mM ammonium bicarbonate for 15 min and once with 100% acetonitrile. Washed gel pieces were allowed to air-dry for 10 min before being rehydrated in 5 μl of 10 $\mu\text{g ml}^{-1}$ trypsin (sequencing grade modified; Promega) in 10 mM ammonium bicarbonate. Digestion (37°C for 4 h) was stopped by the addition of 5 μl of 5% formic acid. Extraction of peptides into this solution was encouraged by a sonic water bath treatment for 20 min. Peptide extract (0.5 μl) was spotted on to a thin layer of α -cyano-4-hydroxycinnamic acid applied to a MALDI-TOF sample template and analysed by MALDI-TOF mass spectrometry (Bruker Reflex III) using an accelerating voltage of 25 kV. Samples were externally calibrated using a standard mixture of six peptides ranging in

mass from 1046.5423 Da to 3494.6500 Da, and the mass accuracy obtained was 60 p.p.m. or better. Identification of proteins from MALDI-TOF peptide mass fingerprint data was performed using the 'Mascot' search engine at <http://www.matrixscience.com> and was based on a positive result using their 'Probability Based MOWSE Score' algorithm. A MOWSE score of 60 or higher is significant at the 5% level or better, and proteins in this work typically gave scores >80 (frequently considerably so). In addition, no identification was accepted unless at least five peptides representing at least 20% of the protein sequence were detected in the MALDI-TOF peptide mass fingerprint.

Assignment of proteins to metabolic pathways

Proteins were assigned to primary metabolic pathways using the 'Protein Classification Scheme' web page of the annotated *Streptomyces coelicolor* genome sequence at http://www.sanger.ac.uk/Projects/S_coelicolor, as it appeared in May 2002. The annotation of the genome, and therefore the Protein Classification Scheme, is likely to be modified as more experimental data are obtained about the function of genes that have currently been assigned on the basis of sequence information alone. It is therefore possible that proteins assigned to metabolic pathways in this study may in future be shown to perform different functions in the cell.

Assignments to gene clusters encoding secondary metabolites are based on an analysis of the genome sequence reported by Bentley *et al.* (2002). In some cases, these assignments are based on extensive experimental data (*act*, *red*: reviewed by Chater and Bibb, 1997; Champness, 2000; *cda*: Chong *et al.*, 1998; Huang *et al.*, 2001; *whiE*: Davis and Chater, 1990).

Acknowledgements

This work was funded by grants (2/FGT11406, 208/FGT11408 and 208/IGF12432) under the BBSRC's Technologies for Functional Genomics and Investigating Gene Function initiatives, and by a competitive strategic grant from the BBSRC to the John Innes Centre. We thank Mike Naldrett for his extensive contribution in establishing technology for proteome analysis at the JIC.

Supplementary material

The following material is available from <http://www.blackwell-science.com/products/journals/suppmat/mole/mole3219/mmi3219sm.htm>

Section 1. Detection of gene products assigned to individual metabolic steps in primary metabolic pathways in *S. coelicolor*.

Section 2. Identification of proteins from individual amino acid biosynthetic pathways.

Section 3. Detection of gene products assigned to individual metabolic steps in secondary metabolic pathways in *S. coelicolor*.

References

- Anderson, T.B., Brian, P., and Champness, W.C. (2001) Genetic and transcriptional analysis of *absA*, an antibiotic gene cluster linked two-component system that regulates multiple antibiotics in *Streptomyces coelicolor*. *Mol Microbiol* **39**: 553–566.
- Bentley, S.D., Chater, K.F., Cerdeno-Tarraga, A.-M., Challis, G.L., Thomson, N.R., James, K.D., *et al.* (2002) Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). *Nature* **417**: 141–147.
- Berdy, J. (1984) New ways to obtain antibiotics. *Chin J Antibiot* **7**: 272–290.
- Berggren, K., Chernokalskaya, E., Steinberg, T.H., Kemper, C., Lopez, M.F., Diwu, Z., *et al.* (2000) Background-free, high sensitivity staining of proteins in one- and two-dimensional sodium dodecyl sulfate-polyacrylamide gels using a luminescent ruthenium complex. *Electrophoresis* **21**: 2509–2521.
- Blackstock, W.P., and Weir, M.P. (1999) Proteomics: quantitative and physical mapping of cellular proteins. *Trends Biotechnol* **17**: 121–127.
- Blattner, F.R., Plunkett, G., Bloch, C.A., Perna, N.T., Burland, V., Riley, M., *et al.* (1997) The complete genome sequence of *Escherichia coli* K-12. *Science* **277**: 1453–1462.
- Buttner, K., Bernhardt, J., Scharf, C., Schmid, R., Mader, U., Eymann, C., *et al.* (2001) A comprehensive two-dimensional map of cytosolic proteins of *Bacillus subtilis*. *Electrophoresis* **22**: 2908–2935.
- Champness, W. (2000) Actinomycete development, antibiotic production and phylogeny. questions and challenges. In *Prokaryotic Development*. Brun, Y.V., and Skimketts, L.J. (eds). Washington, DC: American Society for Microbiology Press, pp. 11–31.
- Chater, K.F., and Bibb, M.J. (1997) Regulation of bacterial antibiotic production. In *Biotechnology*, Vol. 7. *Products of Secondary Metabolism*. Kleinkauf, H., and von Döhren, H. (eds). Weinheim, Germany: VCH, pp. 57–105.
- Chong, P.P., Podmore, S.M., Kieser, H.M., Redenbach, M., Turgay, K., Marahiel, M., *et al.* (1998) Physical identification of a chromosomal locus encoding biosynthetic genes for the lipopeptide calcium-dependent antibiotic (CDA) of *Streptomyces coelicolor* A3(2). *Microbiology* **144**: 193–199.
- de Crecy-Lagard, V., Servant-Moisson, J.V., Grandvalet, C., and Mazodier, P. (1999) Alteration of the synthesis of the Clp ATP-dependent protease affects morphological and physiological differentiation in *Streptomyces*. *Mol Microbiol* **32**: 505–517.
- Davis, N.K., and Chater, K.F. (1990) Spore colour in *Streptomyces coelicolor* A3(2) involves the developmentally regulated synthesis of a compound biosynthetically related to polyketide antibiotics. *Mol Microbiol* **4**: 1679–1691.
- Fernandez-Moreno, F.A., Martinez, E., Caballero, J.L., Ichinose, K., Hopwood, D.A., and Malpartida, F. (1994) DNA sequence and functions of the *actVI* region of the actinorhodin biosynthetic gene cluster of *Streptomyces coelicolor* A3(2). *J Biol Chem* **269**: 24854–24863.
- Fink, D., Falke, D., Wohlheben, W., and Engels, A. (1999) Nitrogen metabolism in *Streptomyces coelicolor* A3(2): modification of glutamine synthase I by an adenyltransferase. *Microbiology* **145**: 2313–2322.
- Godovac-Zimmermann, J., Soskic, V., Posnanovic, S., and Brianza, F. (1999) Functional proteomics of signal transduction by membrane receptors. *Electrophoresis* **20**: 952–961.
- Gorg, A., Obermaier, C., Boguth, G., Harder, A., Scheibe, B., Wildgruber, R., and Weiss, W. (2000) The current state of two-dimensional electrophoresis with immobilized pH gradients. *Electrophoresis* **21**: 1037–1053.
- Gramajo, H.C., Takano, E., and Bibb, M.J. (1993) Stationary-phase production of the antibiotic actinorhodin in *Streptomyces coelicolor* A3(2) is transcriptionally regulated. *Mol Microbiol* **7**: 837–845.
- Gross, C.A., Chan, C., Dombroski, A., Gruber, T., Sharp, M., Tupy, J., and Young, B. (1998) The functional and regulatory roles of sigma factors in transcription. *Cold Spring Harb Symp Quant Biol* **63**: 141–155.
- Hesketh, A., Sun, J., and Bibb, M.J. (2001) Induction of ppGpp synthesis in *Streptomyces coelicolor* A3(2) grown under conditions of nutritional sufficiency elicits *actII-ORF4* transcription and actinorhodin biosynthesis. *Mol Microbiol* **39**: 136–144.
- Hesketh, A., Fink, D., Gust, B., Rexer, H.-U., Scheel, B., Chater, K., *et al.* (2002) The PII nitrogen regulatory protein GlnK in *Streptomyces coelicolor* A3(2) is modified by adenylation rather than uridylation, and the GlnD and GlnK homologues are functionally dissimilar to their nitrogen regulatory system counterparts. *Mol Microbiol* (in press).
- Hood, D.W., Heidstra, R., Swodoba, U.K., and Hodgson, D.A. (1992) Molecular genetic analysis of proline and tryptophan biosynthesis in *Streptomyces coelicolor* A3(2): interaction between primary and secondary metabolism—a review. *Gene* **115**: 5–12.
- Hu, D.S.-J., Hood, D.W., Heidstra, R., and Hodgson, D.A. (1999) The expression of the *trpD*, *trpC* and *trpBA* genes of *Streptomyces coelicolor* A3(2) is regulated by growth rate and growth phase but not by feedback repression. *Mol Microbiol* **32**: 869–880.
- Huang, J., Lih, C.J., Pan, K.H., and Cohen, S.N. (2001) Global analysis of growth phase responsive gene expression and regulation of antibiotic biosynthetic pathways in *Streptomyces coelicolor* using DNA microarrays. *Genes Dev* **15**: 3183–3192.
- Ichinose, K., Taguchi, T., Ebizuka, Y., and Hopwood, D.A. (1998a) Biosynthetic gene clusters of benzoisochromanone antibiotics in *Streptomyces* spp. Identification of genes involved in post-PKS tailoring steps. *Actinomycetologica* **12**: 99–109.
- Ichinose, K., Bedford, D.J., Tornus, D., Bechthold, A., Bibb, M.J., Revill, P.W., *et al.* (1998b) The granaticin biosynthetic gene cluster of *Streptomyces violaceoruber* Tu22: sequence analysis and expression in a heterologous host. *Chem Biol* **5**: 647–659.
- Kelemen, G.H., Brian, P., Flardh, K., Chamberlin, L., Chater, K., and Buttner, M. (1998) Developmental regulation of transcription of *whiE*, a locus specifying the polyketide spore pigment in *Streptomyces coelicolor* A3(2). *J Bacteriol* **180**: 2515–2521.
- Kieser, T., Bibb, M.J., Buttner, M.J., Chater, K.F., and

- Hopwood, D.A. (2000) *Practical Streptomyces Genetics*. Norwich: John Innes Foundation.
- Kim, E.J., Chung, H.J., Suh, B., Hah, Y.C., and Roe, J.H. (1998) Expression and regulation of the *sodF* gene encoding iron- and zinc-containing superoxide dismutase in *Streptomyces coelicolor* Müller. *J Bacteriol* **180**: 2014–2020.
- Kimura, Y., Takaoka, M., Tanaka, S., Sassa, H., Tanaka, K., Polevoda, B., et al. (2000) N α -acetylation and proteolytic activity of the yeast 20S proteasome. *J Biol Chem* **275**: 4635–4639.
- Kunst, F., Ogasawara, N., Moszer, I., Albertini, A.M., Alloni, G., Azevedo, V., et al. (1997) The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*. *Nature* **390**: 249–256.
- Lucchini, S., Thompson, A., and Hinton, J.C.D. (2001) Microarrays for microbiologists. *Microbiology* **147**: 1403–1414.
- Mann, M., Hendrickson, R.C., and Pandey, A. (2001) Analysis of proteins and proteomes by mass spectrometry. *Annu Rev Biochem* **70**: 437–473.
- Miyodah, S. (1993) Research on antibiotic screening in Japan over the last decade: a producing microorganisms approach. *Actinomycetologica* **7**: 100–106.
- Molloy, M.P., Herbert, B.R., Williams, K.L., and Gooley, A.A. (1999) Extraction of *Escherichia coli* proteins with organic solvents prior to two-dimensional electrophoresis. *Electrophoresis* **20**: 701–704.
- Nagy, I., Tamura, T., Vanderleyden, J., Baumeister, W., and De Mot, R. (1998) The 20S proteasome of *Streptomyces coelicolor*. *J Bacteriol* **180**: 5448–5453.
- Neuhoff, V., Arold, N., Taube, D., and Ehrhardt, W. (1988) Improved staining of proteins in polyacrylamide gels including isoelectric focusing gels with clear background at nanogram sensitivity using Coomassie Brilliant Blue G-250 and R-250. *Electrophoresis* **9**: 255–262.
- Nilsson, C.L., and Davidsson, P. (2000) New separation tools for comprehensive studies of protein expression by mass spectrometry. *Mass Spectrom Rev* **19**: 390–397.
- Puglia, A.M., Vohradsky, J., and Thompson, C.J. (1995) Developmental control of the heat-shock stress regulon in *Streptomyces coelicolor*. *Mol Microbiol* **17**: 737–746.
- Rabilloud, T. (1992) A comparison between low background silver diamine and silver nitrate protein stains. *Electrophoresis* **13**: 429–439.
- Robinson, V.L., Buckler, D.R., and Stock, A.M. (2000) A tale of two components: a novel kinase and a regulatory switch. *Nature Struct Biol* **7**: 626–633.
- Schena, M. (2000) *Microarray Biochip Technology*. Natick, MA: Eaton Publishing.
- Spusansky, O., Rezuchova, B., Homerova, D., and Kormanec, J. (2001) Expression of the *gap* gene encoding glyceraldehyde-3-phosphate dehydrogenase of *Streptomyces aureofaciens* requires GapR, a member of the AraC/XylS family of transcriptional activators. *Microbiology* **147**: 1291–1301.
- Taguchi, T., Itou, K., Ebizuka, Y., Malpartida, F., Hopwood, D.A., Surti, C.M., et al. (2000) Chemical characterisation of disruptants of the *Streptomyces coelicolor* A3(2) *acVI* genes involved in actinorhodin biosynthesis. *J Antibiot* **53**: 144–152.
- Takano, E., Gramajo, H.C., Strauch, E., Andres, N., White, J., and Bibb, M.J. (1992) Transcriptional regulation of the *redD* transcriptional activator gene accounts for growth-phase-dependent production of the antibiotic undecylprodigiosin in *Streptomyces coelicolor* A3(2). *Mol Microbiol* **6**: 2797–2804.
- Tonella, L., Hoogland, C., Binz, P.-A., Appel, R.D., Hochstrasser, D.F., and Sanchez, J.-C. (2001) New perspectives in the *Escherichia coli* proteome investigation. *Proteomics* **1**: 409–423.
- Vohradsky, J., Li, X.M., and Thompson, C.J. (1997) Identification of prokaryotic developmental stages by statistical analysis of two-dimensional gel patterns. *Electrophoresis* **18**: 1418–1428.
- Vohradsky, J., Li, X.M., Dale, G., Folcher, M., Nguyen, L., Viollier, P.H., and Thompson, C.J. (2000) Developmental control of stress stimulons in *Streptomyces coelicolor* revealed by statistical analyses of global gene expression patterns. *J Bacteriol* **182**: 4979–4986.