New Insights from Genomics and Functional Genomics in *Streptomyces coelicolor*

Keith Chater, Govind Chandra, Bertolt Gust, Andy Hesketh, Helen Kieser, and Tobias Kieser

John Innes Centre, Norwich NR4 7UH, UK

The complete sequence of the chromosome (8.67 Mb), the linear plasmid SCP1 (365 Kb) and the circular plasmid SCP2* (30Kb), which together make up the complete genome of *Streptomyces coelicolor* A3(2), have been determined. Analysis of the genome (and comparison with other actinomycete genomes) provides new insights into the evolution and adaptation of streptomycetes. The genome also provides a crucial platform from which functional genomics can take off. One important aspect of this is the capability to manipulate the genome readily, through variations on the theme of PCR targeted mutagenesis. The genome also provides information that permits the unambiguous identification of nearly all the protein spots detectable by 2D gel electrophoresis. It appears that at least half of the 7825 predicted genes in the chromosome are expressed as proteins at the transition from growth to stationary phase.

Introduction

Among streptomycetes, *Streptomyces coelicolor* A3(2) has been the leading model representing this important group of antibiotic-producing bacteria for more than 40 years, thanks to the clear-sighted development of its genetics led by D. A. Hopwood (Hopwood, 1999). The culmination of this approach was the determination of the complete chromosome sequence of a plasmid-free derivative (M145) of the A3(2) strain (Bentley *et al.*, 2002). Since then, the two plasmids of the wild-type, SCP1 (356 kb) and SCP2 (31 kb), have also been completely sequenced (S.D. Bentley *et al.*, unpublished; Haug *et al.*, 2003). The SCP1 and chromosome sequences are accessible at http://jic-bioinfo.bbsrc.ac.uk/S.coelicolor/. The availability of this wonderful resource has two major kinds of application, broadly epitomised as functional genomics and comparative genomics. In this brief article, we summarise recent progress in these two directions (particularly the former) in our laboratory.

Functional Genomics

Proteomics. The initial annotation of the *S. coelicolor* genome provided a theoretical proteome of 7825 proteins of predictable mass and isoelectric point (note that this number does not take account of duplicated genes). Separation of proteins on classical two-dimensional polyacrylamide gels depends on exactly these properties. It is therefore possible to use this technique to match up the theoretical proteome with what is actually present in cells and cultures. Moreover, the analysis of tryptic digests of individual protein spots by MALDI-ToF mass spectrometry allows essentially any spot detectable by moderately sensitive stains, such as colloidal Coomassie brilliant blue, to be identified with a gene. We have applied such analysis to *S. coelicolor* cultures grown in liquid medium (Hesketh *et al.*, 2002; http://qbab.aber.ac.uk/s_coeli/referencegel). It turns out that probably half of the entire genome is being expressed at the transition from growth to stationary

phase. Many of the spots identified correspond to genes annotated as "hypotheticals": clearly, the detection of such proteins proves that they are real! No examples have been found of proteins whose sequence implies a significant misinterpretation of, or mistake in, the genome sequence (such as a frameshift error), providing a testament to the accuracy of the sequencing done at the Sanger Institute. Remarkably, about 20% of genes are represented by more than one spot, presumably in many cases because of post-translational processing or modification. Examples of such phenomena include several of the enzymes involved in antibiotic biosynthesis. This is particularly clear in the actinorhodin pathway, in which we can readily detect a high proportion of the predicted gene products (Hesketh *et al*, 2002 and submitted). In one particularly striking case, this has led to the finding that an enzyme ostensibly involved in closure of the ether ring is located in the cell envelope, and has lost an N-terminal secretion signal sequence. Clearly, there is an implication that investigators studying secondary metabolite biosynthesis should not be too naïve in assuming that the dynamics of biosynthesis are determined just by the level of gene expression and the supply of metabolites.

Targeted mutagenesis. The genome of *S. coelicolor* was sequenced by the use of an ordered collection of overlapping cosmid clones, and the resource of these fully sequenced cosmids has become an important tool for genetic manipulation. The basic point is that selectable changes introduced into a cosmid in *E. coli* can easily be moved into the *S. coelicolor* genome, because the long flanking sequences on either side of such changes in a cosmid permit frequent double crossover recombination with homologous sequences in *S. coelicolor*. How can we efficiently do the initial manipulations in *E. coli* to bring about the desired changes? The simplest case is straightforward gene disruption in cosmid clones. Two highly successful approaches have been made to this: PCR targeting, in our laboratory (Gust *et al.*, 2003); and *in vitro* transposon mutagnesis, in the laboratory of Dr Paul Dyson (personal communication).

Our work has taken advantage of the tools developed by Datsenko and Wanner (2000) to permit the targeting of selectable resistance cassettes, amplified by PCR using long oligonucleotide primers homologous to the desired gene, directly into the DNA sequence in question using in vivo recombination. This requires a host capable of bringing about recombination at a useful level when only about 40 bases of sequence identity are provided. In the Datsenko and Wanner sysem, this is made possible by the induction (by arabinose) of the RED system of bacteriophage lambda. Key features of our version of this system are the use of resistance cassettes selectable in both *E. coli* and *Streptomyces*; and the incorporation into the resistance cassettes of a wide host-range origin of DNA transfer, which allows the mutated cosmid to be mobilised directly into *Streptomyces* without the need for DNA isolation and labour- and time-consuming protoplast transformation. Using this system, we have been able to make knockouts in many more than 100 genes, and it has become straightforward to make multiple mutants – something that is of great value, for example, in the analysis of sets of genes with overlapping function, such as the eight genes for a novel family of surface-associated proteins (chaplins: Elliot *et al.*, 2003). There are many variants on this system. For example, we have been able to insert a controllable promoter upstream of a cluster of genes that are normally expressed at an inconveniently low level (C.J. Bruton, pers. comm.), make fusions with fluorescent reporter genes (D. Jakimowicz, pers. comm.), and introduce point mutations into the genome (Tian, Y., pers. comm.). Most versions of the resistance cassettes contain target sites for the yeast Flp recombinase (FRT sites), that permit the use of the recombinase to excise the selectable resistance gene, leaving behind a phenotypically neutral 81-bp "scar" that has no stop codons. This permits the non-polar disruption of genes within operons, or, as in our dissection of the geosmin synthase (Gust *et al.*, 2003), the surgical removal of particular domains of complex proteins.

Comparative Genomics

At the time of writing, 127 bacterial and archaeal genomes have been completely sequenced, making it possible to countenance meaningful global comparisons. One starting point for such analyses is to compare the genome of *S. coelicolor* with that of *Streptomyces avermitilis,* the nearest relative for which a genome sequence is available (Omura *et al.*, 2001; Ikeda *et al.*, 2003). These two species shared a common ancestor probably more than 100 million years ago, yet the central portion (c. 6 Mb) of their genomes shows an extraordinary degree of similarity in gene order. Of course, this is an important aspect of the analysis done by the Kitasato group who sequenced the *S. avermitilis* genome, but here we only wish to point out two ways in which comparisons between genomes can lead to new insights into genome-wide questions, or even to new questions.

We have long been interested in the *bldA* gene of *S. coelicolor.* This gene encodes the only tRNA in this organism able to translate the UUA codon efficiently (Leskiw *et al.*, 1991). Deletion of *bldA* has no effect on vegetative growth, implying that this codon is absent from all essential genes. Strikingly, however, the mutant has multiple defects in its post-growth attributes: on most media, it fails to make at least three different antibiotics, and it does not form aerial mycelium or spores. This is confirmed by genomic analysis. Of the 7825 genes annotated by Bentley *et al.,* (2002), only 145 (1.9%) contain the TTA codon, and none of these had been allocated a function expected to be essential. We find that the *S. avermitilis* genome contains rather more genes with the TTA codon (260, 3.4%). By considering which of these are orthologues of TTA-containing genes of *S. coelicolor,* taking advantage of the conservation of gene order (synteny) between the two genomes as a confirmation of orthology, we are able to narrow down the hunt for *crucial* target genes of *bldA.* Indeed we find that only eleven genes common to both organisms carry the codon, permitting a much more focussed approach to be taken to analysing *bldA*-related phenomenology at the molecular level.

A second question of interest to many investigators is the analysis of non-coding regions of genomes. Using a new approach, we discovered a novel family of small (<200bp) imperfectly repeated sequences in *S. coelicolor*, which have diverged just enough to make it difficult to use hybridisation to find out what the distribution of the sequences is in other streptomycetes. With the availability of the *S. avermitilis* genome sequence, it has become obvious that these elements are absent from equivalent positions in the *S. avermitilis* genome, indicating that they do not carry out a critical role in the expression of the genes neighbouring them. Thus, they can be considered as transposable selfish DNA, even though they do not have any recognisable coding capacity and therefore presumably use a trans-acting transposition machinery.

Closing Remarks

In this brief overview, it is clear that *Streptomyces* research has moved into a new dimension as a result of genome sequence determination. We have not dealt with some applications, notably the use of microarrays for transcriptome analysis (Huang *et al.*, 2001; Chater *et al*, 2002) and for taxonomic and phylogenetic studies (C. Kao, personal communication). We can confidently predict that the next few years will see unprecedented progress towards an integrated view of *Streptomyces* evolution, ecology, cell and developmental biology, secondary and primary metabolism, and homeostatic and stress biology that will revolutionise our exploitation of these wonderful organisms for human welfare.

Acknowledgments

This work was supported by the Biotechnological and Biological Sciences Research Council (UK) through grants FGT11408 and IGF12432. We thank Profesor Haruo Ikeda for pre-publication information on the *S. avermitilis* genome sequence.

References

1. Bentley, S.D., Chater, K.F., Cerdeno-Tarraga, A.M., Challis, G.L., Thomson, N.R., James, K.D., Harris, D.E., Quail, M.A., Kieser, H., Harper, D., Bateman, A., Brown, S., Chandra, G., Chen, C.W., Collins, M., Cronin, A., Fraser, A., Goble, A., Hidalgo, J., Hornsby, T., Howarth, S., Huang, C.H., Kieser, T., Larke, L., Murphy, L., Oliver, K., O'Neil, S., Rabbinowitsch, E., Rajandream, M.A., Rutherford, K., Rutter, S., Seeger, K., Saunders, D., Sharp, S., Squares, R., Squares, S., Taylor, K., Warren, T., Wietzorrek, A., Woodward, J., Barrell, B.G., Parkhill, J., and Hopwood, D.A. (2002) Complete genome sequence of the model actinomycete Streptomyces coelicolor A3(2). Nature 417:141-147.
2. Elliot, M.A., Karoonuthaisiri, N., Huang, J., Bibb, M.J., Cohen, S.N., Kao, C.M., and Buttner, M.J. (2003) The chaplins: a family of secreted, hydrophobic proteins involved in aerial mycelium formation in Streptomyces coelicolor. Submitted for publication.
3. Gust, B., Challis, G.L., Fowler, K., Kieser, T., and Chater, K.F. (2003). PCR-targeted Streptomyces gene disruption identifies a protein domain needed for biosynthesis of the sesquiterpene soil odor geosmin. Proc. Natl. Acad. Sci., USA, 100:1541-1546
4. Chater, K.F., Bucca, G., Dyson, P., Fowler, K., Gust, B., Herron, P., hesketh, A., Hotchkiss, G., Kieser. T., Mersinias, V. and Smith, C.P. (2002) Methods in Microbiology 33, 321-336.
5. Haug, I., Weissenborn, A., Brolle, D., Bentley, S., Keiser, T. and Altenbuchner, J. (2003) Streptomyces coelicolor plasmid SCP2*: deductions from the complete sequence. Microbiology 149, 505-535
6. Hesketh, A.R., Chandra, G., Shaw, A.D. Rowland, J.J., Kell, D.B., Bibb, M.J. and Chater, K.F. (2002) Primary and secondary metabolism, and post-translational protein modifications, as portrayed by proteomic analysis of Streptomyces coelicolor. Mol. Microbiol. 46, 917-932
7. Hopwood, D.A. (1999) 40 years of genetics with Streptomyces: from in vivo to in vitro to in silico. Microbiology 145, 2183-2202
8. Huang, J., Li, 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 & Development 15, 3183-3192
9. Leskiw, B.K., Lawlor, E.J., Fernandez-Abalos, J.M., and Chater, K.F. (1991) TTA codons in some genes prevent their expression in a class of developmental, antibiotic-negative, Streptomyces mutants. Proc Natl Acad Sci USA 88:2461-5.
10. Ikeda H., Ishikawa J., Hanamoto A., Shinose M., Kikuchi H., Shiba T., Sakaki Y., Masahira Hattori, Omura, S. (2003) Complete genome sequence and comparative analysis of the industrial microorganism Streptomyces avermitilis. Nature Biotechnology, 21, 526 – 531
11. Omura S, Ikeda H, Ishikawa J, Hanamoto A, Takahashi C, Shinose M, Takahashi Y, Horikawa H, Nakazawa H, Osonoe T, Kikuchi H, Shiba T, Sakaki Y, Hattori M. (2001) Genome sequence of an industrial microorganism Streptomyces avermitilis: deducing the ability of producing secondary metabolites. Proc. Natl. Acad. Sci. USA, 98, 12215-12220