Plasticity of the *Streptomyces* Genome-Evolution and Engineering of New Antibiotics

Daslav Hranueli[1], John Cullum[2], Bojan Basrak[3], Pavle Goldstein[3] and Paul F. Long*[,4]

[1]*Faculty of Food Technology and Biotechnology, University of Zagreb, Croatia*

[2]*Department of Genetics, University of Kaiserslautern, Germany*

[3]*Department of Mathematics, University of Zagreb, Croatia*

[4]*School of Pharmacy, University of London, UK*

Abstract: *Streptomyces* is a genus of soil dwelling bacteria with the ability to produce natural products that have found widespread use in medicine. Annotation of *Streptomyces* genome sequences has revealed far more biosynthetic gene clusters than previously imagined, offering exciting possibilities for future combinatorial biosynthesis. Experiments to manipulate modular biosynthetic clusters to create novel chemistries often result in no detectable product or product yield is extremely low. Understanding the coupling between components in these hybrid enzymes will be crucial for efficient synthesis of new compounds. We are using new algebraic approaches to predict protein properties, and homologous recombination to exploit natural evolutionary constraints to generate novel functional enzymes. The methods and techniques developed could easily be adapted to study modular, multi-interacting complex systems where appreciable biochemical and comparative sequence data are available, for example, clinically significant non-ribosomally synthesised peptides and polyketides.

Keywords: *Streptomyces*, Hidden Markov models, genetic recombination, enzyme structure-function, novel antibiotics.

INTRODUCTION

The quest for novel chemical entities to develop into pharmaceutical and other healthcare products has never been greater. As a population we are living longer, drugs that prevent the onset of degenerative diseases associated with old age, including many cancers, are seen to be the 'holy-grail' in modern medicine, enabling us to achieve healthier lives [1]. Cures are also needed, especially to combat the growing public health problem of pan-antibiotic resistant pathogens, and the stark threat posed by emerging and re-emerging infectious diseases [2]. Pharmacogenomics now offers the possibility for personalised treatments; while the human genome project allows identification of new targets that may result in cures for previously untreatable diseases [3]. Historically, natural products have been an extraordinarily rich source of new bioactive molecules for the pharmaceutical industry. Despite much interest in alternative approaches to new lead discovery, bioprospecting for lead compounds from nature continues to be a corner stone in drug development [4]. As well as isolating microorganisms from unique or biological diversity 'hotspots', approaches have also been developed to exploit the chemical diversity from slow-growing or uncultivable microbes. These approaches include the heterologous expression of genomic DNA libraries in more physiologically amenable hosts such as *Streptomyces* [5]. Natural products from this genus of bacteria have been an unparalleled source of chemical diversity extensively screened by the pharmaceutical industry [6]. Such biosynthetic repertoires make *Streptomyces* attractive hosts to express exogenous DNA. Advances in functional genomics should allow the engineering of *Streptomyces* 'superhosts' to become a realistic goal, offering a more rational approach to yield optimisation than more traditional strain improvement techniques [7]. We review current understanding of chromosome structure in *Streptomyces* and suggest how a better understanding of the genetic endowment of these bacteria might provide alternative routes to antibiotic discovery. It might be possible to exploit horizontal gene transfer, which seems to play an important role in natural evolution.

BIOSYNTHESIS OF MODULAR POLYKETIDES AND NON-RIBOSOMAL PEPTIDES

Non-ribosomal peptides and modular polyketides represent a large class of structurally diverse natural products much studied over recent years because the enzymes that synthesise them, the Type A non-ribosomal peptide synthetases (NRPSs) and the Type I polyketide synthases (PKSs), share striking architectural similarities that can be exploited to generate 'un-natural' natural products [8]. Modular PKS and NRPS proteins are multifunctional, composed of a co-linear arrangement of discrete protein domains representing each enzymic activity which catalyse chain elongation from either carboxy or amino acid building blocks. Each domain is housed within larger modules that form the complex. In *Streptomyces* and related genera, these assembly lines produce well-known compounds including the polyketide antibiotic erythromycin, the non-ribosomal peptide immunosupressant cyclosporin, and the anti-neoplastic drug bleomycin that is synthesised by a hybrid modular PKS-NRPS system [9].

*Address correspondence to this author at the School of Pharmacy, University of London, 29/39 Brunswick Square, London WC1N 1AX, UK; Tel: +44-(0)20-77535868; Fax: +44-(0)20-77535868; Email: paul.long@ams1.ulsop.ac.uk

0929-8673/05 $50.00+.00 © 2005 Bentham Science Publishers Ltd.

Polyketides are assembled through serial condensations of activated coenzyme-A thioester monomers derived from simple organic acids such as acetate, propionate and butyrate. The choice of organic acid allows the introduction of different chiral centres into the polyketide. The active sites required for condensation include an acyltransferase (AT), an acyl carrier protein (ACP) and a β-ketoacylsynthase (KS) (Fig. **1A**). Each condensation results in a β-keto group that undergoes all, some or none of a series of processing steps. Active sites that perform these reactions are contained within the following domains; ketoreductase (KR), dehydratase (DH) and an enoylreductase (ER). The absence of any β-keto processing results in the incorporation of a ketone group into the growing polyketide chain, a KR alone gives rise to a hydroxyl moiety, a KR and DH produce an alkene, while the combination of KR, DH and ER domains lead to complete reduction to an alkane. After assembly, the polyketide chain typically undergoes cyclisation and post-PKS modifications such as glycosylation to give the final active compound [10].

Non-ribosomal peptides are assembled by the so-called "multiple carrier thio-template mechanism" [11]. Three domains are necessary for an elongation module: an adenylation (A) domain that selects the substrate amino acid, analogous to a polyketide AT domain, and activates it as an amino acyl adenylate; a peptidyl carrier protein (PCP) that binds the co-factor 4'-phosphopantetheine to which the activated amino acid is covalently attached, analogous to the ACP of a PKS; and a condensation (C) domain that catalyses peptide bond formation, again analogous to the KS in modular PKSs (Fig. **1B**). In *Streptomyces*, the NRPSs also contain a TE domain located at the C-terminal of the protein which is essential for release of linear, cyclic or branched cyclic peptides. Auxiliary activities can further enlarge the structural diversity of the peptide, especially common are epimerisation domains that convert the thioester-bound amino acid from an L- to D- configuration. An alternative mode of elongation has recently been observed in some NRPS systems. Cyclisation domains can substitute for C-domain in modules where the A-domain incorporates the amino acids cysteine, serine or threonine residues forming thiazoline and oxazoline heterocyclic rings which can then undergo further oxidation [12].

In principle, the modular nature of PKSs and NRPSs should lend itself favourably to rational genetic engineering to produce novel 'un-natural' natural products with predicted structures. These strategies include module swaps and deletions to control chain length, domain substitutions to alter incorporation of different building blocks, and inactivation or addition of auxiliary domains to alter processing of the growing acyl- or peptide chain [13, 14]. Although in many cases the predicted product can be isolated, the yield of such compounds is usually extremely low. In other cases no product can be detected. From the point of view of producing novel compounds commercially, this means that the efforts to date have been largely unsuccessful. Thus, although the role of individual domains and modules may be well understood, the interactions between domains in a module and between modules are not yet tractable. These problems have directed attention to the "linkers" between domains and "dockers" between modules [15]. However, the linker and docker sequences are very variable and are functionally poorly defined. One solution

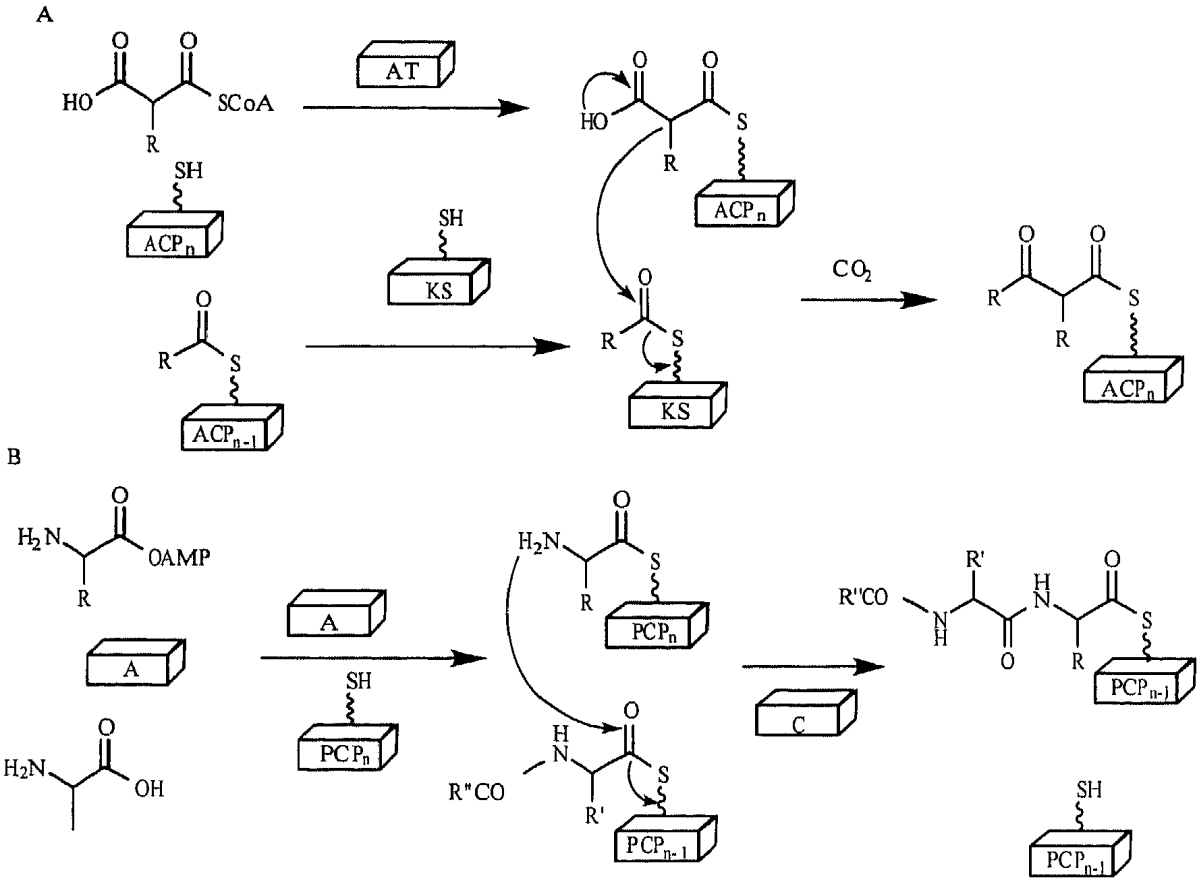

Fig. (1). Schematic presentations of acyl-bond formation catalysed on a polyketide synthase (A) and peptide bond formation on a non-ribosomal peptide synthase (B).

would be to develop statistically robust methods to assign weightings to groups of amino acids that share similar biochemical properties, rather than computing the frequency at which an individual amino acid occurs within a protein alignment of linkers or dockers.

Current approaches have used empirical design rules to define new biosynthetic gene clusters for modular polyketide synthases and non-ribosomal peptide synthases. Viewed from an abstract standpoint, the problem is that of predicting protein function from the primary sequence. A major assumption of this approach is that success can be achieved by local engineering and that it will not be necessary to redesign the whole cluster architecture or content. However, it is very likely that the rules will involve a region of the cluster larger than a single catalytic domain. Questions like universality/non-universality of the ACP and KS domains, upstream/downstream compatibility, linker-domain compatibility and similar should be addressed in a systematic fashion. The sequences of natural clusters provide examples of successful solutions to the problem which have evolved by unique contingent events. This suggests that a new comprehensive approach to extract this information from the sequences of natural clusters might allow the development of design rules for engineering clusters. We propose a probabilistic model of modular PKS and NRPS systems that can be used *in silico* to design genetically engineered systems with a greater than average chance of giving significant product yield.

Our model is sequence-based and was inspired by Hidden Markov model (HMM) family profiles [cf. 16, and references therein]. A family profile is just a convenient probabilistic description of a multiple alignment of a protein family. (Fig. **2A**) shows a HMM as described in the literature [16]. The model allows different sorts of behaviour in the alignment. The most typical behaviour corresponds to the "match" states, where amino acids are matched to the family pattern. There are also "insert" and "delete" states corresponding to extra amino acids or missing amino acids. The model assumes that there are different probabilities corresponding to moving between the different states. As in any HMM, the sequences of known proteins are used as training data to estimate the probabilities. The model can be used subsequently to evaluate new sequences.

Unfortunately, PKS modules are too large to be conveniently modelled in their entirety and our approach was to first exploit the hierarchical structure of a PKS module to divide the model into more tractable components. These

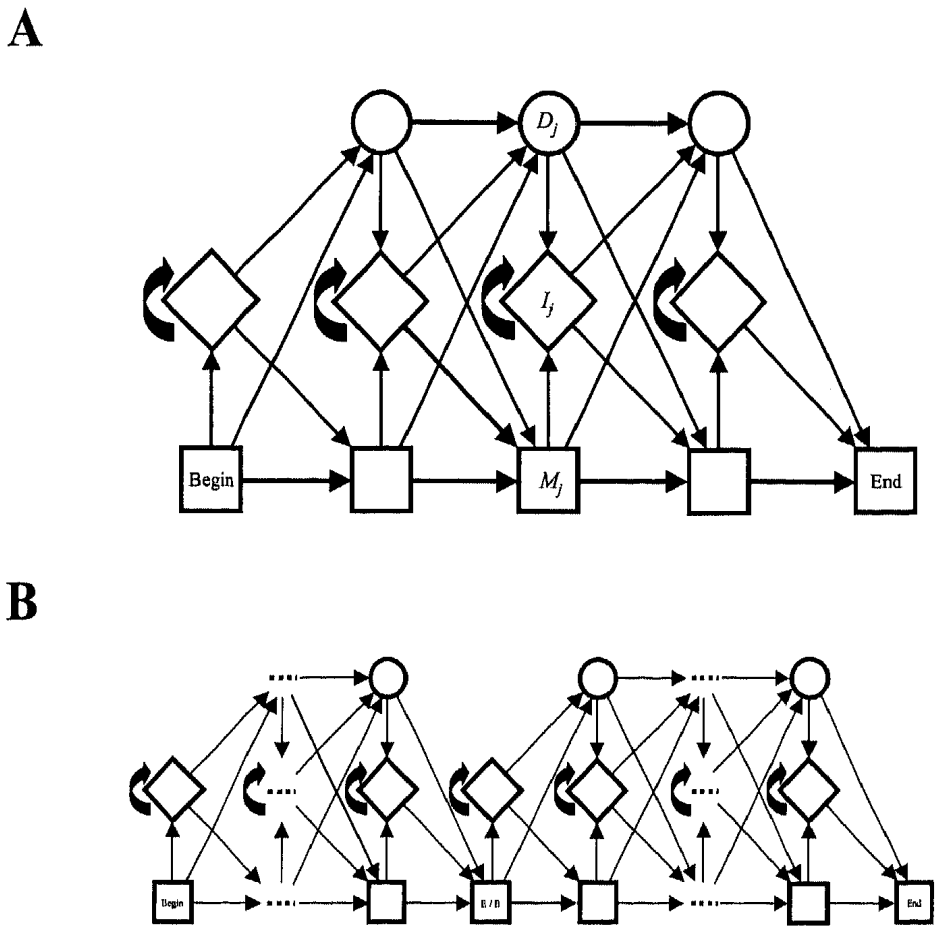

Fig. (2). HMM, used for modelling protein families. Here, the squares **M** correspond to the *match* states, diamonds **I** to the *insert* states and circles **D** to the silent *delete* states. The model is determined in terms of parameters – emission probabilities $e_S(\mathbf{b})$, probability of state **S** emitting the amino acid **b**, for **S** being **M** or **I**, and transition probabilities a_{ST}. Note that the fully connected standard model is shown, and that more restrictive model topology is likely in applications (A). HMM obtained by adjoining models of various components from Fig. 2 A (B). This is performed by merging the end state of a component with the begin state of the next one, and greatly reduces the complexity of the model.

components, i.e. domains and linkers, were first modelled separately using HMMs (Fig. **2A**) and concatenated into the model shown in (Fig. **2B**).

A family profile assumes that each column in multiple alignments is independent from the others. As already shown, these assumptions are adequate when studying properties of a single component, for example AT domains [17]. However, the interactions between individual components are likely to be crucial when distinguishing between successful and unsuccessful clusters. Descriptive statistical methods were used to adapt the neighbourhood structure of the model so that the interdependence between various sites within a single component as well as between sites from different components within a functional module could be considered. (Fig. **3**) shows such a model with internal arrows showing correlations between sites within the 2nd and nth component as well as correlations between sites in the 1st component and sites in the 2nd and nth component. The model in (Fig. **3**) is, thus, capable of modelling interactions between different domains and linkers, even though the underlying HMM in (Fig. **2B**) was developed by considering all the components separately. The HMM in (Fig. **2**) are not capable of modelling interactions, because, by definition, the family profiles consider each column (i.e. amino acid position) to be independent.

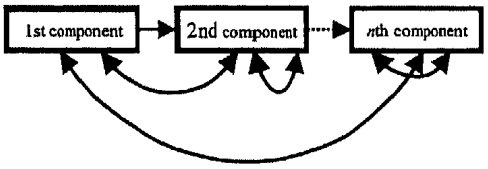

Fig. (3). The full model of a PKS module. Boxes indicate various components, while arrows describe the new neighbourhood structure and indicate correlations between sites within a single component, as well as sites from different components.

It is interesting to note that, in natural clusters, there are a small number of examples for reduction domains that are apparently non-functional, but there is still a good product yield. It is possible that such domains are necessary for the structural integrity of the protein even if they play no enzymatic role. This might make it difficult to predict the efficiency of a hybrid protein in terms of yield, although a reductive domain with a structural function should stand apart in our model from reductive domains with catalytic activity.

THE GENOME OF *STREPTOMYCES* AND EVOLUTION OF SECONDARY METABOLITE CLUSTERS

Streptomyces and related genera synthesise a plethora of biologically active secondary metabolites and any one strain can often synthesise several chemically unrelated compounds. *Streptomyces* are ubiquitous in soil habitats, forming a branched, mould-like vegetative mycelium that grows showing apical dominance following the germination of reproductive spores. The mycelium is multinucleated, but the spores themselves are haploid and are formed in chains by differentiation at the distal ends of aerial hyphae, in response to a number of environmental triggers. Differentiation is accompanied by the production of secondary metabolites [18]. Many secondary metabolites are active as antibiotics and it is likely that this activity confers a selective advantage in the soil environment. A novel antibiotic would be particularly advantageous so that frequency-dependent selection should result in the observed diversity of antibiotics. The structure of modular biosynthetic clusters lends itself to the generation of diversity. Duplication and deletion of modules accompanied by sequence divergence could lead to the evolution of new activities. In addition to such clonal mechanisms, it is also conceivable that recombination between clusters can generate new activities. Phylogenetic trees have been constructed for domains in PKS and NRPS clusters [19]. These analyses are consistent with both types of evolutionary process operating, but do not give a rigorous answer to these questions.

The genome sequences of two *Streptomyces* strains are in the public domain: the model organism *S. coelicolor* A3 (2) [20] and the avermectin producer *S. avermitilis* [21]. The genomes are large (8.7 Mb and 9.0 Mb respectively) with over 7600 protein-coding sequences, which is about twice the size and double the number of genes found in *Escherichia coli* [22]. The G+C content is also extreme, greater than 70%. The *Streptomyces* genomes, like those of fungi, are linear. There are terminal inverted repeats with primer proteins bound at the telomeres preventing loss of coding sequences during each round of replication [23]. A comparison of the genome of *S. coelicolor* A3(2) with those of the two pathogenic actinomycetes *Mycobacterium tuberculosis* and *Corynebacterium diphtheriae* [20] showed that most of the essential genes are present in blocks of synteny located within a central core of about 4.9 Mb. There are about 20 secondary metabolite clusters and most (about 75%) reside in the two arms (of about 1.5 Mb and 2.3 Mb) outside this core region. A comparison of the genomes of *S. coelicolor* and *S. avermitilis* [21] showed that the genetic structure in a 6.5 Mb central region is well conserved. However, there is little conservation of genes in the terminal arms. Thus, the genomes of the two sequenced *Streptomyces* seem to possess a biphasic architecture. Although most genes in the arms are probably non-essential in the laboratory they are likely to confer selective advantages to the organism in the natural environment [24]. At least 30% of the protein coding genes in each species do not have a homologue in the other species. Thus there is considerable genetic diversity in the genus *Streptomyces*.

Conjugation systems are very common in *Streptomyces* and have been used for constructing genetic maps [25]. Linkage analysis shows that large pieces of chromosome of several megabases may be transferred between strains. Furthermore unstable partial diploids called heteroclones may be formed in *S. coelicolor* [26, 27] and unstable apparently diploid clones called heterogenotes were observed in *S. rimosus* [28]. The clones eventually segregate as stable haploids. The physical basis of these phenomena is unknown, but it is conceivable that it may be associated with the multinucleoid compartments present in the mycelium. Many *Streptomyces* species show genetic instability affecting the ends of the linear chromosome [e.g. 29 – 31]. In both *S. ambofaciens* and *S. rimosus* [32 - 34] deletions were found that removed one chromosome end.

Closer analysis showed that two copies of the chromosome had recombined in inverted orientation to give a diploid chromosome with the deletion of some terminal sequences (Fig. 4). Therefore, it seems that duplication of the *Streptomyces* chromosome in inverted orientation might be a general phenomenon for this genus of bacteria. This is supported by the observation that there is relatively frequent recombination between linear replicons with the exchange of one replicon end with another [35, 36]. Genetic recombination within a particular species of *Streptomyces* is well established. There have also been attempts to generate recombinants between species using protoplast fusion [37, 38]. With the exception of recombinants between *S. coelicolor* A3(2) and *S. lividans* 66 [39], which are very closely related and taxonomically belong to the same species, there has not yet been a physical analysis to prove that genuine interspecies recombinants have been formed. This is a very interesting possibility that could be exploited for the generation of new products as well as a model for evolution of *Streptomyces* species. β-lactam biosynthesis genes in fungi were probably gained by horizontal transfer from actinomycetes [40]. Fungi and actinomycetes share many common features of morphology and habitat. It will be interesting to see whether further genome sequencing reveals a more intimate relationship at the genetic level than suspected up to now.

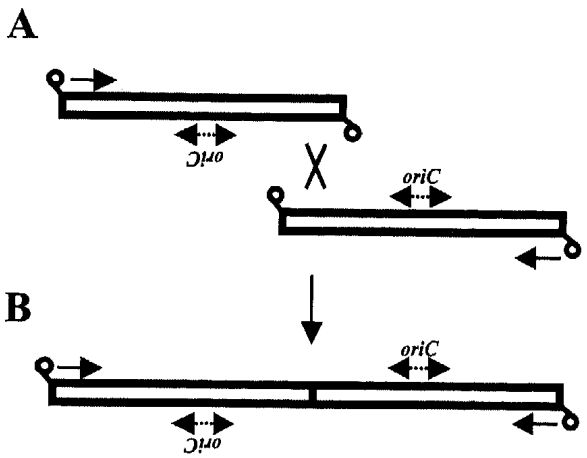

Fig. (4). Model for the formation of inverted fusion chromosomes in *S. rimosus* and *S. ambofaciens*. Two copies of the chromosome undergo a recombination in inverse orientation **(A)**, which generates a double chromosome with two origins of replication **(B)**. The terminal sequences, which are deleted, do not carry any known origin of replication although they may be retained at very low copy number in the mycelium of *S. rimosus*.

RECOMBINATION BETWEEN BIOSYNTHETIC GENE CLUSTERS

A single crossover between two linear DNA molecules will exchange the ends and generate two linear recombinant molecules. Such an exchange has been observed between the linear plasmid pPZG101 and the linear chromosome of *S. rimosus*, as well as between linear plasmid SCP1 and the linear chromosome of *S. coelicolor* [35, 36]. It has been suggested that recombination between the chromosome and a linear plasmid (Fig. 5) could allow convenient selection of recombinants between two modular biosynthetic clusters [41]. A major question is the efficiency of recombination between such sequences, which have a considerable degree of sequence divergence. Little is known of recombination pathways in *Streptomyces*. In Enterobacteria mutations in the mismatch repair genes *mutS* or *mutL* [42] or in the *recD* gene [43] have been shown to increase the efficiency of such recombination. The two *Streptomyces* genome sequences do not show the presence of homologues of these three genes. However, experiments in *S. roseosporus* suggested that it was fairly easy to isolate a mutant that gave increased recombination frequencies between diverged sequences [44].

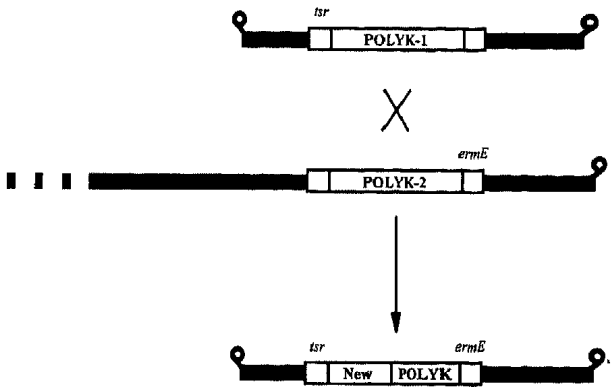

Fig. (5). Suggested strategy to obtain recombination between polyketide clusters in *Streptomyces*. One cluster (Polyk-1) should be cloned on a linear plasmid with an adjacent thiostrepton resistance gene (*tsr*). The other cluster (Polyk-2) should be cloned near one end of the linear chromosome with an adjacent erythromycin resistance gene. A single cross over between the clusters to generate the novel polyketide gene-cluster (New Polyk) can be selected by transfer of the recombinant plasmid carrying both resistance markers to a second *Streptomyces* strain.

As homologous recombination should favour recombination sites that show little sequence divergence, it is likely that recombination between clusters will be strongly biased to more closely related modules. It is conceivable that such recombinants will suffer fewer problems of low productivity than recombinants generated by *in vitro* manipulation to fuse more distantly related modules. The modelling of recombination between clusters *in silico* should reveal the spectrum of potential products that can be produced by recombination and also provide a framework for evaluating the evolution of natural clusters. A fundamental question in such modelling is the choice of recombination sites. Any recombination between clusters involves homeologous sites i.e. homologous non-identical sequences. The parameters governing selection of recombination sites are not fully understood in any system and there is little systematic work in *Streptomyces*. This means that the modelling results will only be useful if they are relatively insensitive to the exact parameters used to define the recombination sites. The consequence of recombination between two modules will depend on the site of recombination. The order of domains in PKS and NRPS modules is normally: KS-AT-reduction domains (if present)-ACP and C-A-PCP-modifying domains (if present), respectively (Fig. 6). As the AT- and A-domains determine the nature of the extender unit (e.g. if a C_2 or a C_3 unit is added in PKS systems), recombination events in the AT-

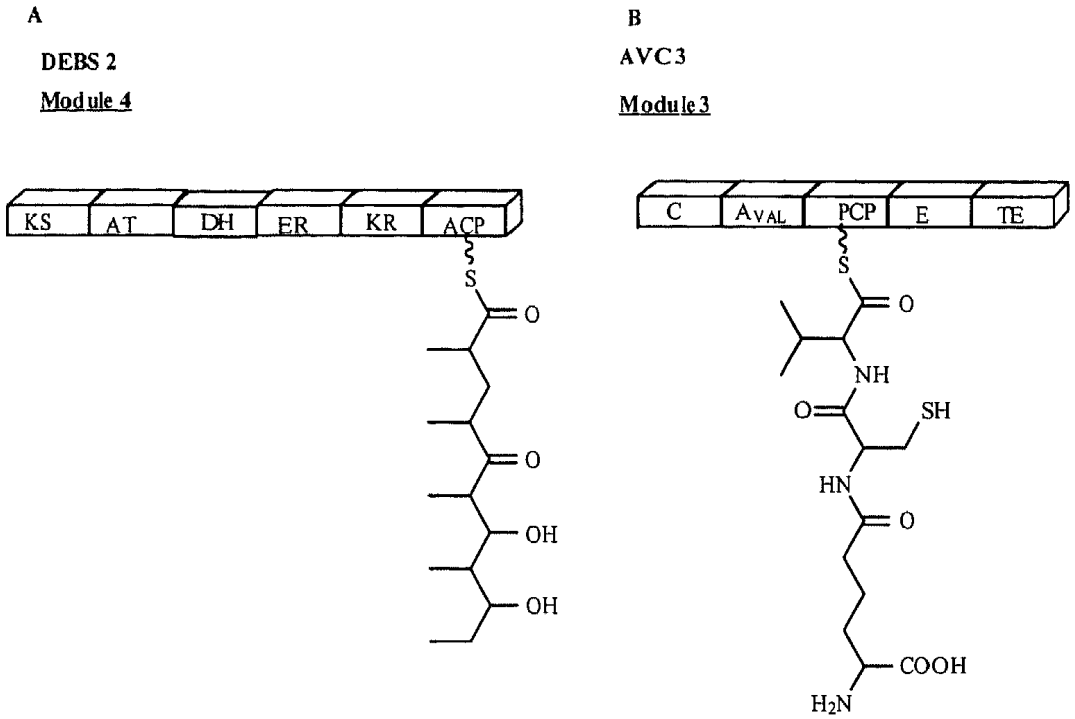

Fig. (6). Schematic presentations of DEBS 2 modull 4 from erythromycin A polyketide gene-cluster (A) and of AVC 3 modull 3 from penicillin non-ribosomal peptide gene-cluster (B). Linear polyketide and peptide products are shown.

and A-domains may well produce a novel function in comparison with the two parental modules. Recombination in the reduction and modification domains may also create a new product, as the domains may not cope with a novel substrate. In contrast, recombination events in the KS (C)- or ACP (PCP)-domains may not change the activity of the module.

In order to model homologous recombination it is necessary to define criteria for selecting recombination sites. The BLAST algorithm for similarity searching [45] suggests a convenient way of doing this. High scores can arise from relatively short sequences that have nearly perfect matching or longer sequences with more mismatches. In the absence of more detailed knowledge of the selection of recombination sites, it seems reasonable to model recombination by assuming that sequences that achieve good enough BLAST scores will undergo recombination and different BLAST parameters can be used to model different assumptions about the degree of matching needed to allow homologous recombination. Computer programs were written [46, 47] to model recombination between PKS clusters. This showed that most of the recombination events (> 90%) would occur in the KS-domains and changes in the BLAST parameters had little effect on the pairs of modules that could undergo recombination. This simplifies the analysis, because, as noted above, the recombinant module should retain the activity of one of the parents. This analysis ignores the question of stereochemistry during the KS reaction, but it is plausible that the most similar modules that undergo recombination will also show the same stereochemistry.

The recombination model can produce the predicted DNA sequences of recombination events. It is also useful to predict the products of the recombinant clusters, because different recombinants may produce the same product. The main difficulty in this exercise is that the polypeptides involved in many biosynthesis clusters are not encoded co-linearly. The clusters often have PKS genes on both DNA strands. Although the modules are used in order in each polypeptide, the order of encoding the polypeptides in the DNA is often different from the order they are used in the enzyme. A description of the clusters was developed that includes information about the order of use of the polypeptides in the enzyme [47] and this was used to develop an algorithm that predicts the polyketide structures that are produced by recombinant clusters. The approach above only produces linear polyketide products. However, in practice, most such products will undergo a spontaneous chemical reaction to form ring structures. In principle, it is possible to use standard chemical computation methods to generate three-dimensional structures that can be used as a virtual compound library for *in silico* screening. A virtual library of structures that could be potentially generated by *in vitro* genetic manipulation has been used for *in silico* screening [48]. However, it is not yet clear how accurate the prediction of the final chemical structures of polyketides will be.

Most naturally occurring polyketides undergo modification after the synthesis of the polyketide backbone. In particular, glycosylation is very common. It is not known how often the modifying enzymes present in PKS clusters will recognise the novel substrates produced by recombination and this is likely to be an intractable problem.

CONCLUSIONS

Current attempts to reengineer modular biosynthetic clusters have usually resulted in very low product yield, which makes them unsuitable for any commercial development. It seems unlikely that traditional strain improvement of such modified clusters will result in acceptable yield, because very specific multiple mutation

steps within affected modules will be necessary. Although novel products have also been obtained by precursor-directed feeding of exogenous substrates, including expression in "superhosts" [13], any general method to produce new structures is likely to involve redesigning the junctions of the modified regions. It is conceivable that the generation of modified clusters using homologous recombination would favour biologically active junctions. Other approaches are to try to gain a better understanding of the architecture of the clusters. It has been suggested that the problems might be solved by a better understanding of the linker and docker sequences coupling domains and modules. This approach is limited, because of the extreme diversity of these sorts of sequences so that there is unlikely to be a universal solution. An alternative approach is to use hidden Markov models to learn the rules applied in "successful" natural clusters. This will probably result in a redefinition of the units involved different from the present domain/module/cassette architecture. It seems possible that interactions between altered intermediates and functional domains might also be a major factor in reduced product yield. However, this scenario has yet to be addressed seriously and seems unlikely to yield solutions in the absence of knowledge of three-dimensional structure of the synthetases.

Streptomyces produce more antibiotics than any other taxonomic group of microorganism, with World markets worth upwards of $25.2 billion in 2001 and projections forecast to reach nearly $28 billion by 2007. Powerful genomic and post-genomic technologies may cast light in the future as to why these bacteria are such prolific producers of blockbuster drugs. The future looks promising, new antibiotics and other drugs could be made by a number of genetic engineering strategies, which when taken together we refer to as combinatorial biosynthesis [13, 14]. Using DNA and protein sequence information from *Streptomyces,* we are developing *in silico* systems to predict the likely function of novel hybrid proteins created by homologous recombination between chromosomes. Chromosome behaviour and genetic structure of *Streptomyces,* as well as morphology, ecology and metabolic potential seem more akin to fungi than bacteria. We suggest this could represent an earlier state in the evolution of diploidy, perhaps found in the ancestors of modern *Streptomyces.* This departs from most theories of the origin of eukaryosis [e.g. 49, 50] that suggest such chromosome behaviour occurred first in the haploid condition followed by diploidy; clearly it is an issue that warrants closer scrutiny. Perhaps one day development of genetic recombination and protein expression methods between very biologically diverse organisms might be possible [51], which could be exploited as a new tool in natural products lead discovery.

REFERENCES

[1] Beckman, J.A.; Ames, B.N. *Phsiol. Rev.*, **1998**, *78*, 547.
[2] Cohen, M.L. *Nature,* **2000**, *406*, 762.
[3] Evans, W.E.; Relling, M.V. *Nature,* **2004**, *429*, 464.
[4] Drews, J. *Science,* **2000**, *287*, 1960.
[5] Rondon, M.R.; Goodman, R.M.; Handelsman, J. *Trends Biotechnol.*, **1999**, *17*, 403.
[6] Watve, M.G.; Tickoo, R.; Jog, M.M.; Bhole, B.D. *Arch. Microbiol.*, **2001**, *176*, 386.
[7] Khosla, C.; Keasling, J.D. *Nat. Rev. Drug Discov.*, **2003**, *2*, 1019.
[8] Cane, D.E.; Walsh, C.T.; Khosla, C. *Science,* **1998**, *282*, 63.
[9] Cane, D.E.; Walsh, C.T. *Chem. Biol.*, **1999**, *6*, R319.
[10] Staunton, J.; Weissman, K.J. *Nat. Prod. Rep.*, **2001**, *18*, 380.
[11] Stein, T.; Vater, J.; Kruft, V.; Otto, A.; Wittmann-Liebold, B.; Franke, P.; Panico, M.; McDowell, R.; Morris, H.R. *J. Biol. Chem.*, **1996**, *271*, 15428.
[12] Schwarzer, D.; Finking, R.; Marahiel, M.A. *Nat. Prod. Rep.*, **2003**, *20*, 275.
[13] Kumar, P.; Khosla, C.; Tang, Y. *Methods Enzymol.*, **2004**, *388*, 269.
[14] Finking, R.; Marahiel, M.A. *Ann. Rev. Microbiol.*, **2004**, *58*, 453.
[15] Broadhurst, R.W.; Nietlispach, D.; Wheatcroft, M.P.; Leadlay, P.F.; Weissman, K.J. *Chem. Biol.*, **2003**, *10*, 723.
[16] Durbin, R.; Eddy, S.; Krogh, A.; Mitchison, G. *Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids*, Cambridge University Press: Cambridge, New York, Port Melbourne, Madrid, Cape Town, **2001**.
[17] Yadav, G.; Gokhale, R.S.; Mohanty, D. *J. Mol. Biol.*, **2003**, *328*, 335.
[18] Flärdh, K. *Curr. Opin. Microbiol.*, **2003**, *6*, 564.
[19] Lopez, J.V. *Mol. Genet. Genomics,* **2003**, *270*, 420.
[20] 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.; Hopwood, D.A. *Nature,* **2002**, *417*, 141.
[21] Ikeda, H.; Ishikawa, J.; Hanamoto, A.; Shinose, M.; Kikuchi, H.; Shiba, T.; Sakaki, Y.; Hattori, M.; Omura, S. *Nat. Biotechnol.*, **2003**, *21*, 526.
[22] Blattner, F.R.; Plunkett, G. 3rd; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.; Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.; Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao, Y. *Science,* **1997**, *277*, 1453.
[23] Bao, K.; Cohen, S.N. *Genes Dev.*, **2003**, *17*, 774.
[24] Challis, G.L.; Hopwood, D.A. *Proc. Natl. Acad. Sci. USA,* **2003**, *100* Suppl *2*, 14555.
[25] Hopwood, D.A. *Microbiology,* **1999**, *145*, 2183.
[26] Sermonti, G.; Mancinelli, A.; Spada-Sermonti, I. *Genetics,* **1960**, *45*, 669.
[27] Hopwood, D.A.; Sermonti, G.; Spada-Sermonti, I. *J. Gen. Microbiol.*, **1963**, *30*, 249.
[28] Hranueli, D.; Pigac, J.; Smokvina, T.; Alacević, M. *J. Gen. Microbiol.*, **1983**, *129*, 1415.
[29] Volff, J.N.; Altenbuchner, J. *Mol. Microbiol.*, **1998**, *27*, 239.
[30] Fischer, G.; Wenner, T.; Decaris, B.; Leblond, P. *Proc. Natl. Acad. Sci. USA,* **1998**, *95*, 14296.
[31] Chen, C.W.; Huang, C.H.; Lee, H.H.; Tsai, H.H.; Kirby, R. *Trends Genet.*, **2002**, *18*, 522.
[32] Gravius, B.; Bezmalinović, T.; Hranueli D.; Cullum, J. *Appl. Environ. Microbiol.*, **1993**, *59*, 2220.
[33] Denapaite, D.; Paravić Radičević, A.; Cajavec, B.; Hunter, I.S.; Hranueli, D.; Cullum, J. *Food Technol. Biotechnol.*, **2005**, *43*, 9.
[34] Wenner, T.; Roth, V.; Fischer, G.; Fourrier, C.; Aigle, B.; Decaris, B.; Leblond, P, *Mol. Microbiol.*, **2003**, *50*, 411.
[35] Pandza, S.; Biuković, G.; Paravić, A.; Dadbin, A.; Cullum, J.; Hranueli, D. *Mol. Microbiol.*, **1998**, *28*, 1165.
[36] Yamasaki, M.; Kinashi, H. *J. Bacteriol.*, **2004**, *186*, 6553.
[37] Ikeda, H.; Inoue, M.; Tanaka, H.; Omura, S. *J. Antibiot.*, **1984**, *37*, 1224.
[38] Qi, H.Y.; Zheng, Y.X. *Chin. J. Biotechnol.*, **1990**, *6*, 139.
[39] Wang, S.J.; Chang, H.M.; Lin, Y.S.; Huang, C.H.; Chen, C.W. *Microbiology,* **1999**, *145*, 2209.
[40] Coque, J.J.; Martin, J.F.; Calzada, J.G.; Liras, P. *Mol. Microbiol.*, **1991**, *5*, 1125.
[41] Hranueli, D.; Cullum, J. In *Current Studies of Biotechnology: Food*; Z. Kniewald *et al.*, Eds.; Croatian Society of Biotechnology: Zagreb, Croatia, **2003**, Vol. *3*, pp. 333-340.
[42] Rayssiguier, C.; Thaler, D.S.; Radman, M. *Nature,* **1989**, *342*, 396.
[43] Zahrt, T.C.; Maloy, S. *Proc. Natl. Acad. Sci. USA*, **1997**, *94*, 9786.
[44] Hosted, T.J.; Baltz, R.H. *Microbiology,* **1996**, *142*, 2803.

30. Juni 2005

[45] Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W.; Lipman, D.J. *J. Mol. Biol.*, **1990**, *215*, 403.

[46] Tupath, H.; Pfeiffer, J.; Pfeifer, I.; Deckbar, D.; Fleige, T.; Peitz, H.; Cullum, J. In: Proceedings of the 4th Croatian Congress of Food Technologists, Biotechnologists and Nutritionists (Central European Meeting); V. Lelas et al., Eds.; Faculty of Food Technology and Biotechnology, University of Zagreb: Zagreb, **2002**, pp. 291-295.

[47] Peitz, H. Simulation homologer Rekombination von modularen Biosynthase Genclustern. **2004**. *Diplom thesis, University of Kaiserslautern.*

[48] Siani, M.A.; Skillman, A.G.; Carreras, C.W.; Ashley, G.; Kuntz, I.D.; Santi, D.V. *J. Mol. Graph. Model.*, **2000**, *18*, 497.

[49] Cavalier-Smith T. *Ann N Y Acad Sci.* **1987**, *503*, 17.

[50] Margulis L.; Dolan M.F.; Guerrero R. *Proc. Natl. Acad. Sci. USA*, **2000**, *20*, 6954.

[51] Bode, H.B.; Muller, R. *Plant Physiol.*, **2003**, *132*, 1153.