Available Pathways Database (APD): An Essential Resource for Combinatorial Biology

Michael C. Pirrung,*,† Chris M. Silva,‡ and John Jaeger§

Department of Chemistry, Levine Science Research Center, Duke University, Durham, North Carolina 27708-0317, ChromaXome Corporation, TerraGen Discovery, Inc., 300-2386 East Mall, Vancouver, British Columbia, V6T 1Z3, Canada, and Trega Biosciences, Inc., 9880 Campus Point Drive, San Diego, California 92121

Received May 17, 2000

A relational database, the Available Pathways Database (APD), has been constructed of microbial natural products, their producing strains, and their biosynthetic pathways. The database allows the ready selection of donor strains for combinatorial biology experiments. It provides the same type of resource for combinatorial biology as the Available Chemicals Directory (ACD) does for combinatorial chemical library generation. Its cataloging ability can also provide insight into novel aspects of biosynthetic routes. In particular, no 10-unit Type I polyketides were found in the compilation of this edition of the APD (Version I).

Medicine currently faces serious problems in infectious diseases. At one time it may have been believed that antibacterial research could be de-emphasized since treatments for many bacterial infections were available,[1] but the emergence of antibiotic-resistant strains of pathogens in hospital-acquired infections (methicillin-resistant *Staphylococcus aureus* (MRSA),[2] vancomycin-resistant *Enterococcus* (VRE),[3] and other multidrug-resistant bacteria[4]) and the increased incidence of mycobacterial diseases such as tuberculosis[5] (due both to drug-resistant strains and increasing numbers of immunocompromised patients) have re-energized antibacterial research. While microbial genomics of pathogens offers an avenue to novel targets for antibiotic development,[6] chemical diversity for infectious diseases can arise from the recent advances in synthetic library methods or from equally impressive advances in diversity-oriented natural products research. Considerations favoring natural products as a source of chemical diversity in antibiotic testing include their functional role in microbial ecology and their ∼70% greater structural diversity, as evidenced by a recent statistical comparison of natural products versus available synthetic compounds.[7] Amplification of the already superior diversity of natural products can be accomplished by harnessing the full genetic potential of antibiotic-producing microorganisms. The cloning, engineering, and transgenic expression of complex biosynthetic pathways can lead to structurally novel and complex secondary metabolites.[8] Great strides have been made in the analysis of large genetic loci that encode pathways to antibiotics, such as the macrolides and anthracyclines, and in manipulation of the cognate synthases. This emerging technology of combinatorial biology has the potential to radically alter the search for natural product drug leads for the pharmaceutical industry.

Combinatorial biology concerns cloning of the gene clusters that code for the biosynthesis of secondary metabolites and expressing these DNAs in an alternative host. Complex loci can be isolated from known organisms, uncultivable microbes, or difficult-to-ferment isolates. This process of expressing pathways isolated from diverse sources in a panel of well-characterized hosts obviates the need for the arduous process of optimizing idiosyncratic growth conditions for each antibiotic-producing strain. Perhaps the most exciting prospect for combinatorial biology is that it is possible to combine genes from different biosynthetic pathways and express them within the same cell, thus creating entirely novel structures that can be screened for biological activity. Their increased structural diversity and novelty should permit the discovery of new modes of action in biologically active natural products.

The combinatorial approach to biosynthetic diversity relies upon the fact that, although natural products show a great deal of structural diversity, almost all of them are generated using a modest number of similar classes of biosynthetic pathways. In many cases it appears that the enzymes involved in the biosynthesis of these secondary metabolites show a considerable degree of promiscuity in the range of substrates they will accept. The similarities between these biosynthetic pathways are such that it is possible to substitute certain modules from distinct pathways, creating novel biosynthetic pathways that produce novel natural products. This approach has been well demonstrated.[9]

The vast variety of natural compounds available for human health care is due in large part to the biosynthetic versatility of the Actinomycetes. These bacteria have provided many thousands of structurally diverse low molecular weight chemicals that are currently being exploited in both medicine and agriculture. Many of the regulatory networks involved in antibiotic biosynthesis are often unique to the Actinomycetes and are conserved throughout the group.[10] This facilitates heterologous expression of complex biosynthetic clusters among Actinomycetes. Furthermore, if (1) all the genes necessary for the biosynthesis of a product are clustered at a single locus; (2) the enzymes involved in these biosynthetic pathways are able to recognize a variety of substrates; and (3) the mechanisms of global regulation of antibiotics are conserved throughout the Actinomycetes group, it should be possible to isolate, clone, and recombine different biosynthetic loci to create novel natural products. The clustering of biosynthetic pathways at a single locus has been demonstrated repeatedly for almost all biosynthetic pathways studied in Actinomycetes.[11] This clustering allows the simple cloning and

* To whom correspondence should be addressed. Fax: (919) 660-1591. E-mail: pirrung@chem.duke.edu.
† Duke University.
‡ ChromaXome Corporation.
§ Trega Biosciences.

10.1021/np000244x CCC: $19.00
Published on Web 09/22/2000

expression of complete pathways as single clones in an appropriate heterologous host. The lack of substrate specificity for many biosynthetic enzymes has been demonstrated both directly and indirectly over the past few years; this observation permits the mixing of genes from distinct or related pathways with a reasonable expectation that the biosynthetic enzymes will have enough substrate flexibility to recognize and catalyze anabolic reactions on the intermediates and products from other biosynthetic systems. The conservation of global regulators of antibiotics among the actinomycetes has only recently been investigated directly.[12] However, there have been many studies in which biosynthetic pathways have been expressed without significant modification in a heterologous host to improve expression. These experiments demonstrate indirectly the similarities between actinomycete species in the global regulation of antibiotics. Combinatorial biology relies upon the expression host's own regulatory elements being functionally similar enough to those in the original host to allow expression of the heterologous biosynthetic pathway.

A key raw material for combinatorial biology is DNA that encodes biosynthetic pathways, which can arise from the vast culture collections that have been compiled over many decades. Just as it is important for the development of synthetic chemical pharmaceuticals to access diverse starting materials for combinatorial chemical libraries, it is essential to the success of combinatorial biology to have access to a diverse and high-quality biosynthetic pathway DNA collection. It is therefore necessary to analyze and classify/catalog DNA samples based on known biosynthetic pathways. Any DNA sample can be analyzed for a biosynthetic pathway (such as polyketide, terpenoid, or aminoglycoside) present in it based on gene probes. For many sources of DNA though, it is not necessary to probe for pathways because the biosynthetic potential contained within a DNA sample is already signaled by the natural product(s) that its host organism produces. It is only necessary to analyze their chemical structure(s) based on paradigmatic biosynthetic pathways and whatever literature information is available regarding pathways in a specific organism.

An initial compilation of microbial strains and their biosynthetic pathways has been developed in this work in the form of the Available Pathways Database (APD), which provides essential bioinformatics support for combinatorial biology experiments. The APD provides the same type of resource for combinatorial biology that the Available Chemicals Directory (ACD)[13] provides for combinatorial chemistry. For the biosynthetic pathway to a secondary metabolite to be *available* (in the form of DNA), the strain producing it should come from a public domain culture collection. For the purposes of the compilation of this database, a *pathway* may represent all of the DNA (and of course the derived enzymes) needed to make a *particular* biosynthetic intermediate from primary metabolites (e.g., glucose → *p*-hydroxyphenylpyruvic acid by the shikimate pathway), or it could represent a general capability not bearing further delineation, such as the glycosylation of a macrolide.

Results and Discussion

Information concerning available microorganism strains, the compounds they produce, and the steps/intermediates involved in their biosynthesis were collected in a relational database in which it is possible to access any one part and find all others through well-constructed queries. The structure of this database, the APD, is given in Figure 1.

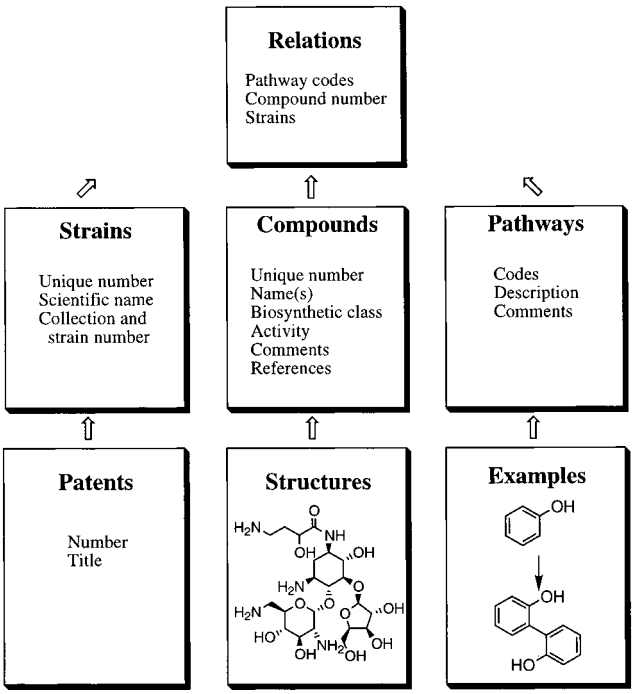


Figure 1. Database structure for the Available Pathways Database (APD).

Initial work involved searching the electronic version of the *Dictionary of Natural Products* (DNP)[14] for compounds of known structure from Actinomycetes, a group known to produce a significant number of structurally varied and complex secondary metabolites with diverse biological activities. The DNP is not limited to biologically active compounds, which might work to advantage in the long term, since the goal of combinatorial biology is first to access novel and diverse structures, not merely to rehash known antibiotics. The results are given in Figure 2. The dominance of *Streptomyces* is obvious and interesting, which may be related to its ease of culture. However, because the DNP does not generally list the producing strain or collection for its entries, it does not serve as a useful entrée to available strains.

Another information source was the electronic version of the American Type Culture Collection (ATCC) catalog, from which were identified 949 strains producing at least 680 different compounds in classes such as polyethers/ionophores, macrolides, ansa macrolactams, polyenes (Type I polyketides), anthracyclines (Type II polyketides), β-lactams, aminoglycosides, nucleosides, and enediynes, with very diverse biological activities, including anti-Gram(+), anti-Gram(−), antimycobacterial, antifungal, antiviral, antitumor, antiprotozoan, anticoccidial, antimycoplasmal, antihelmintic, immunosuppressant, and even antimosquito activities (i.e., many of the known biological activities of secondary metabolites).

The final information source was the U.S. patent literature from 1975 to 1998. Over 1700 patents were identified using the key word *Streptomyces*, and over 750 patents were identified using the non-*Streptomyces* Actinomycetes. Copies of all of these patents were obtained, and several pieces of information for each compound were abstracted from them: strain collection, biosynthetic class, and structure. One advantage of this strategy is that strains producing patented secondary metabolites must be deposited in a culture collection, which provides the ability to locate via patents strains that may not appear in the general catalog of a collection.

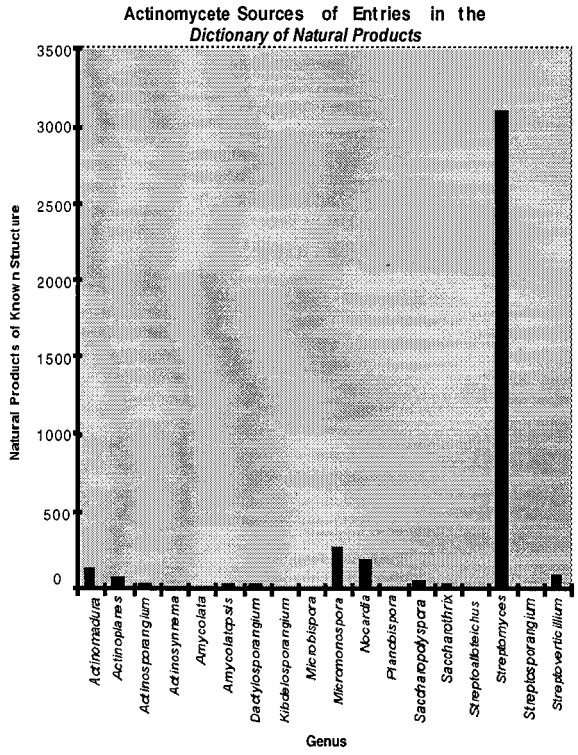


Figure 2. Entries in the *Dictionary of Natural Products*[21] derived from Actinomycetes.

With these raw materials for the compilation of the database in hand, a protocol was developed for making entries to it, particularly for the analysis of natural products to construct the pathways. This can be accomplished by examining the structures of compounds produced by available strains and listing most of the steps (putatively, since almost all certainly have not been studied, yet rationally, based on known biosynthetic paradigms) involved in the biosynthesis of each. It was important to form classes of transformations that were neither too broad or narrow. Excessive specificity would make each biosynthetic process unique. Lax specificity would group functionally unrelated strains, genes, and enzymatic activities (vide infra).

An example will serve to illustrate this process. Chromomycin A_3[15] is a cancerostatic isolated from *Streptomyces griseus* ssp. *griseus* and *Streptomyces cavourensis* ssp. *washingtonensis*. Its subclass of the Type II polyketides is known[16] to arise from a decaketide precursor **B**, folded and cyclized as shown to **A**, which is cleaved (likely by a dioxygenase) and decarboxylated to give the general olivomycin skeleton seen in chromomycin.[17] Other steps needed to assemble this molecule include *C*-methylation (methyl groups found on the aromatic nucleus of Type II polyketides generally arise not from propionate extender units but from *S*-adenosyl methionine methylations), hydroxylation of the polyketide (whenever a carbon that did not arise from C-1 of an extender unit is oxidized, postsynthesis oxidation likely occurred), glycosyl transfer to the polyketide, glycosyl transfer to a carbohydrate, *O*-methylations on the polyketide and the sugars, and acylation on the sugars. These intermediates (**B**) or processes can be encoded as a string of mnemonics [pks210, fld001, omt007, gtx006, omt002, gtx004, act002, hox006, dec001, cmt002]. The meaning of these particular mnemonics is cataloged in Table 1. While certainly not specifying all of the steps in the production of as complex a molecule as chromomycin, and while saying nothing about the order in which the steps occur, this string does reflect some of the biosynthetic potential contained in the genome of this *Streptomyces griseus* strain that can be exploited for combinatorial biology.

Table 1. Mnemonics for Some of the Steps in the Biosynthesis of Chromomycin.

pks210	polyketide synthase, Type II, 10 acetate units
fld001	polyketide fold, mithramycin/tetracenomycin-like
omt007	*O*-methyl transferase to a carbohydrate
gtx006	glycosyl transferase to a carbohydrate acceptor
omt002	*O*-methyl transferase to a Type II polyketide
gtx004	glycosyl transferase to Type II polyketide acceptor
act002	acyl transfer to a carbohydrate acceptor
hox006	hydroxylation of a Type II polyketide
dec001	decarboxylation of a Type II polyketide, as shown in **A**, Figure 3
cmt002	*C*-methyl transferase to a Type II polyketide

This analysis of natural product structures and encoding of biosynthetic pathways may be conducted with a good but not encyclopedic knowledge of biosynthetic pathways, such as arises from reading any of the treatises in the area.[18] We thereby identified >200 pathways and/or biosynthetic intermediates and >400 natural products whose biosynthetic pathways, in the form of DNA, are available from public domain collections such as ATCC, NRRL, and others. The general classes of pathways that were identified are as follows: Type I polyketide synthase, Type II polyketide synthase, nonribosomal polypeptide synthase, shikimate, monosaccharide synthase, amino acid synthase, mevalonate, methyltransferase, glycosyltransferase, acyltransferase, hydroxylase, epoxidase, chloroperoxidase, oxidases, and miscellaneous transformations.

The definition of these mnemonics, or "license plates", incorporated when possible specific knowledge of the transformation(s) involved or their genetics. Polyketides were identified on the basis of the total number of units, that is, a starter unit plus extender units, regardless of whether they were acetate, propionate, butyrate, or others. In addition, in the Type II (aromatic) polyketides (which

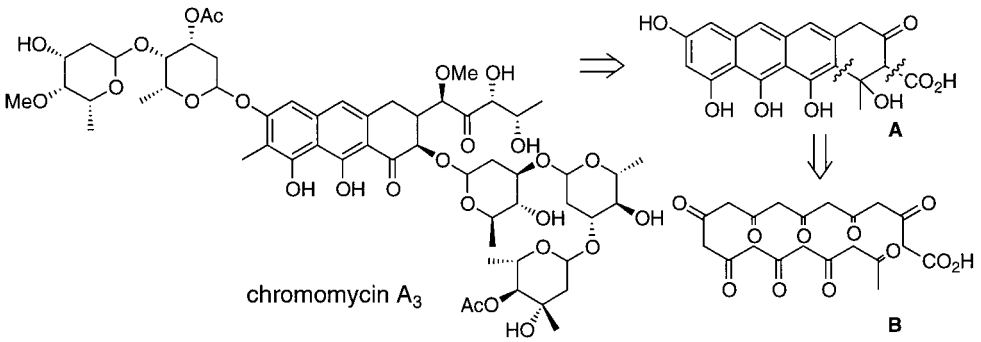


Figure 3. Example biosynthetic pathway analysis. Chromomycin A3 is derived from a putative tetracyclic precursor A, which can be formed by cyclization of polyketide B.

generally involve only acetate), it was necessary to specify the specific folding of the putative polyketide in preparation for its polycondensation/cyclization, in concert with research showing that specific genetic loci determine the folding of Type II polyketides.[19] Many transformations were classified by the natural product substrates operated upon. In addition to substrate, methyl transferases were classified based on the atom methylated, which show functional/genetic differences.[20]

Structure. The database was constructed using Accord for Access as shown in Figure 1, with a Relations table linking the compounds to the strains that produce them and the pathways that are involved. Because of software requirements, the chemical structures of the compounds are housed in a separate Structures table linked to the Compounds table. Likewise, chemical structures or reactions illustrating some pathways are housed in the Examples table separate from Pathways. Some fields within a table may be empty; it is common that the activity of a compound is unavailable, for example.

For each compound, there may be many strains that produce it, many patents on it, and many steps involved in its biosynthesis. Likewise, each patent may refer to several compounds or several strains, and each strain may produce several different compounds. This is evident from examining some of the "flat" file (spreadsheet) compilations that were used as raw material for the APD. As the database is currently constituted, some of that information was omitted in order to avoid repetitive data entry for each patent or strain. For each compound analyzed, however, at least one strain has been linked to it through the Relations table.

Conventions. Names. Names of compounds are problematic, as different uses prevail over time, and in some cases compounds were isolated by different groups and given different names. While all known names are listed in this table, one name must be chosen to appear first and therefore be most readily identified from alphabetical lists. If the current name of interest is not immediately found alphabetically, a Find function can be applied to the Name(s) field in the Compounds table to discover its listing as a synonym. Alphanumeric names are also common. For the most part, general and noninformative terms in these names such as "factor", "antibiotic", and "substance" were eliminated from the Compounds table. Because there is wide variation in punctuation among alphanumeric compound names, even in different publications concerning the same compound, compounds such as "A-963" may be found under that name as well as A_963 and A963. The most prevalent usage was generally adopted, though a wise future policy may be to omit all nonessential punctuation. The use of unique identifiers, such as the *Chemical Abstracts* registry number, was considered, though this information is generally not available from the natural products databases used to initiate this program. Listings of closely related compounds can entail an individual entry for each member of a compound group (67-121 A, 67-121 B, etc.), or the compound name can be given followed by a string of alphanumeric descriptors, and entries to the Relations table are needed only for the first example of a compound of a given group. This is no disadvantage for the intended use of the APD, as often many or all of the members of a compound group are isolated from the *same* strains, which bear many or all of the same pathways.

Pathways. For the purposes of the compilation of this database, a *pathway* represents all of the DNA (and of course the derived enzymes) needed to make a *particular* biosynthetic intermediate from primary metabolites (e.g., glucose → *p*-hydroxyphenylpyruvic acid by the shikimate pathway), or it could represent a general capability not bearing further delineation, such as the glycosylation of a macrolide. The mnemonics (which we colloquially call "license plates") encoding pathways are six characters in length, beginning with three letters incorporating a mnemonic of the particular process/compound. They are followed by numbers that can also encode information [i.e., pks115 means polyketide synthase, Type I, 15 units (initiator + extenders)], or they may simply reflect different subclasses of processes such as acyl transfer, numbered in the order that they were identified during database compilation. These mnemonics are readily perceived by an examination of the Pathways table. These *are all unique*, as they are the primary key for the Pathways table.

Database Accretion. The database enforces referential integrity among the three tables that are linked to Relations, which are Compounds, Pathways, and Strains. Each of these tables has enforced referential integrity between itself and its linked tables: Compounds⇔Structures, Strains⇔Patents, and Pathways⇔Examples. Therefore, entries in the Compounds, Strains, and Pathways tables *must* be present before Relations can be established. For the most part, the Pathways table contains all of the already identified steps in the biosynthesis of the major natural product classes. The formulation of new license plates may be required if a novel structure is added. Their addition to the Pathways table may be done directly. Single new compounds may be entered directly into the Compounds table. For additions to the tables Examples and Structures, forms with windows for chemical structure depiction were developed. These Forms (nuStructures and vuExamples) scroll through the tables, displaying structures, and permit cutting and pasting of structures. Structures are readily added to the database by simultaneously running a chemical drawing program and alternating between it and Access. Some of the Examples are chemical reactions rather than structures, and many chemical drawing programs are unable to communicate with Accord about reactions. However, it has a Reaction Builder facility that enables reactions consisting of one starting material and one product to be readily encoded. The nuStructures form accesses the Compounds table; once a compound is located, its structure is pasted into the form and added to the Structures table.

Queries are also used to add to the APD. The query qryStrn/Pat simultaneously makes additions to the Patents and Strains tables so that the link between them is automatically established. This query provides a dynaset of all Strains and Patents, which is also an updatable query. Additions to this dynaset table update both Strains and Patents tables and create the required link between them.

New information can be added to the database in large blocks, as from a spreadsheet, using Append queries. These can make additions to appropriate fields in the Compounds, Strains, and Patents tables. With sophisticated query design, unique entries can be assured (i.e., spreadsheets often have multiple entries for a single Patent or Strain because that is required by their structure), leading and trailing spaces can be eliminated, and "grouped" fields [ATCC 0000 → ATCC and 0000] can be separated.

The most important step in construction of the database is additions to the Relations table, which can be accomplished through frmRelations. This form simultaneously scans the Pathway, Strain, and Compound tables

and establishes the links when the correct entries are present.

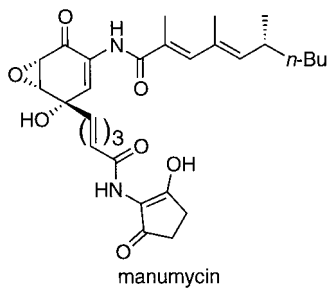
Database Interrogation. Queries have been developed to use the information in the APD. The query PathsOccurrence lists Pathways in decreasing order of their appearance in the database. This dynaset can be easily viewed through a filter to give ranked lists of different groups of transformations (i.e., polyketide synthases). The pathways that appear most frequently or least frequently is a point of interest. The latter constitute unique biosynthetic transformations that can provide highly unusual structures, while the former are most likely to lead to biosynthetic intermediates that can be utilized by other pathways. The query qryStrn/Path lists strains and the scientific name of all microbes in the database possessing a particular pathway, i.e., a given license plate from the Pathways table.

Electronic forms have also been developed to use the information in the APD. The form FrmCmpdSheet collects all of the information that has been entered for an individual compound so that it may be easily checked. Combination of the nuStructures form described above and the FindChemistry function of Accord enables natural products with particular structural elements to be found by interrogating the Structures table based on exact, substructure, or similarity criteria.

Other structure-based searches are created by choice of a *particular* structure and whether it will be used in an exact, substructure, or similarity query. Creation of an Accord query, which is then used in the creation of an Access query, delivers the information sought.

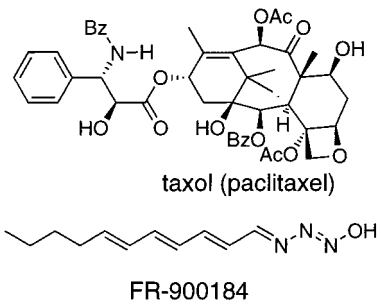
Status. The APD currently includes 934 entries in Strains, 1108 entries in Patents, 267 entries in Pathways, 241 entries in Examples, 965 entries in Compounds, 613 entries in Structures, and 1745 entries in Relations. Approximately 410 compounds have been analyzed at least partially and their pathways entered into the database.

Typical Use. One intended use of the APD is in selecting donor strains for directed combinatorial biology experiments. For example, perhaps a halogenated version of a nonribosomal polypeptide (NRPP) is desired, but is unknown in nature. A Boolean query combining a NRPP chloroperoxidase activity with strains producing the desired NRPP would provide a comprehensive strain list that could be further filtered by criteria such as known biological activities or taxonomy. Combining mnemonics for Type I polyketide glycosyltransferases with Type I polyketide synthases in a Boolean database query would list all strains that produce a polyketide that could be a substrate for the glycosyl transferases in other strains. Another approach is to identify structurally novel or otherwise desirable units, such as the 2-acylamino-3-hydroxycyclopentenone moiety present in manumycin, and search for all microbes that produce it or all natural products in which it is found.



manumycin

Prospects. Creation of the APD has analyzed in a novel way the biosynthetic potential in microbial genomes. It has made manifest unrecognized traits of natural biosynthetic pathways that can enable the directed biosynthesis of novel natural products that of necessity will constitute new chemical entities (vide infra) and may, therefore, possess novel biological activities. Particularly, such molecules may be unrecognized by extant pathways of resistance and may, thereby, circumvent antibiotic resistance. For example, it has been found that no 10-unit (1 initiator unit + 9 extenders units) Type I polyketides are known, at least in this *Streptomyces* database.



taxol (paclitaxel)

FR-900184

Figure 4. Natural product diversity.

Refinements to the APD that can be envisioned include the incorporation of data from specific feeding experiments that have been performed toward secondary metabolites, information on the genetic locus and structure of cloned biosynthetic genes, functional assignments of those genes (which will relate directly to pathway codes), and GenBank links to gene probes for particular pathways. A very small percentage of the biosynthetic pathways to secondary metabolites have actually been established experimentally, at the crudest level by feeding/incorporation experiments. When facts about the biosynthetic pathway to a secondary metabolite are known, this information may be included in the database. The rate at which biosynthetic clusters are being cloned is accelerating, so likewise, if the locus of the biosynthetic pathway is known, it will be directly referenced in the APD. Functional assignments for the different ORFs in a locus will come more slowly, but a number of these are already available in the polyketide area, and this information should be directly accessible in the APD.

We expect that the frequency of occurrence of pathways that are identified in the construction of the database will range from pervasive to unique. It is worth emphasizing that both have value. The pervasive pathways provide an opportunity to identify those strains whose pathways are most likely to intersect at key intermediates, providing true combinatorial biology. Those pathways may even provide taxonomic information on the host strains. The unique pathways provide entry to provocative natural product structures, such as the oxetane ring of a taxol, the hydroxytriazene functionality of a FR-900184 (Figure 4), or the trisulfide group of an esperamicin, which must reflect rather unique genes/enzymes that could confer a heretofore unknown structure and hence biological activity to the product of an existing pathway.

A potential application of the database beyond those so far delineated includes the ready identification of all microbes that can accomplish a particular biosynthetic transformation, e.g., the glycosylation of a macrolide, that could be used for custom biotransformation development.

It is notable that a structural class from which few representatives have so far been identified in this compilation is terpenoids. Certainly, many terpenoids are of plant, not microbial, origin, but fungal terpenes do occur naturally. There are many significant collections of fungi that

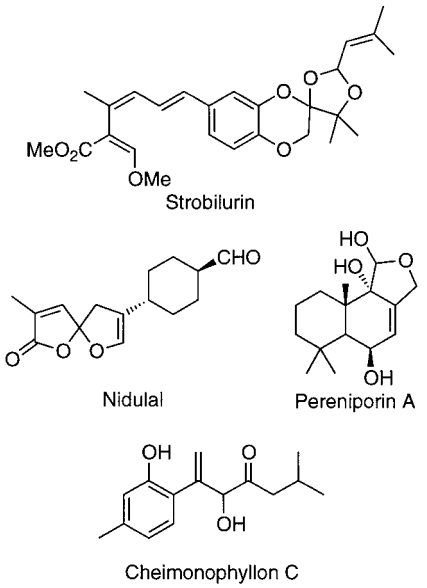


Figure 5. Terpenoid natural products from Basidiomycetes.

produce polyketides and other structural classes in addition to the terpenoids and may be highly useful to incorporate into the APD. Also worthy of mention are the Basidiomycetes, which produce some fascinating antibiotic compounds, examples of which are collected here (Figure 5). Certainly, the examples used in the creation of the database so far, microbial metabolites, are a narrow subset of the known natural products. Yet, as microbial DNA is also the most accessible, the database should be a superior resource for accessing this genetic diversity.

Experimental Section

The software used for the construction of the APD is Access97, a Windows-based relational database software program that is accessible to the basic personal computer user and has advanced SQL querying capabilities. In addition, the companion product Accord for Access97 was used to provide in the database the capability for chemical structure presentation and searching, including substructure searching. The resulting database contains natural products, pathways, and underlying DNA/cultures from culture collections and enables rapid access to DNA raw materials for combinatorial biology experiments.

Acknowledgment. Financial support was provided by NSF CHE-9742005, NSF CHE-9632047, and Trega Biosciences.

References and Notes

(1) Tomasz, A. *New Eng. J. Med*. **1994**, *330*, 1247−1251.
(2) Davies, J. G. *Science* **1994**, *264*, 375−380.
(3) Lai, M. H.; Kirsch, D. R. *Antimicrob. Agents Chemother*. **1996**, *40*, 1645−1648.
(4) Levy, S. B. *Trends Microbiol*. **1994**, *2*, 341−342.
(5) (a) Humma, L. M. *Am. J. Health-Syst. Pharm*. **1996**, *53*, 2291−2298. (b) Heym, B.; Philipp, W.; Cole, S. T. *Curr. Top. Microbiol. Immunol*. **1996**, *215*, 49−69. (c) Blanchard, J. S. *Annu. Rev. Biochem*. **1996**, *65*, 215−239.
(6) Rosamond, J.; Allsop, A. *Science* **2000**, *287*, 1973−1976.
(7) Henkel, T.; Brunne, R. M.; ¨Müller, H.; Reichel, F. *Angew. Chem., Int. Ed*. **1999**, *38*, 643−647.
(8) (a) Cane, D. E. *Chem. Rev*. **1997**, *97*, 2463−24??. (b) Grabley, S.; Thiericke, R. *Adv. Biochem. Eng. Biotechnol*. **1999**, *64*, 101−154. (c) Khosla, C. *Curr. Opin. Biotechnol*. **1996**, *7*, 219−222. (d) Tsoi, C. J.; Khosla, C. *Chem. Biol*. **1995**, *2*, 355−362. (e) Hopwood, D. A.; Khosla, C. *Ciba Found. Symp*. **1992**, *171*, 88−106. (f) Hopwood, D, A.; Chater, K. F.; Bibb, M. J. *Biotechnology* **1995**, *28*, 65−102. (g) Hopwood, D. A. *Curr. Opin. Biotechnol*. **1993**, *4*, 531−537. (h) Hopwood, D. A.; Sherman, D. H. *Annu. Rev. Genet*. **1990**, *24*, 37−66. (i) Hopwood, D. A. *Philos. Trans. R. Soc. London B Biol. Sci*. **1989**, *324*, 549−562. (j) Leadlay, P. F.; Staunton, J.; Aparicio, J. F.; Bevitt, D. J.; Caffrey, P.; Cortes, J.; Marsden, A.; Roberts, G. A. *Biochem. Soc. Trans*. **1993**, *21*, 218−222. (k) Leadlay, P. F. *Curr. Opin. Chem. Biol*. **1997**, *1*, 162−168. (l) Cane, D. E.; Walsh, C. T.; Khosla, C. *Science* **1998**, *282*, 63−68. (m) Roy, R. S.; Gehring, A. M.; Milne, J. C.; Belshaw, P. J.; Walsh, C. T. *Nat. Prod. Rep*. **1999**, *16*, 249−263. (n) Walsh, C. T.; Gehring, A. M.; Weinreb, P. H.; Quadri, L. E.; Flugel, R. S. *Curr. Opin. Chem. Biol*. **1997**, *1*, 309−315. (o) Hutchinson, C. R. *Curr. Opin. Microbiol*. **1998**, *1*, 319−329. (p) Hutchinson, C. R.; Fujii, I. *Annu. Rev. Microbiol*. **1995**, *49*, 201−238.
(9) (a) Kuhstoss, S.; Huber, M.; Turner, J. R.; Paschal, J. W.; Rao, N. *Gene* **1996**, *183*, 231−236. (b) Jacobsen, J. R.; Hutchinson, C. R.; Cane, D. E.; Khosla, C. *Science* **1997**, *277*, 367−369.
(10) (a) Adamidis, T.; Champness, W. C. *J. Bacteriol*. **1992**, *174*, 4622−4628. (b) Adamidis T.; Riggle P.; Champness W. C. *J. Bacteriol*. **1990**, *172*, 2962−2969.
(11) Chater, K. F. *Ciba Found. Symp*. **1992**, *71*, 144−156.
(12) Ishizuka, H.; Horinouchi, s.; Kieser, H. M.; Hopwood, D. A.; Beppu, T. *J. Bacteriol*. **1992**, *174*, 7585−7594.
(13) (a) Christie, B. D.; Nourse, J. G. *Annu. Rep. Comb. Chem. Mol. Diversity* **1997**, *1*, 267−272. (b) Bauer, B. E. *High Throughput Screening*; Devlin, J. P., Ed.; Dekker: New York, NY, 1997; pp 551−581. (c) Franke, R.; Collins, C. *Am. Lab*. **1997**, *29*, 25−26, 28.
(14) The DNP also provided source data for the statistical analysis in ref 7.
(15) Gause, G. F. *Adv. Chemother*. **1965**, *2*, 179−95.
(16) Blanco, G.; Fu, H.; Mendez, C.; Khosla, C.; Salas, J. A. *Chem. Biol*. **1996**, *3*, 193−196.
(17) Montanari, A.; Rosazza, J. P. *J. Antibiot*. **1990**, *43*, 883−9.
(18) Herbert, R. B. *The Biosynthesis of Secondary Metabolites*; Chapman and Hall: New York, 1989.
(19) Meurer, G.; Gerlitz, M.; Wendt-Pienkowski, E.; Vining, L. C.; Rohr, J.; Hutchinson, C. R. *Chem. Biol*. **1997**, *4*, 433−443.
(20) Weinshilboum, R. M.; Otterness, D. M.; Szumlanski, C. L. *Annu. Rev. Pharmacol. Toxicol*. **1999**, *39*, 19−52.
(21) Buckingham, J., Ed. *Dictionary of Natural Products*; Chapman & Hall: New York, 1994.

NP000244X