Biochemical and Biophysical Research Communications **266,** 62–65 (1999)
Article ID bbrc.1999.1775, available online at http://www.idealibrary.com on IDEAL®

# Efficient Discovery of Inhibitory Ligands for Diverse Targets from a Small Combinatorial Chemical Library of Chimeric Molecules

David S. Thorpe,*,†,[1] A. W. Edith Chan,†,[2] Alan Binnie,* L. Charlie Chen,* Anna Robinson,* James Spoonamore,*,† David Rodwell,† Shelly Wade,‡ Sydney Wilson,*,† Martha Ackerman-Berrier,* Helen Yeoman,* Stefan Walle,‡ Qinyuan (Quincey) Wu,‡,[3] and Kenneth F. Wertman*,†

**Department of Discovery Biology, †Department of Chemistry, and ‡Department of Analytical Chemistry, Selectide Corporation, Tucson, Arizona 85737*

Received November 1, 1999

**Living systems are mainly composed and regulated by compounds in four biochemical classes and their polymers—nucleotides, carbohydrates, lipids, and amino acids. Early combinatorial chemistry libraries consisted of peptides. The present report describes the general bioactivity and biophysical properties of a combinatorial chemical library that used glyco, nucleotidyl, and lipid building blocks. The resulting chimeric combinatorial library of 361 compounds had a confirmed cumulative hit rate of 0.16%, which is 8-fold higher than a commonly claimed industrial benchmark of 0.02%. It produced 7 structurally confirmed hits for a third of 12 proprietary drug discovery projects, and these comprised a variety of molecular targets. Diversity analyses demonstrated that despite the small number of compounds, a wider range of diversity space was covered by this library of biochemical chimeras than by a branched tripeptide library of the same size and similar generic formula.**



Nearly any textbook of biochemistry divides the topic along four classes of molecular building blocks and their polymers: nucleotides, amino acids, carbohydrates, and lipids. Indeed, living things constitute an organized and regulated collection of these four biochemical classes in a complex system of molecular interactions. Historically, combinatorial chemistry certainly covered the peptide and amino acids (e.g., 1), but what about the other classes of biomolecules? Sometimes, in order to find useful properties, investigators have combined materials into chimeric structures. Engineers make alloys that have unexpected or beneficial attributes, molecular biologists generate chimeric proteins to explore and harness cellular signaling events (e.g., 2), and medicinal chemists make dual inhibitors to obtain greater efficacy or other pharmaceutical advantage (e.g., 3, 4). So, what happens if chimeras are made using analogs of the basic building blocks of life?

We report here the production and evaluation of a small chimeric library. A collection of 361 chimeric compounds was tested in 12 proprietary drug discovery assays, and it was found to be fairly effective at providing inhibitory ligands of reasonable selectivity, structural specificity and affinity. Interestingly enough, the simple design encompassed a large amount of chemical diversity space compared to a branched tripeptide library, yet the molecules were paradoxically not “drug like.” Such compounds may nevertheless prove to be useful reagents in the biological sciences.

## MATERIALS AND METHODS

*Materials.* Hydroxy TentaGel was from Rapp Polymere (Tubingen, Germany). HOBt and Boc-K(Fmoc)-OH were from Novabiochem (San Diego, CA). Diisopropyl carbodiimide and 1-methylimidazole were from Aldrich. The solvents, DMF, DCM, ACN, and DMSO, were from Aldrich or Burdick & Jackson. The combinatorial building block acids were generally from Aldrich (Milwaukee, WI) or Acros (Pittsburgh, PA).

Abbreviations used: HOBt, hydroxybenzotriazole; DMF, dimethylformamide; DCM, dichloromethane; DMSO, dimethylsulfoxide; ACN, acetonitrile; K, L-lysine; R1, R2, first and second positions for randomized acylations, respectively.

[1] To whom correspondence should be addressed at Selectide Corporation, a subsidiary of Hoechst Marion Roussel, 1580 E. Hanley Boulevard, Tucson, AZ 85737. Fax: (520) 575-8283. E-mail: David.Thorpe@hmrag.com.

[2] Present address: Inpharmatica, 60 Charlotte Street, London W1P2AX, UK.

[3] Present address: Bayer AG, Central Research/Chemistry for Life Sciences, Building Q-18, D-51368 Leverkusen, Germany.

0006-291X/99 $30.00

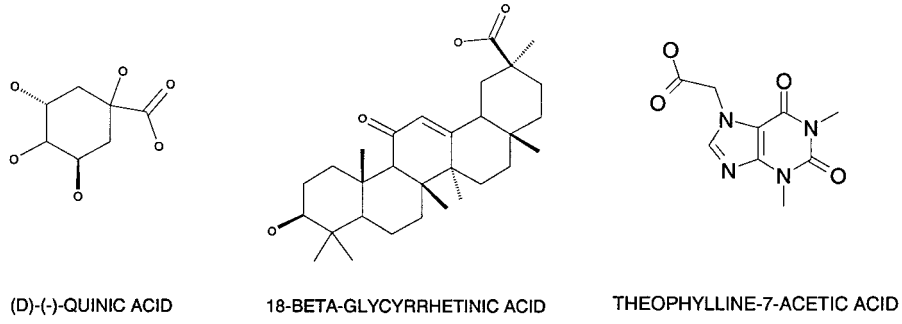


**FIG. 1.** Examples of the building blocks used in the chimeric diamide library. Hydrogens are not shown.

*Methods.* The design of the combinatorial library was done in a few steps. Building blocks were selected by searching the Available Chemical Directory (ACD, MDL Information Systems, Inc., San Leandro, CA) with the following criteria: substructure search, or 50% similarity to nucleotidyl bases (purine, pyrimidine or xanthine), carbohydrates, or hormonal lipids (e.g., steroidal, sterol, or other lipid analogs). All had to have a free carboxylic acid functional group (obtained as either the free acid form, or as a salt). They also had to be anhydrous (water content $< 0.01\%$), or simply made so by addition of molecular sieves, and cost $<$\$100/g. By way of illustration, some examples of the 19 building blocks selected are shown in Fig. 1. The free acids of some building blocks were prepared from their salts by TFA acid precipitation from an aqueous solution. Furthermore, some building blocks had precedence in the medicinal chemistry repertoire. For example, the carbohydrate analog quinic acid is a constituent of various bioactive compounds—e.g., agents derived from the natural product, propolis, and enzymatic inhibitors (e.g., 5, 6). The steroidal building block, glycyrrhetinic acid, has been widely described as having interesting activities alone or as part of a bioactive molecule (e.g., 7). Theophylline acetic acid is an analog of the methylxanthine class of drugs known since the dawn of medicinal chemistry. Finally, to be included in the library, each building block had to yield satisfactory products upon synthesis of model compounds. The products were evaluated by using TLC (8), HPLC, MS and MS/MS. After all of these steps, 8 lipids, 6 nucleo analogs and 5 carbohydrate analogs were used in the present chimeric library. For production of the library, L-lysine was selected as the combinatorial "scaffold" onto which the building blocks would be coupled through amide chemistry. The choice of this scaffold was based upon the pragmatic considerations of availability, steric accessibility and being the first chimeric library to be produced. The formula for this library is R1-K(R2)-OH, where R1 is the first randomized acid coupled to the alpha amine, and R2 is the second randomized acid coupled to the epsilon amine. Note that all compounds are at least a hybrid of two classes since all of them are coupled to an amino acid, lysine.

The chimeric library was made using a split, 3-way pool, split method to couple the selected 19 acids to each of the orthogonally protected amines of the lysine scaffold. The total complexity of the chimeric library was 361 ($19 \times 19$) intended compounds. In the appropriate anhydrous solvent, 5 equivalents of each acid along with DIC and HOBt were used to acylate the reactive amine. The extent of coupling was monitored by Ninhydrin testing of resin samples, and all acylations went to 95% completion, or better. Conventional Boc/Fmoc chemistry steps were used during the synthesis of the library in polypropylene syringes. The resin was in 57 different syringes after the second randomization ($3 \times 19$). To facilitate structural determination, the resin was kept in separate syringes after the second acylation so that the acid used was known for the R2 position. The combinatorial products were cleaved off of the resin with aqueous base and collected, and the resin was washed with 50% ACN in water and collected. These two eluates were kept as separate solutions thereby providing aqueous and organic extracts. There were 57 aqueous extracts and 57 organic extracts in the final library. Depending upon the particular source syringe, there were 6 or 7 expected chimeric compounds. Following cleavage, they partitioned into the aqueous or organic fractions according to their solubilities.

These soluble compounds were tested in 12 proprietary drug discovery bioassays of both molecular and cellular types. All projects were looking for inhibitors. The resin loading gave an upper limit estimate of the concentration of combinatorial compounds being tested in the assays. Appropriate solvent controls were used, and the extracts were generally diluted 100-fold into biological reactions. Screening hits had statistically significant inhibitory effects compared to control incubations (e.g., $>4$ fold the standard deviation of controls). The active compound was determined by either synthetic deconvolution or by purification followed by appropriate microanalytical methods (e.g., FT-MS/MS). Data from dose-response curves were fitted using nonlinear regression software (9).

*Computational methods.* The Legion (10) combinatorial builder module in SYBYL was used to create two libraries, first the chimeric diamide library described above, and second a theoretical branched tripeptide library with the same generic formula R1-K(R2)-OH, in which 19 amino acids (from the 20 common amino acids, except lysine) at R1 and R2 were randomly combined to yield 361 virtual tripeptides. Thus, each library consisted of 361 compounds. The Tanimoto coefficient (11, 12) was used to assess the dissimilarity between any two compounds from the two libraries. The work was done with a SGI Octane machine (SGI, Mountain View, CA). Another method, the Auto Correlation Technique Package (ACT), was used to generate graphical output. The basic theory of ACT has been published (13–16). Simply put, this method uses an autocorrelation technique to transform a 2D- or 3D-molecular structure, including its local properties, into a vector, that is, a point in an N-dimensional space. Then statistical techniques, including principal component analyses, are used to reduce the whole collection to a smaller number of molecules. The parameters used are atom connectivity, Van der Waals volume, pi functionality, electronegativity of an atom, heteroatomicity, hydrogen bond donor or acceptor and atomic contribution to LogP.

## RESULTS AND DISCUSSION

Extracts from the chimeric combinatorial library were tested in 12 drug discovery projects. Eight assays did not find any hits from the chimeric library. The other 4 bioassays found at least one hit. In all, 7 compounds with desirable bioactivities were identified. A screen for a lipid enzyme discovered one hit, three hits were found for two kinases, and three hits were found for a protease. None of these compounds were crossre-

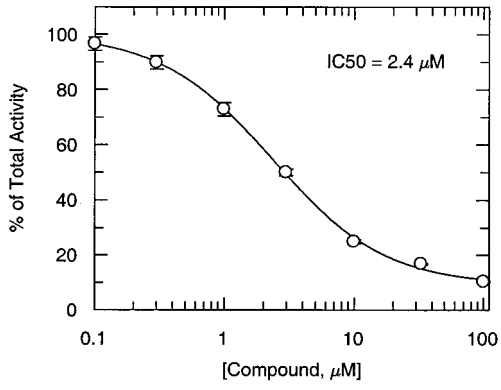


**FIG. 2.** Effect of a dinucleotidyl chimeric molecule on the phophorylating activity of a protein kinase. Open circles represent the mean values expressed as a percentage of control activity. The standard deviation is indicated by bars. The data are representative of three determinations.

active, indicating biological selectivity. For example, neither of the two hits found for the first kinase affected the second kinase, and the one for the second kinase did not affect the first kinase. Also, the hits came from extracts that were active in only one bioassay, again indicating selectivity dependent upon molecular recognition. All of these hits had reasonable affinities with potencies $<15$ $\mu$M. For each hit, analogs were made and tested to demonstrate structural specificity as well, thereby fulfilling the requirements for a true ligand:target type of molecular interaction. Finally, for the projects described, none of the building blocks were significantly active by themselves indicating a need to be coupled in a chimeric molecule.

The biological properties of the chimeric compounds will be illustrated by describing the hits found for two proprietary screening projects. Figure 2 shows a dose-response curve for a dinucleotidyl inhibitor of a protein kinase. The curve shape was normal with a slope of 1. The $IC_{50}$ was 2.4 $\mu$M. Ligand specificity was confirmed by an inactive analog that had two atoms removed compared to the hit. Furthermore, this hit and a nucleosterol hit did not affect the phosphorylation activity of other protein kinases, such as cAMP dependent protein kinase, or two other proprietary protein kinases, demonstrating selectivity in their inhibitory effects. Figure 3 shows a dose–response curve for a nucleosterol inhibitor discovered for a lipid enzymatic assay. The best fitting curve had a slope of 1.5, and the $IC_{50}$ was 13 $\mu$M. Structural specificity of the ligand was established using isomeric and spacing analogs (L-diaminopropionic acid substituted for the L-Lys scaffold). Furthermore, this compound was not active in any of the other 12 screens. Inactive building block deletions, inversions between the alpha and epsilon amines, and analogs using isomeric building blocks all emphasized the structural requirements of the hits for these two projects. These two examples help demonstrate that true ligands were found that fullfill basic requirements for molecular recognition: affinity, structural specificity and biological selectivity.

The chimeric library was fairly efficient at providing inhibitory ligands. Seven confirmed hits from 361 compounds translates into 52 combinatorial compounds made per hit (nearly 2% were ligands found for 4 screening projects). This indicates a highly concentrated amount of bioactivity in the chimeric collection. Despite loading the building block set with a few analogs of known therapeutic agents or substructures, only one of the 7 hits had such a building block. The target it hit, moreover, was counterintuitive. A benchmark commonly cited in the drug discovery industry is a hit rate of 0.02% (calculated as 100 $\times$ the total number of hits divided by the product of the total number of compounds tested $\times$ the number of screens). The chimeric library described here, as small as it was, had a confirmed hit rate eight times better than the stated benchmark.

To better understand the efficiency of the chimeric library, computational chemistry tools were applied to evaluate its properties. Common sense suggested that the chimeric library would be very different from peptide libraries, and diversity analysis quantified this expected difference. When a branched tripeptide library of the same R1-K(R2)-OH formula as the chimeric library was compared, there was 0% redundancy between these two libraries based on a Tanimoto Index of 0.85. Figure 4 shows the first two principal components axes (Factors F1 and F2). F1 represents mainly the size of the molecules. F2 is mainly related to the polarity and unsaturation of the molecules. On the lower rim of the cloud are lipophilic molecules, such as saturated hydrocarbons; by moving towards the positive value of F2, one finds unsaturated hydrocarbons, and then, more and more heteroatoms within unsaturated molecules. From our experience, factors F3, F4,

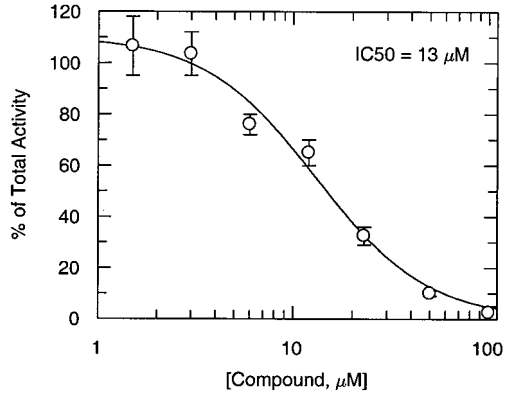


**FIG. 3.** Effect of a nucleosterol on an enzyme that handles lipids. Open circles represent the mean values expressed as a percentage of control activity. The standard deviation is indicated by bars. The data are representative of three determinations.

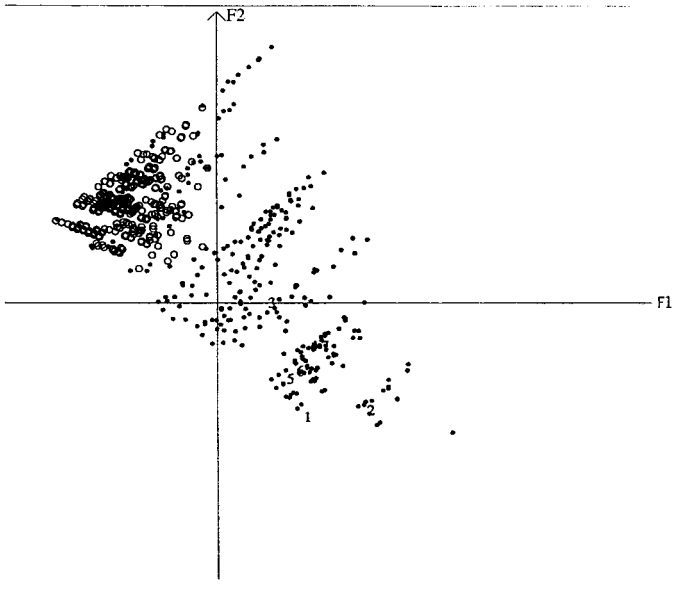


**FIG. 4.** Diversity comparison of the chimeric diamide and virtual tripeptide libraries. The libraries are projected onto the F1–F2 plane. Open circles denote the virtual branched tripeptide library; filled circles represent molecules of the chimeric diamide library. The positions of seven chimeric hits are plotted as numbers 1 through 7. The hits are all in the lower right quadrant.

F5, etc., are more suitable for the observation of structural family cluster. The comet shape of the cloud of the F1–F2 plane is typical of molecular sets containing small and large molecules. The smaller ones are at the left tip, and as the size increases along F1, diversity along F2 also increases. The molecular size range of the tripeptide library is smaller than the chimeric library, the molecular weights of the former being in the range of 400 to 500, while the latter is from 300 to 1000, respectively. The seven hits were in a diversity space that is not covered by the tripeptide library.

The chimeric library also was not drug like. Its physical properties fall out of the Lipinski's "Rule of Five" (17). The molecular weight was from 300 to 1000 (Mode of 700 Da); hydrogen-bond donor was from 3 to 10 (Mode of 6); hydrogen-bond acceptor was from 4 to 14 (Mode of 9); cLogP was from −1 to 13 (Mode of 6). Nevertheless, this library provided 7 hits.

Four of 12 screens were successful in finding one or more hits from the library giving a project success rate of 33%. Although the screening projects were diverse, they were not randomly distributed in character. The targets were classified in terms of the biochemical class of their natural ligand(s). Nine (75%) of the proprietary screens had ligands that were nucleotidyl or peptidic in type. Six of the confirmed hits were evenly distributed among the targets handling nucleotidyl and peptidic ligands. Only one target within an engineered cell line handled lipids, which found a single hit from the chimeric library. One project had a target that handled carbohydrates, but it did not yield a confirmed hit. Thus, from among the four classical biochemical classes, targets that handle carbohydrates or lipids were under-represented in the screening set. To date, it may be concluded that the chimeric library design succeeded in providing ligands for targets that normally handle nucleotidyl, peptidic and lipid molecules.

From acetylcholine to Zovirax, medicinal chemists have drawn inspiration from biological factors. The studies described here show that the chimeric design was quite relevant and efficient in the discovery of inhibitory ligands, yet it was not designed for any particular target apart from using analogs of building blocks commonly found in the molecular recognition of biological systems. This report is the first to describe an early attempt to splice together the major biochemical classes to generate chimeric molecules and the biological consequences of such unions.

## ACKNOWLEDGMENT

The authors thank Daniel Schirlin for constructive criticism and editing of the manuscript.

## REFERENCES

1. Lam, K. S., Salmon, S. E., Hersh, E. M., Hruby, V. J., Kazmierski, W. M., and Knapp, R. J. (1991) *Nature* **358,** 82–84.
2. Riedel, H., Dull, T. J., Schlessinger, J., and Ullrich, A. (1986) *Nature* **324,** 68–70.
3. Fink, D. M., Palermo, M. G., Bores, G. M., Huger, F. P., Kurys, B. E., Merriman, M. C., and Olsen, G. E. (1996) *Bioorg. Med. Chem. Lett.* **6,** 625–630.
4. Mehdi, S., Angelastro, M. R., Burkhart, J. P., Koehl, J. R., Peet, N. P., and Bey, P. (1990) *Biochem. Biophys. Res. Commun.* **166,** 595–600.
5. Basnet, P., Matsushige, K., Hase, K., Kadota, S., and Namba, T. (1996) *Biol. Pharm. Bull.* **19**(4), 655–657.
6. Schaub, C., Muller, B., and Schmidt, R. R. (1998) *Glycoconj. J.* **15,** 345–354.
7. Harada, S., Maekawa, T., Haneda, E., Morikawa, Y., Nagata, N., and Ohtsuki, K. (1998) *Biol. Pharm. Bull.* **21**(12), 1282–1285.
8. Thorpe, D. S., Niu, S., and Morkin, E. (1996) *Biochem. Biophys. Res. Commun.* **218,** 670–673.
9. Leatherbarrow, R. J. (1992) GraFit Version 3.0, Erithacus Software Ltd., Staines, U.K.
10. Legion, SYBYL, UNITY, and Selector are from Tripos, Inc., St. Louis, MO.
11. Willett, P., and Winterman, V. A. (1986) *Quant. Struct.-Act. Relat.* **5,** 18–25.
12. Willett, P. (1987) Letchworth Research Studies Press.
13. Moreau, G., and Broto, P. (1980) *Nouv. J. Chim.* **4,** 359–360.
14. Moreau, G., and Broto, P. (1980) *Nouv. J. Chim.* **4,** 757–764.
15. Moreau, G., and Turpin, C. (1996) *Analusis Magazine* **24,** 17–21.
16. Broto, P., Moreau, G., and Vandycke, C. (1984) *Eur. J. Med. Chem. Chim. Ther.* **19,** 66–70.
17. Lipinski, C. A., Lombardo, F., Dominy, B. W., and Feeney, P. J. (1997) *Adv. Drug Delivery Rev.* **23,** 3–25.