Journal of Computer-Aided Molecular Design **19:** 47–63, 2005.
DOI 10.1007/s10822-005-0097-6

A very large diversity space of synthetically accessible compounds for use with drug design programs

Sergey Nikitin[1,*], Natalia Zaitseva[1], Olga Demina[1], Vera Solovieva[1], Evgeny Mazin[1], Sergey Mikhalev[1], Maxim Smolov[1], Anatoly Rubinov[1], Peter Vlasov[1], Dmitry Lepikhin[1], Denis Khachko[1], Valery Fokin[2], Cary Queen[1] & Viktor Zosimov[1]
[1]*Algodign LLC, Bolshaya Sadovaya 8, Moscow 103001, Russia;* [2]*Department of Chemistry, The Scripps Research Institute, 10550 N. Torrey Pines Rd., La Jolla, CA 92037, USA*

Received 16 September 2004; accepted in revised form 3 January 2005

Key words: combinatorial library, *de novo* drug design

Summary

We have constructed a very large virtual diversity space containing more than 10^{13} chemical compounds. The diversity space is built from about 400 combinatorial libraries, which have been expanded by choosing sizeable collections of suitable R-groups that can be attached to each link point of their scaffolds. These R-group collections have been created by selecting reagents that have drug-like properties from catalogs of available chemicals. As members of known combinatorial libraries, the compounds in the diversity space are in general synthetically accessible and useful as potential drug leads. Hence, the diversity space can be used as a vast source of compounds by a *de novo* drug design program. For example, we have used such a program to generate inhibitors of HIV integrase enzyme that exhibited activity in the micromolar range.

Introduction

Structure-based drug design programs are potentially an important complement to experimental approaches such as high-throughput screening for discovery of new drug leads (recently reviewed in [1]). Such programs typically have three major components: (1) one or more scoring functions to evaluate quality of binding between a compound and the target protein; (2) an algorithm to search for a conformation and orientation of a ligand or a fragment that yields an optimal score; and (3) a conceptual *diversity space* of chemical compounds from which candidate ligands are selected or generated. When the diversity space is not simply a predetermined list of compounds, a fourth component is needed – an algorithm for constructing members of the space – which may be integrated with other components of the program. We have developed a de novo drug design program, to be described in detail elsewhere, that uses a novel scoring function, search algorithm and diversity space. Here we describe construction of the diversity space, which may have applications to other structure-based drug design programs.

For the docking or virtual screening class of drug design programs, which evaluate compounds one-by-one for binding to the target, the diversity space generally consists of a list of particular compounds, for example a subset of the Available Chemicals Directory [2] or one or more previously described combinatorial libraries. Of course, different diversity spaces may be chosen for various applications of a given program. One advantage of such a diversity space is that it can be selected to contain only compounds that are sufficiently "drug-like" or have other desirable properties. Another important advantage is that it can be chosen so that each compound has already been or

*To whom correspondence should be addressed. Fax: +7-95-209-1415; E-mail: Sergei.Nikitin@algodign.com

can readily be synthesized. However, because docking programs usually require 100 s or more to evaluate each compound, diversity spaces of this type generally contain fewer than 100,000 compounds so they can be screened in a reasonable amount of time. This is no greater than the number of compounds that can be handled by high-throughput experimental methods, which limits the advantages of the programs, although of course it may still be considered preferable to screen a diversity space by computer rather than experimentally.

On the other hand, for drug design programs of the de novo class, which build candidate ligands from defined components according to certain rules, the size of the diversity space may be essentially infinite. For example, LEGEND [3] and MCDNLG [4] build the compounds from individual atoms. Other programs such as LigBuilder [5] and Pro_Ligand [6] construct compounds from a defined set of functional groups. However, a significant issue with such programs is that it may not be possible to synthesize many or even most of the candidate ligands in any straightforward way. Indeed, the programs build up the compounds with little or no attention to whether known chemical reactions can create the desired bonds, and moreover without altering parts of the compounds already built. In principle, it may be possible to develop a *de novo* drug design program having an extremely large diversity space of synthetically accessible compounds by building in expert knowledge about acceptable chemical reactions, retrosynthetic analysis, etc. However, it is likely to be very challenging to ensure that most of the compounds can really be synthesized in the time that a pharmaceutical industry chemist is likely to devote to them (one to several weeks).

Other programs such as PRO_SELECT [7] and CombiDOCK [8], which have aspects of both the screening and the *de novo* approaches, avoid the synthesis issue by constructing the diversity space for a particular target from one template (scaffold) and a set of functional groups known to be linkable to it, i.e., from a single combinatorial library. Unfortunately, this approach severely limits the diversity space and may thus be most appropriate when the chosen library is already known to have members that bind to the protein. More generally, extensive work has been devoted to developing programs that construct and manipulate combinatorial libraries, for example [9–15]. A major goal of many such programs is to select relatively small subsets of large combinatorial libraries that retain high diversity while maximizing one or more desirable properties such as drug-likeness or similarity to a known pharmacophore. Sophisticated methods such as D-optimal algorithms [10], tree-sorting and hierarchical selection [11], genetic algorithms [12], and Monte-Carlo [13] techniques have been employed for this purpose. However, these approaches are most relevant when it is necessary to drastically reduce the size of a potential combinatorial library, e.g. from billions to thousands of compounds, because it is going to be actually synthesized and used in high-throughput screening.

In contrast, the very large diversity space described here is intended for use with drug design programs capable of handling such a large diversity space. Hence, while we also eliminate compounds that are likely to be useless because of insufficient drug-likeness or other undesirable properties, it is not necessary to prune the libraries so severely. Indeed, our purpose is different than the prior work cited – it is to generate a very large and highly diverse space of readily synthesized compounds, from which a *de novo* drug design program will have the greatest opportunity to generate ligands having high affinity for any target protein. In our approach, the diversity space is composed of hundreds of different combinatorial libraries, each of which contains from a small number up to billions of compounds.

Concepts and definitions

Before describing construction of the diversity space, it will be helpful to provide several definitions, which are illustrated schematically in Figure 1. A *combinatorial library* is a set of compounds consisting of a common element of chemical structure called a *scaffold* to which collections of chemical fragments called *R-groups* are attached at designated atoms (*link points*). Importantly, the scaffold and R-groups are not the same as the actual compounds that would be used to chemically synthesize members of the library, because those compounds are modified in the chemical synthesis reactions. We call R-reagents

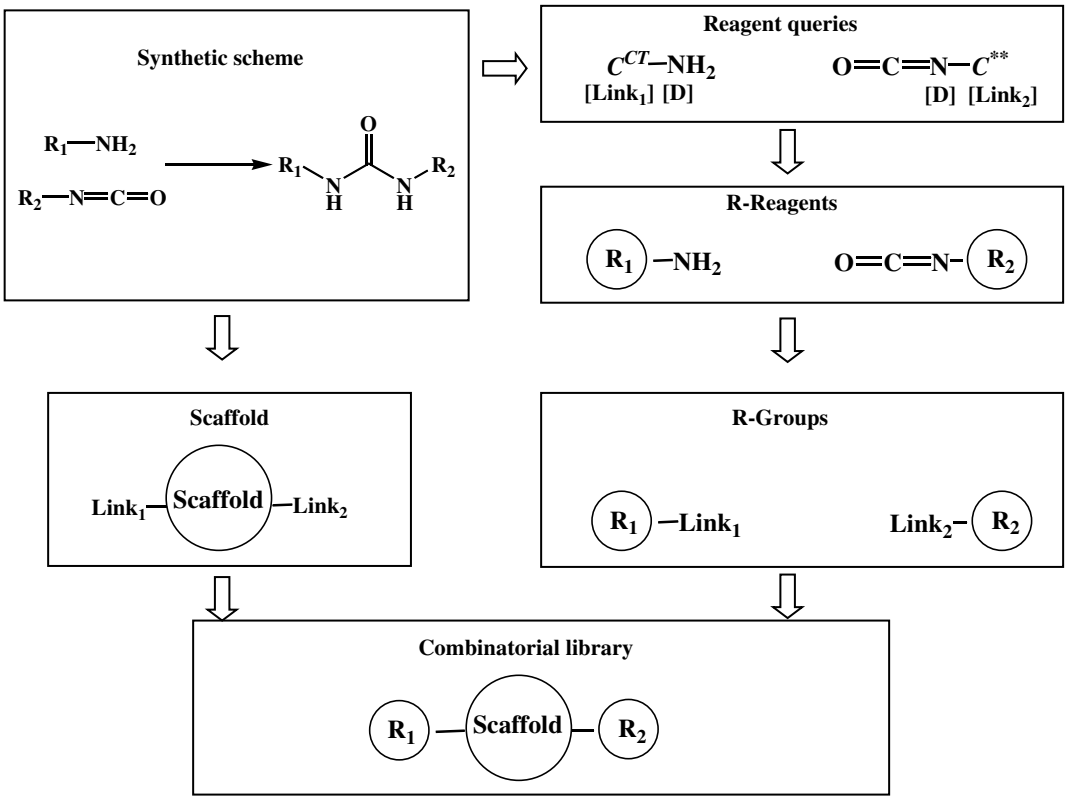


Figure 1. Schematic illustration of construction of a combinatorial library.

the compounds which become the R-groups. The R-groups have often been referred to by others as *clipped* reagents. For many libraries, not only the R-groups, but part or even the entire scaffold originate from the R-reagents during synthesis. For example, in the combinatorial library shown in Figure 1, the urea scaffold is constructed completely from the two R-reagents [16]. A combinatorial library or diversity space which exists in the abstract rather than as a collection of real, existing compounds is often called a *virtual* combinatorial library or a *virtual* diversity space, but here we will usually omit the *virtual* as being implied.

Associated with each combinatorial library is a synthetic scheme, a series of chemical reactions which can be used to synthesize each member of the library. To make the same reactions applicable to the synthesis of every member of the library, the collection of R-reagents that are the source of R-groups attached to a given link point must all have a common reactive functionality, for example an amine or an alcohol. However, as that functional group is typically destroyed during the synthesis, the corresponding R-groups are not necessarily related. The scientific publications that present combinatorial libraries generally describe the synthetic scheme as well as the limitations on the R-reagents, but typically only provide a limited number of examples of acceptable R-reagents.

Libraries

Two major schemes have previously been used to enumerate large combinatorial libraries in a form that can manipulated by a computer, without the need to store information about each compound separately in computer memory: ''fragment marking'' and ''reaction transform'' [9]. In the ''fragment marking'' method, the scaffold and lists of the R-groups at each link point are stored, and library members are generated as needed by combining the scaffold and respective R-groups. In the reaction transform approach, the lists of input

reagents are specified together with rules for combining them that correspond to the chemical reactions used to synthesize a library. The latter approach may provide additional information to the chemist, but is more difficult to set up and can be very computationally intensive. For example, one such program [9] enumerated 640,000 compounds in 18 h. Clearly, such a program would not be effective to generate combinatorial libraries of 10^9 or more compounds. Hence, to construct the very large diversity space described here, we use the fragment marking method. A previously reported limitation of this approach, that it does not adequately handle the situation when part of the scaffold itself is formed from an R-reagent such as by the Diels–Alder reaction [9, 10], is addressed by a reinterpretation of what constitutes the R-group.

We began construction of the diversity space by scanning the scientific literature for descriptions of combinatorial libraries, taking as a starting point R.E. Dolle's excellent reviews [17–20] of combinatorial library synthesis. Other reviews [21, 22] were also useful. Libraries were selected with the aim of providing a structurally diverse set of pharmaceutically relevant scaffolds. In general, scaffolds of relatively low molecular weight but having several points for linking of R-groups were preferred, in order to maximize the diversity provided by the R-groups. Predicted ease and reliability of the synthetic scheme was also taken into account. The libraries selected to date for the diversity space come from about 200 publications (Appendix A), but in many cases one publication yielded more than one library. That was either because the article actually described more than one scaffold, or because the same scaffold gave rise to several libraries when different classes of R-group (e.g., alkyl halides and carboxylic acids) were attached to a given link point using different transformations.

For each library, a file was created that contains the structure of the scaffold with designations of the atoms to which R-groups could be attached (the link points) The file also contains information about the type of R-reagent that can be attached at each linkage point (amine, alcohol, etc.). We also recorded the synthetic scheme used to make the library in the corresponding publication and, following the fragment marking approach, a list of the particular R-groups described there. To date, about 400 separate libraries have been used to construct the diversity space, and additional libraries are currently being processed.

Expansion of combinatorial libraries

The libraries described above, each made from a scaffold and the particular R-groups described in the corresponding publication, already constitute together a diversity space. Indeed, initially our de novo drug design program used only this diversity space, which was calculated as containing about 4.5 million individual compounds. However, to increase the probability of obtaining ligands with high affinity against a given protein, the libraries were vastly expanded by creating much larger collections of R-reagents and therefore R-groups.

3D reagent database

For this purpose, a large database of available chemical reagents was first generated by extracting the organic compounds with low molecular weight from the Sigma-Aldrich, Acros, Lancaster and TCI catalogs [23–26]. If a compound was an organic salt, it was converted into the corresponding acid or base and included in the database as such. In order to avoid duplication of compounds, the database was first divided into groups of compounds with the same empirical formula. Within each group, a search for duplicated structures was made by attempting to embed each structure into all structures of the group using a published algorithm [27], and the duplicates were eliminated. Compounds with atoms rarely found in drugs (i.e., other then C, H, N, O, B, F, Cl, Br, I, P, and S) were shunted into a separate database used in rare cases. For example metallo-organic fragments are generally undesirable in drugs, whereas metallo-organic compounds are sometimes used as synthetic reagents. So in order to utilize these reagents, we can scan this separate database when desired.

Our drug design program, like most such programs, requires knowledge of the 3D structures of the members of the diversity space, which can be provided as the 3D structures of the scaffolds and R-reagents. While the R-reagent structures could be obtained for each individual library, it was more convenient to determine the 3D structures of

all compounds in the database at once, so that any later corrections of compound structures could be readily propagated to each library. The 3D structures were prepared using HyperChem release 7 [28] with geometry optimization according to the Amber99 force field and atomic charges calculated according to the PM3 method. The structure of each compound was then included in the database to create a 3D reagent database containing approximately 100,000 structures (one conformation per one reagent). Thus the retained information consists of the reagent name, molecular formula and weight, supplier, catalog number, and atomic coordinates, charges, AMBER types, and connectivity.

Selection of R-reagents by reagent queries

In programs to construct combinatorial libraries, the selection of reagents is a critical step. Generally this process is implemented by means of a substructure search, for example as realized within the ADEPT [9] or MERLIN [10] programs. At a minimum, substructure searches are required to identify R-reagents that can participate in the reaction scheme for a combinatorial library, but they have also been used to identify R-reagents having characteristics desired in a particular targeted library. Our method of reagent selection has three distinguishing features. First, as mentioned above all reagents have 3D structures, although the substructure search is carried out on the basis of 2D structures. Second, the search is performed taking into account AMBER atom types. Leach et al. [9] and others use bond order for this purpose, but in some cases this does not provide a fine enough distinction. For example, nitrogen in amines and amides has the same bond order but different AMBER type. Since, amines and amides can participate in different chemical reactions and serve as R-reagents for different libraries, it is very important to distinguish them. And third, as described below, there is a capability for more sophisticated reagent searches such as for dependent R-groups and for conditionally prohibited fragments. This facilitates the exclusion of R-reagents containing toxic or unstable structures.

To create the collections of R-groups, we performed the following steps as illustrated in Figure 1: (1) for each library, using knowledge of the requirements for the R-reagents at each link point, special *reagent queries* were prepared as described below; (2) using proprietary software, all reagents in the 3D reagent database fulfilling the reagent queries were selected; (3) functional groups of these R-reagents were eliminated in accordance with the chemical reactions used to attach the R-reagents to the scaffold at link points, and were replaced by "dummy atoms" needed for saving the location and direction of the bonds from R-groups to the scaffold; (4) the structures of the R-groups thus generated were stored in lists for each library.

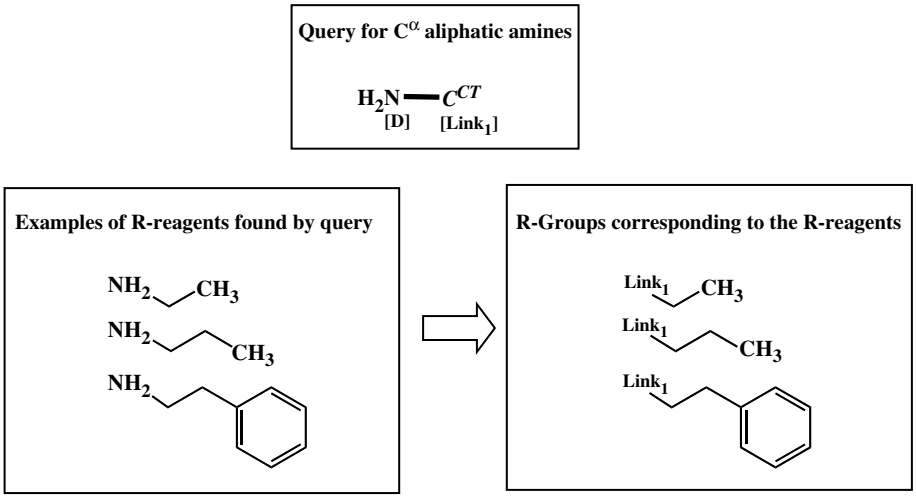


Figure 2. Example of reagent query for aliphatic amines.

Information on the R-reagents themselves was also retained.

A reagent query includes the functional group needed for an R-reagent to participate in a combinatorial library synthetic scheme, and special labels to indicate how the R-reagent is converted to an R-group. For example, the query for C-alpha aliphatic amines is represented in Figure 2, where [D] is a label for the deleted group (NH_2) and [$Link_1$] is a label showing the atom which will be bonded to the scaffold at link point 1. During transformation of R-reagent to R-group, the [$Link_1$] label is moved to a dummy atom, which represents the scaffold atom to which the R-group will be attached. An in-house program based on the referenced algorithm [27] attempts to embed the reagent query successively into each structure in the 3D reagent database. When a correct embedding in a compound is found, that compound is selected as a candidate R-reagent. The program attaches [D] and [Link] labels to the R-reagent using the query information and then generates an R-group from the reagent (step 3 above).

Our methodology allows but does not require that the type of any atom be specified, for example, aliphatic carbon, aromatic carbon, etc. The atom type is indicated in the reagent queries (e.g., C^{CT} for aliphatic carbon and C^{CA} for aromatic carbon), whereas ** indicates that any type is permissible (e.g., C** for any carbon). Thus, the reagent queries for all carboxylic acids and all aromatic carboxylic acids are respectively, C^{**}–(COOH) and C^{CA}–(COOH). The symbol H** is used to represent any atom from the acceptable list. For example, Figure 3 shows a query which covers benzenes, pyridines, triazines, and other six-member aromatic rings. The reaction scheme for a combinatorial library may require that the R-reagents have more than one functional group,

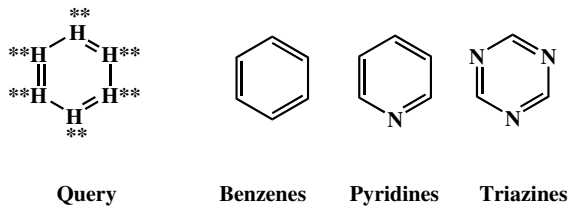


Figure 3. Generalized reagent query for six-membered aromatic rings.

for example an amino group to attach to a link point of the scaffold and a carboxylic acid group to attach to a solid support. Our program supports such queries, for example the query for R-reagents having an amino and carboxylic acid group (amino acids) is shown in line 4 of Table 1. More generally, Table 1 shows the list of reagent queries used to create the collections of R-reagents used for most of the combinatorial libraries.

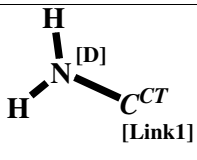
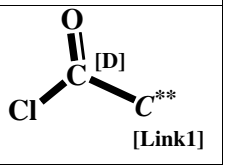
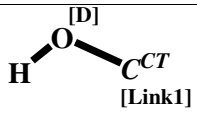
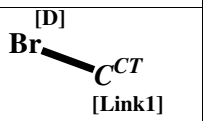
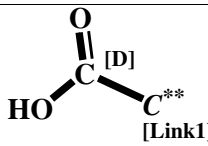
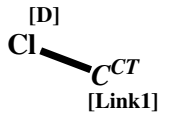
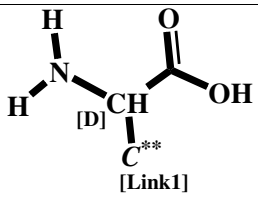
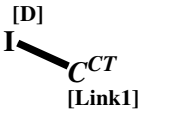
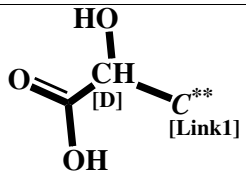
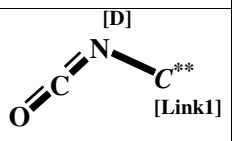
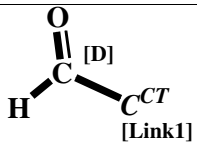
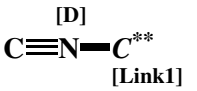
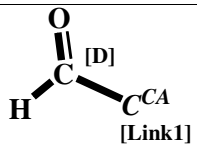
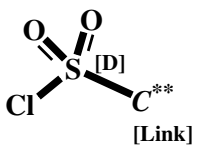
Finally, our methodology can handle synthetic schemes in which a single R-reagent becomes two or more ''dependent'' R-groups attached to the scaffold. An example is the library of 2,3-disubstituted indoles [29] in which the alkyne reagent R_1–C ≡ C–R_2 becomes separate groups R_1 and R_2 attached to the indole scaffold (Figure 4). For each such R-reagent found by the reagent query, R_1 and R_2 are included in a collection of dependent R-groups that can be attached to link points 1 and 2 of the scaffold, in such a way that when the *de novo* drug design program generates members of the library, R_1 and R_2 must be used together.

Filtering of R-groups

Ideally, each member of the diversity space should not only be synthetically accessible, but should be adequately ''drug-like''. While the same goal could be achieved by discarding non-drug-like compounds *after* they have been generated by the de novo drug design program, that would be inefficient because computer time and memory would have been wasted in generating those compounds. Moreover, our de novo drug design program and other such programs generate a limited number of candidate compounds with top scores, and if these compounds are non-drug-like, other lower scoring but more drug-like compounds might be missed. Hence, to avoid these problems we instead filtered the lists of R-groups to exclude those likely to give rise to non-drug-like compounds.

Specifically, our aim was to use only R-groups that (i) contain elements commonly found in drugs, (ii) do not contain functional groups with known toxicity or stability problems, and (iii) are likely to give rise to compounds that fulfill the rule-of-five of Lipinski et al. [30] for drug-likeness. We therefore retained in the collections only R-groups satisfying the following criteria:

Table 1. The most commonly used reagent queries used to expand libraries.

	Reagent	Query		Reagent	Query
1.	Primary aliphatic amines		8.	Acyl-chlorides	
2.	Aliphatic alcohols		9.	Aliphatic bromides	
3.	Carboxylic acids		10.	Aliphatic chlorides	
4.	α-Amino acids (except for Gly and Pro)		11.	Aliphatic iodides	
5.	α-Hydroxy acids		12.	Iso-cyanates	
6.	Aliphatic aldehydes		13.	Iso-cyanides	
7.	Aromatic aldehydes		14.	Sulfonyl-chlorides	

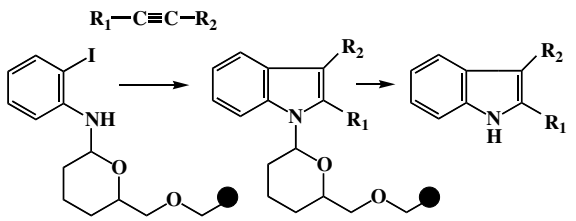


Figure 4. Example of an R-reagent that generates two dependent R-groups in construction of a library of 2,3-disubstituted indoles [29].

Contains only elements: H, C, N, O, S, P, F, Cl.
Molecular weight ≤ 250
Number of rotatable bonds ≤ 5
Total number of N's and O's ≤ 5
Total number of OH, NH and SH groups < 5

Notably, a criterion for clogP was not included, because a high clogP for one R-group could be readily compensated by a low clogP of other R-groups and/or the scaffold.

In addition, R-groups containing known toxic or unstable fragments were eliminated. The set of prohibited fragments currently includes about 50 structures, but can be readily expanded or reduced as desired. Some of them are unconditionally prohibited (including acylchlorides, isocyanates, dichlorosubstituted vinyl derivatives), while others may be acceptable if certain additional conditions are satisfied. For example p-dihydroxybenzene fragments are undesirable because of transformation to quinones during metabolism. However, additional carboxylate functionalities result in alteration of chemical properties and metabolism, which may make this fragment satisfactory as part of a larger R-group. We used an in-house program to determine if each prohibited fragment could be embedded in each R-group. Those R-groups containing unconditionally prohibited fragments were eliminated. For those R-groups containing conditionally prohibited fragments, the program determined whether other functionalities making the fragment acceptable were present, and retained or discarded the R-group accordingly.

In general, the overall filtering procedure reduced the number of R-groups by a factor of about three relative to the original lists of R-groups derived from the 3D database. Hence, this procedure substantially reduced the number of undesirable compounds in the diversity space, especially considering the multiplicative nature of the number of R-groups attached to different link points on the same scaffold. However, the diversity space certainly still contains some undesirable compounds because when several R-groups, each independently satisfying the listed rules, are combined with a scaffold, the resulting compound may exceed the allowed molecular weight, number of hydrogen bond donors or acceptors, etc. Such compounds can be discarded after they have been generated by the drug design program, either in an automated fashion or by the supervising chemist. Of the candidate ligands generated by the program, the chemist can also choose for synthesis only those compounds having preferred properties such as ability to serve as a good drug lead.

Characterization of the diversity space

To evaluate the size and extent of the diversity space, its constituent combinatorial libraries were first numbered 1, 2, 3 ... according to ascending size (i.e., number of compounds in the library). Figure 5 shows a plot of the numbers of the combinatorial libraries against the size of each library and the total size (cumulative number of compounds) in all libraries having numbers less than or equal to it. As can be seen from the graph, the first few libraries contain only a few compounds each – these special libraries were generally created for testing purposes. The next 150 or so libraries each contain less than 10^3 compounds; such libraries typically have only one link point for R-groups. The next approximately 150 libraries each contain between 10^3 and 10^6 compounds, and the final, largest 75 or so libraries each contain more than 10^6 compounds. Hence, the libraries of the diversity space have a smooth size distribution varying from minimal to extremely large.

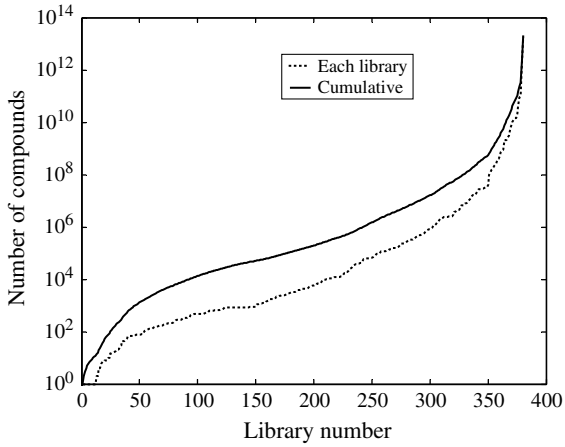


Figure 5. Graph of the combinatorial libraries in ascending size against the number of compounds in each library (...) and the cumulative number of compounds (—).

The total size of the whole diversity space is $>10^{13}$ compounds, as may be seen from the cumulative line in Figure 5. The largest library in the diversity space has a diaminoalcohol scaffold [31] and four link points to which collections of 2936, 2555, 2936 and 860 R-groups can respectively be attached, so the library contains about 1.9×10^{13} compounds. Excluding this library, the diversity space still contains $>10^{12}$ compounds. The next largest library has a galactose scaffold [32], four link points, and about 1.8×10^{12} compounds; excluding this library as well as the largest leaves $>10^{11}$ compounds in the diversity space. Finally, if the ten largest libraries are all removed,

the diversity space still contains about 10^{10} compounds. Hence, it is clear that the space derives its enormous diversity from contributions by a large number of combinatorial libraries rather than from just a few.

An important question in evaluating the quality of the diversity space is the extent to which it contains known drugs, or at least analogues of them. We therefore compared the diversity space to the MDDR database [33] of 8800 marketed and clinical-stage drugs. Analysis showed that about one fifth of the scaffolds used in the libraries of the diversity space can be embedded in at least one drug in the MDDR database. Conversely, about half of the drugs (4879) contain one of the scaffolds. This high concordance between the diversity space and known drugs is not surprising, since many of the published combinatorial libraries that we used in constructing the diversity space were devised precisely because of their relation to known classes of drugs. The good ability of the diversity space to replicate known drugs suggests that the space contains an abundance of drug-like compounds, from which a de novo drug design program can generate candidates against new target proteins.

Brief description of *de novo* design algorithm

Our own *de novo* drug design program will be described in detail elsewhere. Here we briefly describe the algorithm used by the program in order to show how it is able to utilize the very large diversity space we have constructed. A key feature is that the algorithm does not screen compounds one by one, an impossible task for a diversity space of $>10^{13}$ compounds, but instead constructs high-scoring compounds from within the space. Compounds are "created" and scored only when needed, which is made possible by the fragment marking approach to library enumeration mentioned above.

For a given target protein, we first create a hydrogen bond map of the active site using statistical data from the PDB database of protein structures, coupled with some quantum chemistry calculations. This map shows locations in the active site where hydrogen bonds with specified ligand atom types can be formed with high probability, and is used to find good initial placements of each scaffold. In addition, we calculate the scoring function on a fine 3D grid in the active site. Use of the precalculated grid, which is possible because the scoring function is additive on atoms, accelerates calculation of the binding score of compound fragments by 1000-fold.

For each selected position and conformation of a scaffold, the link points are treated one at a time. For a given link point, the allowed R-groups are successively attached. When an R-group is reached, a *hypothetical* search tree is constructed for that R-group, with each leaf consisting of a particular conformation of all the rotatable bonds in the R-group and of the bond between the scaffold and the R-group. The number of branches at each internal node is the number of allowed angles for the corresponding bond, which depends on the chemical nature of the bond and the specified angle step. Since the size of the tree can be huge, the algorithm does not visit all the nodes. Rather, a key feature of the algorithm is the use of *backtracking*. As each internal node is reached, a decision is made whether to go deeper along that branch; if not, the algorithm moves back to a previous node. To make the decision, a rapid heuristic evaluation is made of the likely score of the *tail* of the R-group, which is added to the calculated score of the already placed *head* of the R-group. If that sum is better then already achieved, the algorithm moves deeper along the branch. At the end of the procedure, the visited leaf, i.e., R-group conformation, with the best score is retained for that R-group.

These R-group's conformations are then combined with the best R-groups/conformations at the other link points, with R-group collisions eliminated, to generate complete conformations of the scaffold and attached R-groups, i.e., of a library compound. An important point is that because the best R-groups are chosen at each link point before they are combined, it is not necessary to consider each possible combination of R-groups (i.e., each compound in the library), thus changing a multiplicative problem into an additive one. This idea has been used previously [15] and is possible because our scoring function is additive [34] as noted above. Finally, after all scaffolds and their combinatorial libraries are searched, a specified number of compounds/conformations having the highest scores are provided as the output of the program.

The result of the various algorithmic techniques is that the time required to "search" a library scales only approximately as the Log of library size, making it possible to handle a very large diversity space. About three months are required on a single microprocessor (2 GHz) to generate candidate ligands from the diversity space of $>10^{13}$ compounds against a given target protein. However, the calculation can be naturally and easily parallelized by separately running the program with different libraries on multiple processors, and can also be accelerated by using reduced libraries as described below.

Reduced libraries

For the quick sampling of each target, smaller versions of the combinatorial libraries (and therefore of the diversity space) were prepared. Specifically, the collection of R-groups for each link point was clustered according to chemical and structural similarity, and only one representative member of each cluster was retained. Similarity was estimated by means of the Tanimoto coefficient [35], utilizing an in-house universal fragment library which can be used to assemble almost all known drugs. The library diversity was estimated by the total length of the minimal spanning tree [36], and clustering was performed using maximum dissimilarity algorithms [37]. The size of the clusters depends on the bound of dissimilarity ε. We used different values of ε for R-group collections of different initial sizes, with the aim of preparing more homogeneously sized reduced libraries. By reducing in this manner the size of most R-group collections by a factor between 3 and 10, we reduced the size of libraries by a factor between 3 and 10,000 (depending on the number of link points), without severely impairing library diversity. The run-time of our drug design program using the reduced libraries was accelerated to several weeks.

An initial concern with using the reduced diversity space was that the faster run-time would come at the expense of missing high affinity compounds from the full diversity space. However, in a number of test runs of the de novo drug design program against various proteins, we found that the score of the top compound from the reduced library approximated the score of the top compound from the full library, to within the accuracy of the scoring function in predicting binding affinity [34]. Hence, the reduced diversity space can be used in an initial pass at generating candidate binding compounds to a particular protein. The program can then be run again, but using only those full libraries corresponding to reduced libraries which yielded compounds having good affinity in the first pass. This procedure will still generate the best binding compounds, as measured by the scoring function, but with considerable savings in time. In the same way an optimization of chosen compound can be performed.

Validation of the diversity space

To further validate the diversity space as a source of drug candidates, we used our de novo drug design program to generate inhibitors of HIV-1 integrase, an important therapeutic target for treatment of AIDS and HIV infection [38]. This work will be described in detail elsewhere. Briefly, a "control" library with a styrylquinoline scaffold was first included in the diversity space, based on the β-diketone motif found in known inhibitors of HIV integrase [39], in order to determine whether the de novo design program would generate and present candidate ligands from this library. Out of the 800 ligand candidates generated by the program from the styrlquinoline library a known inhibitor was listed at the 4th position, representing a striking enrichment factor. Three styrylquinoline structures from the best scoring ligand candidates were selected for synthesis.

Other candidate inhibitors on the list of top compounds from the full diversity space of $>10^{13}$ compounds were also selected for synthesis from the α-ketoamide [40], benzimidazole [41], and triazine [42] libraries. Encouragingly, out of 22 compounds chosen for synthesis, 20 compounds (91%) were in fact synthesized without undue difficulty, verifying the synthetic accessibility of a sample of compounds from the diversity space. In experimental assays, one previously unknown compound from the styrylquinoline library inhibited the first (strand scission) step of HIV integrase with an IC50 of 2.2 μM. One compound (compound **1**; Figure 6) from the benzimidazole library inhibited with an IC50 of 15 μM, and several other compounds of the 20 synthesized

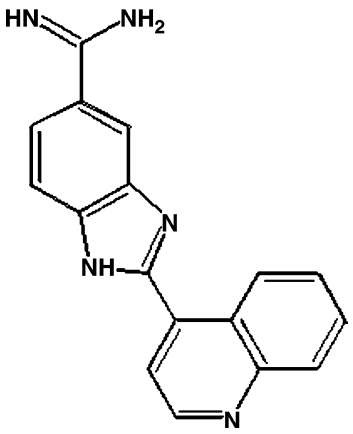


Figure 6. Compound **1**, an inhibitor of HIV-1 integrase generated from within the diversity space by the *de novo* drug design program.

showed lower activity. Importantly, the benzimidazole library was not previously known to contain inhibitors of HIV integrase. Preliminary experiments with compound **1** show that it is reasonably specific for HIV integrase, in that it does not inhibit strand scission reactions of other enzymes tested. Interestingly, this compound has some similarity with known integrase inhibitors such as DAPI (CAS 47165-04-8), but DAPI is not a member of the benzimidazole library. Moreover, DAPI is known to inhibit integrase by binding with DNA rather than integrase, whereas compound **1** appears to be a very weak binder to DNA. Compound **1** has a MW of less than 300, clogP of 3.07 and five hydrogen bond acceptors, so is very drug-like. Hence, this set of experiments proves that the diversity space contains synthesizable, drug-like candidates against a medically relevant target.

Improvement of the diversity space

We are continuing to add new libraries to the diversity space to enhance the variety of scaffolds. In addition, while most compounds in the space are synthetically accessible as intended, a certain percentage require more synthetic effort than may be considered desirable, or in some cases cannot be synthesized at all without unreasonable effort. For example, an R-reagent may contain more than one instance of the functional group used to link it to the scaffold (e.g., two amine groups or two hydroxyl groups). This may lead to a complex mixture of reaction products, with decreased yield of the desired product, or may not preserve the desired R-group at all due to side reactions. In other cases, the chemistry used to link one R-group to the scaffold may alter part of another R-group, requiring time-consuming protection and subsequent deprotection steps. Work is underway to detect and flag such undesirable R-groups and combinations of R-groups, using an advanced version of the reagent queries described above. The *de novo* drug design program can then be adjusted to either include or exclude compounds comprising such R-groups, depending on whether the chemist is or isn't willing to make extra synthetic efforts.

Conclusions

We have presented here an approach to creating an ultra-large but synthetically accessible diversity space of drug-like compounds, for use with *de novo* drug design programs. The diversity space created with this method currently comprises about 400 combinatorial libraries, many of them "expanded" from published libraries, and contains $>10^{13}$ distinct compounds. While conceptually straightforward, this approach requires the concerted efforts of a team of knowledgeable chemists with programming assistance to implement, and thus has not previously been exploited in published work. The utility of the diversity space has been demonstrated by its ability to provide new, drug-like HIV integrase inhibitors with μM potency.

Appendix A

1. An, H., Haly, B.D. and Cook, P.D., J. Med. Chem., 41 (1998) 706.
2. An, H., Haly, B.D. and Cook, P.D., Bioorg. Med. Chem. Lett., 8 (1998) 2345.
3. Andres, C.J., Bronson, J.J., D'Andrea, S.V., Deshpande, M.S., Falk, P.J., Grant-Young, K.A. and Harte, W.E., Bioorg. Med. Chem. Lett., 10 (2000) 715.
4. Barn, D.R. and Morphy, J.R., J. Comb. Chem., 1 (1999) 151.
5. Basso, A., Pegg, N., Evans, B. and Bradley, M., Eur. J. Org. Chem., (2000) 3887.

6. Bergnes, G., Gilliam, C.L., Boisclair, M.D., Blanchard, J.L., Blake, K.V., Epstein, D.M. and Pal, K., Bioorg. Med. Chem. Lett., 9 (1999) 2849.
7. Bhandari, A., Li, B. and Gallop, M.A., Synthesis, 11 (1999) 1951.
8. Bhat, A.S., Whetstone, J.L. and Brueggemeier, R.W., J. Comb. Chem., 2 (2000) 597.
9. Bhatnagar, P., Nguyen, T.T. and Lago, M.A., Tetrahedron Lett., 39 (1998) 3885.
10. Bilodeau, M.T. and Cunningham, A.M., J. Org. Chem., 63 (1998) 2800.
11. Blackburn, C., Tetrahedron Lett., 39 (1998) 5469.
12. Blondelle, S.E., Crooks, E., Ostresh, J.M. and Houghten, R.A., Antimicrob. Agents Chemother., 43 (1999) 106.
13. Bohm, H.-J., Banner, D.W. and Weber, L., J. Comput.-Aided Mol. Design, 13 (1999) 51.
14. Bolton, G.L. and Hodges, J.C., J. Comb. Chem., 1 (1999) 130.
15. Boymond, L., Rottlander, M., Cahiez, G. and Knochel, P., Angew. Chem. Int. Ed. Engl., 37 (1998) 1701.
16. Brady, S.F., Stauffer, K.J., Lumma, W.C., Smith, G.M., Ramjit, H.G., Lewis, S.D., Lucas, B.J., Gardell, S.J., Lyle, E.A., Appleby, C.D., Cook, J.J., Holahan, M.A., Stranieri, M.T., Lynch Jr., J.J., Lin, J.H., Chen, I.-W., Vastag, K., Naylor-Olsen, A.M. and Vacca, J.P., J. Med. Chem., 41 (1998) 401.
17. Brain, C.T., Paul, J.M., Loong, Y. and Oakley, P.J., Tetrahedron Lett., 40 (1999) 3275.
18. Brandli, C. and Ward, T.R., J. Comb. Chem., 2 (2000) 42.
19. Bräse, S., Dahmen, S. and Heuts, J., Tetrahedron Lett., 40 (1999) 6201.
20. Bräse, S., Dahmen, S. and Pfefferkorn, M., J. Comb. Chem., 2 (2000) 710.
21. Bräse, S., Köbberling, J., Enders, D., Lazny, R. and Wang, M., Tetrahedron Lett., 40 (1999) 2105.
22. Brennan, T., Biddison, G., Frauendorf, A., Schwarcz, L., Keen, B., Ecker, D.J., Davis, P.W., Tinder, R. and Swayze, E.E., Biotechnol. Bioeng., 61 (1998) 33.
23. Brown, R.C.D., Castro, J.L. and Moriggi, J.D., Tetrahedron Lett., 41 (2000) 3681.
24. Burkett, B.A. and Chai, C.L.L., Tetrahedron Lett., 40 (1999) 7035.
25. Burns, C.J., Groneberg, R.D., Salvino, J.M., McGeehan, G., Condon, S.M., Morris, R., Morrissette, M., Mathew, R., Darnbrough, S., Neuenschwander, K., Scotese, A., Djuric, S.W., Ullrich, J. and Labaudiniere, R., Angew. Chem. Int. Ed. Engl., 37 (1998) 2848.
26. Carboni, B., Pourbaix, C., Carreaux, F., Deleuze, H. and Maillard, B., Tetrahedron Lett., 40 (1999) 7979.
27. Carroll, C.D., Patel, H., Johnson, T.O., Guo, T., Orlowski, M., He, Z.-M., Cavallaro, C.L., Guo, J., Oksman, A., Gluzman, I.Y., Connelly, J., Chelsky, D., Goldberg, D.E. and Dolle, R.E., Bioorg. Med. Chem. Lett., 8 (1998) 2315.
28. Carroll, C.D., Tao, S., Orlowski, M., Johnson, T.O., Gluzman, I.Y., Lauri, G., Dolle, R.E. and Goldberg, D.E., Bioorg. Med. Chem. Lett., 8 (1998) 3203.
29. Chamoin, S., Houldsworth, S., Kruse, C.G., Bakker, W.I. and Snieckus, V., Tetrahedron Lett., 39 (1998) 4179.
30. Chandrasekhar, S., Padmaja, M.B. and Raza, A., J. Comb. Chem., 2 (2000) 246.
31. Chang, Y.-T., Gray, N.S., Rosania, G.R., Sutherlin, D.P., Kwon, S., Norman, T.C. and Sarohia, R., Chem. Biol., 6 (1999) 361.
32. Chaturvedi, S., Otteson, K. and Bergot, J., Tetrahedron Lett., 40 (1999) 8205.
33. Chen, C. and Munoz, B., Tetrahedron Lett., 39 (1998) 3401.
34. Chen, J., Dixon, B.R., Dumas, J. and Brittelli, D., Tetrahedron Lett., 40 (1999) 9195.
35. Combs, J.D., Jonnalagadda, S., Gordon, E.M., Gordeev, M.F., England, B.P. and Patel, D.V., Bioorg. Med. Chem., 6 (1998) 883.
36. Craig, D., Robson, M.J. and Shaw, S.J., Synlett., (1998) 1381.
37. de Julian-Ortiz, J.V., Galvez, J., Munoz-Collado, C., Garcia-Domenech, R. and Gimeno-Cardona, C., J. Med. Chem., 42 (1999) 3308.
38. Dodd, D.S. and Kozikowski, A.P., Tetrahedron Lett., 35 (1994) 977.
39. Dodd, D.S. and Wallace, O.B., Tetrahedron Lett., 39 (1998) 5701.
40. Drewry, D.H. and Ghiron, C., Tetrahedron Lett., 41 (2000) 6989.
41. Ermann, M., Simkovsky, N.M., Roberts, S.M., Parry, D.M., Baxter, A.D. and Montana, J.G., Tetrahedron Lett., 41 (2000) 2483.

42. Far, A.R. and Tidwell, T.T., J. Org. Chem., 63 (1998) 8636.
43. Farrant, E. and Rahman, S.S., Tetrahedron Lett., 41 (2000) 5383.
44. Filigheddu, S.N., Masala, S. and Taddei, M., Tetrahedron Lett., 40 (1999) 6503.
45. Flygare, J.A., Fernandez, M. and Kearney, P.C., Tetrahedron Lett., 39 (1998) 2663.
46. Fournier, E.J.-L., Nilsson, U.J. and Hindsgaul, O., Bioorg. Med. Chem., 6 (1998) 1563.
47. Garibay, P., Toy, P.H., Hoeg-Jensen, T. and Janda, K.D., Synlett., 9 (1999) 1438.
48. Gelb, M.H., Stoddard, B.L., Wilton, D.C., Flick, K.E., Bryant, M.D. and Koduri, R.S., Bioorg. Med. Chem. Lett., 9 (1999) 1097.
49. Gelens, E., Koot, W.J., Menge, W.M.P.B., Ottenheijm, H.C.J. and Timmerman, H., Bioorg. Med. Chem. Lett., 10 (2000) 1935.
50. Gibson, C. and Kessler, H., Tetrahedron Lett., 41 (2000) 1725.
51. Goff, D. and Fernandez, J., Tetrahedron Lett., 40 (1999) 423.
52. Gomez, L., Gellibert, F., Wagner, A. and Mioskowski, C., J. Comb. Chem., 2 (2000) 75.
53. Gordeev, M.F., Biotechnol. Bioeng., 61 (1998) 13.
54. Gordeev, M.F., Luehr, G.W., Hui, H.C., Gordon, E.M. and Patel, D.V., Tetrahedron, 54 (1998) 15879.
55. Gray, N.S., Wodicka, L., Thunnissen, A.-M.W.H., Norman, T.C., Kwon, S., Espinoza, F.H., Morgan, D.O., Barnes, G., LeClerc, S., Meijer, L., Kim, S.-H., Lockhart, D.J. and Schultz, P.G., Science, 281 (1998) 533.
56. Grellier, P., Vendeville, S., Sergheraert, C., Deprez, B., Davioud-Charvet, E. and Bourel, L., Bioorg. Med. Chem. Lett., 9 (1999) 437.
57. Grigg, R., Major, J.P., Martin, F.M. and Whittaker, M., Tetrahedron Lett., 40 (1999) 7709.
58. Grosche, P., Holtzel, A., Wolk, T.B., Trautwein, A.W. and Jung, G., Synthesis, 11 (1999) 1961.
59. Guillier, F., Roussel, P., Moser, H., Kane, P. and Bradley, M., Chem. Eur. J., 5 (1999) 3450.
60. Habermann, J., Ley, S.V., Scicinski, J.J., Scott, J.S., Smits, R. and Thomas, A.W., J. Chem. Soc., Perkin Trans. 1, (1999) 1253.
61. Habermann, J., Ley, S.V. and Smits, R., J. Chem. Soc., Perkin Trans. 1, (1999) 2421.
62. Hamper, B.C., Gan, K.Z. and Owen, T.J., Tetrahedron Lett., 40 (1999) 4973.
63. Hamper, B.C., Jerome, K.D., Yalamanchili, G., Walker, D.M., Chott, R.C. and Mischke, D.A., Biotechnol. Bioeng., 71 (2000) 28.
64. Han, Y., Roy, A. and Giroux, A., Tetrahedron Lett., 41 (2000) 5447.
65. Haq, W., Chauhan, P.M.S. and Srivastava, S.K., Bioorg. Med. Chem. Lett., 9 (1999) 965.
66. Haque, T.S., Skillman, A.G., Lee, C.E., Habashita, H., Gluzman, I.Y., Ewing, T.J.A., Goldberg, D.E. and Kuntz, I.D., J. Med. Chem., 42 (1999) 1428.
67. Harikrishnan, L.S. and Showalter, H.D.H., Tetrahedron, 56 (2000) 515.
68. Haunert, F., Bolli, M.H., Hinzen, B. and Ley, S.V., J. Chem. Soc., Perkin Trans. 1, (1998) 2235.
69. Havez, S., Begtrup, M. and Vedso, P., J. Org. Chem., 63 (1998) 7418.
70. Heizmann, G., Hildebrand, P., Tanner, H., Ketterer, S., Pansky, A., Froidevaux, S., Beglinger, C. and Eberle, A.N., J. Recept. Signal Transduction Res., 19 (1999) 449.
71. Higaki, J.N., Chakravarty, S., Bryant, C.M., Cowart, L.R., Harden, P., Scardina, J.M., Mavunkel, B., Luedtke, G.R. and Cordell, B., J. Med. Chem., 42 (1999) 3889.
72. Hu, Y., Baudart, S. and Porco Jr., J.A., J. Org. Chem., 64 (1999) 1049.
73. Hulme, C. and Cherrier, M.-P., Tetrahedron Lett., 40 (1999) 5295.
74. Hulme, C., Ma, L., Romano, J.J., Morton, G., Tang, S.-Y., Cherrier, M.P., Choi, S., Salvino, J. and Labaudiniere, R., Tetrahedron Lett., 41 (2000) 1889.
75. Josey, J.A., Tarlton, C.A. and Payne, C.E., Tetrahedron Lett., 39 (1998) 5899.
76. Kallus, C., Opatz, T., Wunberg, T., Schmidt, W., Henke, S. and Kunz, H., Tetrahedron Lett., 40 (1999) 7783.
77. Katritzky, A.R., Belyakov, S.A. and Tymoshenko, D.O., J. Comb. Chem., 1 (1999) 173.
78. Katritzky, A.R., Fang, Y., Kiely, J.S. and Belyakov, S.A., Tetrahedron Lett., 39 (1998) 8051.
79. Katritzky, A.R., Qi, M., Feng, D., Zhang, G., Griffith, M.C. and Watson, K., Org. Lett., 1 (1999) 1189.
80. Kim, S.W., Hong, J.S., Koh, C.Y., Lee, E.J. and Lee, K., Mol. Diversity, 3 (1998) 133.
81. Kim, S.W., Koh, J.S., Lee, E.J. and Ro, S., Mol. Diversity, 3 (1998) 129.

82. Kobayashi, S. and Akiyama, R., Tetrahedron Lett., 39 (1998) 9211.
83. Kobayashi, S., Furuta, T., Sugita, K., Okitsu, O. and Oyamada, H., Tetrahedron Lett., 40 (1999) 1341.
84. Koch, K., Piscopio, A.D. and Miller, J.F., Tetrahedron, 55 (1999) 8189.
85. Kondo, Y., Komine, T., Fujinami, M., Uchiyama, M. and Sakamoto, T., J. Comb. Chem., 1 (1999) 123.
86. Kraxner, J., Arlt, M. and Gmeiner, P., Synlett., 1 (2000) 125.
87. Krchnak, V., Szabo, L. and Vagner, J., Tetrahedron Lett., 41 (2000) 2835.
88. Kuang, R., Epp, J.B., Ruan, S., Yu, H., Huang, P., He, S., Tu, J., Schechter, N.M., Turbov, J., Froelich, C.J. and Groutas, W.C., J. Am. Chem. Soc., 121 (1999) 8128.
89. Kulkarni, B.A. and Ganesan, A., Tetrahedron Lett., 40 (1999) 5637.
90. Kung, P.-P. and Cook, P.D., Biotechnol. Bioeng., 61 (1998) 119.
91. Lam, P.Y.S., Ru, Y., Jadhav, P.K., Aldrich, P.E., DeLucca, G.V., Eyermann, C.J., Chang, C.-H., Emmett, G., Holler, E.R., Daneker, W.F., Li, L., Confalone, P.N., McHugh, R.J., Han, Q., Li, R., Markwalder, J.A., Seitz, S.P., Sharpe, T.R., Bacheler, L.T., Rayner, M.M., Klabe, R.M., Shum, L., Winslow, D.L., Kornhauser, D.M., Jackson, D.A., Erickson-Viitanen, S. and Hodge, C.N., J. Med. Chem., 39 (1996) 3514.
92. Lee, H.B. and Balasubramanian, S., Org. Lett., 2 (2000) 323.
93. Lee, J., Gauthier, D. and Rivero, R.A., J. Org. Chem., 64 (1999) 3060.
94. Lee, K.J., Angulo, A., Ghazal, P. and Janda, K.D., Org. Lett., 1 (1999) 1859.
95. Lepore, S.D. and Wiley, M.R., J. Org. Chem., 64 (1999) 4547.
96. Lepore, S.D. and Wiley, M.R., J. Org. Chem., 65 (2000) 2924.
97. Lynas, J.F., Martin, S.L., Walker, B., Baxter, A.D., Bird, J., Bhogal, R., Montana, J.G. and Owen, D.A., Comb. Chem. High Throughput Scr., 3 (2000) 37.
98. Mackman, R.L., Katz, B.A., Breitenbucher, J.G., Hui, H.C., Verner, E., Luong, C., Liu, L. and Sprengeler, P.A., J. Med. Chem., 44 (2001) 3856.
99. Makino, S., Suzuki, N., Nakanishi, E. and Tsuji, T., Tetrahedron Lett., 41 (2000) 8333.
100. Maloney, P.R., Parks, D.J., Haffner, C.D., Fivush, A.M., Chandra, G., Plunket, K.D., Creech, K.L., Moore, L.B., Wilson, J.G., Lewis, M.C., Jones, S.A. and Willson, T.M., J. Med. Chem., 43 (2000) 2971.
101. Marder, M., Viola, H., Bacigaluppo, J.A., Colombo, M.I., Wasowski, C., Wolfman, C., Medina, J.H., Ruveda, E.A. and Paladini, A.C., Biochem. Biophys. Res. Commun., 249 (1998) 481.
102. Masala, S. and Taddei, M., Org. Lett., 1 (1999) 1355.
103. Matthews, D.P., Green, J.E. and Shuker, A.J., J. Comb. Chem., 2 (2000) 19.
104. Mayer, J.P., Lewis, G.S., McGee, C. and Bankaitis-Davis, D., Tetrahedron Lett., 39 (1998) 6655.
105. McKibben, B.P., Cartwright, C.H. and Castelhano, A.L., Tetrahedron Lett., 40 (1999) 5471
106. Melikian-Badalian, A., Hauske, J.R., Kumaravel, G. and Hoemann, M.Z., Tetrahedron Lett., 39 (1998) 4749.
107. Miel, H. and Rault, S., Tetrahedron Lett., 39 (1998) 1565.
108. Miller, M.W., McCombie, S.W. and Vice, S.F., Tetrahedron Lett., 39 (1998) 3429.
109. Miller, P.C., Owen, T.J., Molyneaux, J.M., Curtis, J.M. and Jones, C.R., J. Comb. Chem., 1 (1999) 223.
110. Miyabe, H., Fujishima, Y. and Naito, T., J. Org. Chem., 64 (1999) 2174.
111. Mjalli, A.M.M. and Cheng, J.-F., Tetrahedron Lett., 39 (1998) 939.
112. Mohan, R., Yun, W., Buckman, B.O., Liang, A., Trinh, L. and Morrissey, M.M., Bioorg. Med. Chem. Lett., 8 (1998) 1877.
113. Mohan, V., Wang, T., Griffey, R.H., An, H. and Cook, P.D., Tetrahedron, 54 (1998) 3999.
114. Munson, M.C., Cook, A.W., Josey, J.A. and Rao, C., Tetrahedron Lett., 39 (1998) 7223.
115. Nefzi, A., Dooley, C., Ostresh, J.M. and Houghten, R. A., Bioorg. Med. Chem. Lett., 8 (1998) 2273.
116. Nefzi, A., Giulianotti, M. and Houghten, R.A., Tetrahedron Lett., 39 (1998) 3671.
117. Nefzi, A., Giulianotti, M.A. and Houghten, R.A., Tetrahedron, 56 (2000) 3319.

118. Nefzi, A., Ostresh, J.M., Giulianotti, M. and Houghten, R.A., J. Comb. Chem., 1 (1999) 195.
119. Nicolaou, K.C., Cao, G.Q. and Pfefferkorn, J.A., Angew. Chem. Int. Ed. Engl., 39 (2000) 734.
120. Nicolaou, K.C., Pfefferkorn, J.A., Roecker, A.J., Cao, G.-Q., Barluenga, S. and Mitchell, H.J., J. Am. Chem. Soc., 122 (2000) 9939.
121. Nicolaou, K.C., Snyder, S.A., Bigot, A. and Pfefferkorn, J.A., Angew. Chem. Int. Ed. Engl., 43 (2000) 2971.
122. Nicolaou, K.C., Winssinger, N., Vourloumis, D., Ohshima, T., Kim, S., Pfefferkorn, J., Xu, J.-Y. and Li, T., J. Am. Chem. Soc., 120 (1998) 10814.
123. Organ, M.G. and Dixon, C.E., Biotechnol. Bioeng., 71 (2000) 71.
124. Ostresh, J.M., Schoner, C.C., Hamashin, V.T., Nefzi, A., Meyer, J.-P. and Houghten, R.A., J. Org. Chem., 63 (1998) 8622.
125. Park, K.-H. and Kurth, M.J., Tetrahedron Lett., 40 (1999) 5841.
126. Parlow, J.J. and Flynn, D.L., Tetrahedron, 54 (1998) 4013.
127. Parrot, I., Wermuth, C.-G. and Hibert, M., Tetrahedron Lett., 40 (1999) 7975.
128. Pirrung, M.C. and Tumey, L.N., J. Comb. Chem., 2 (2000) 675.
129. Pons, F.J., Mishir, Q., Nouvet, A. and Brookfield, F., Tetrahedron Lett., 41 (2000) 4965.
130. Qian, X. and Liang, G.-B., Bioorg. Med. Chem. Lett., 9 (1999) 2101.
131. Ramdas, L., Bunnin, B.A., Plunkett, M.J., Sun, G., Ellman, J., Gallick, G. and Budde, R.J.A., Arch. Biochem. Biophys., 368 (1999) 394.
132. Richter, H., Walk, T., Holtzel, A. and Jung, G., J. Org. Chem., 64 (1999) 1362.
133. Rohrer, S.P., Birzin, E.T., Mosley, R.T., Berk, S.C., Hutchins, S.M., Shen, D.-M., Xiong, Y., Hayes, E.C., Parmar, R.M., Foor, F., Mitra, S.W., Degrado, S.J., Shu, M., Klopp, J.M., Cai, S.-J., Blake, A., Chan, W.W.S., Pasternak, A., Yang, L., Patchett, A.A., Smith, R.G., Chapman, K.T. and Schaeffer, J.M., Science, 282 (1998) 737.
134. Rottlander, M. and Knochel, P., J. Comb. Chem., 1 (1999) 181.
135. Rudolph, J., Tetrahedron, 56 (2000) 3161.
136. Saha, A.K., Liu, L., Simoneaux, R.L., Kukla, M.J., Marichal, P. and Odds, F., Bioorg. Med. Chem. Lett., 10 (2000) 2175.
137. Salvino, J.M., Kumar, N.V., Orton, E., Airey, J., Kiesow, T., Crawford, K., Mathew, R. and Krolikowski, P., J. Comb. Chem., 2 (2000) 691.
138. Sams, C.K. and Lau, J., Tetrahedron Lett., 40 (1999) 9359.
139. Scharn, D., Wenschuh, H., Reineke, U., Schneider-Mergener, J. and Germeroth, L., J. Comb. Chem., 2 (2000) 361.
140. Schlienger, N., Bryce, M.R. and Hansen, T.K., Tetrahedron, 56 (2000) 10023.
141. Schurer, S.S., Ankener, W., Kogan, T.P., Virgilio, A.A., Souers, A.J., Vanderslice, P., West, H.E. and Ellman, J.A., Bioorg. Med. Chem. Lett., 8 (1998) 2297.
142. Schwartz, R.S., Jordan, D.B., Wawrzak, Z., Jennings, L.D. and Amorose, D., Bioorg. Med. Chem. Lett., 9 (1999) 2509.
143. Schwarz, M.K., Tumelty, D. and Gallop, M.A., J. Org. Chem., 64 (1999) 2219.
144. Scicinski, J.J., Barker, M.D., Murray, P.J. and Jarvie, E.M., Bioorg. Med. Chem. Lett., 8 (1998) 3609.
145. Shestopalov, A.M., Kislyi, V.P., Kruglova, E.Y., Nikishin, K.G. and Semenov, V.V., J. Comb. Chem., 2 (2000) 24.
146. Shi, S., Xiao, X.-Y. and Czarnik, A.W., Biotechnol. Bioeng., 61 (1998) 7.
147. Sieber, F., Wentworth Jr., P. and Janda, K.D., J. Comb. Chem., 1 (1999) 540.
148. Siev, D.V., Gaudette, J.A. and Semple, J.E., Tetrahedron Lett., 40 (1999) 5123.
149. Silen, J.L., Lu, A.T., Solas, D.W., Gore, M.A., Maclean, D., Shah, N.H., Coffin, J.M., Bhinderwala, N.S. and Wang, Y., Antimicrob. Agents Chemother., 42 (1998) 1447.
150. Sim, M.M., Lee, C.L. and Ganesan, A., Tetrahedron Lett., 39 (1998) 6399.
151. Smith, A.L., Stevenson, G.I., Lewis, S., Patel, S. and Castro, J.L., Bioorg. Med. Chem. Lett., 10 (2000) 2693.
152. Smith, A.L., Stevenson, G.I., Swain, C.J. and Castro, J.L., Tetrahedron Lett., 39 (1998) 8317.
153. Smith, J.M. and Krchnak, V., Tetrahedron Lett., 40 (1999) 7633.

154. Smith, P.W., Sollis, S.L., Howes, P.D., Cherry, P.C., Starkey, I.D., Cobley, K.N., Weston, H., Scicinski, J., Merritt, A., Whittington, A., Wyatt, P., Taylor, N., Green, D., Bethell, R., Madar, S., Fenton, R.J., Morley, P.J., Pateman, T. and Beresford, A., J. Med. Chem., 41 (1998) 787.
155. Souers, A.J., Virgilio, A.A., Rosenquist, A., Fenuik, W. and Ellman, J.A., J. Am. Chem. Soc., 121 (1999) 1817.
156. South, M.S., Dice, T.A. and Parlow, J., Biotechnol. Bioeng., 71 (2000) 51.
157. Stenzel, W., Konig, W.A., Schotten, T. and Blettner, C.G., Tetrahedron Lett., 40 (1999) 2101.
158. Stephensen, H. and Zaragoza, F., Tetrahedron Lett., 40 (1999) 5799.
159. Stieber, F., Grether, U. and Waldmann, H., Angew. Chem. Int. Ed. Engl., 38 (1999) 1073.
160. Sturino, C.F. and Labelle, M., Tetrahedron Lett., 39 (1998) 5891.
161. Sullivan, R.W., Bigam, C.G., Erdman, P.E., Palanki, M.S.S., Anderson, D.W., Goldman, M.E., Ransone, L.J. and Suto, M.J., J. Med. Chem., 41 (1998) 413.
162. Sun, S. and Murray, W.V., J. Org. Chem., 64 (1999) 5941.
163. Sunami, S., Sagara, T., Ohkubo, M. and Morishima, H., Tetrahedron Lett., 40 (1999) 1721.
164. Suszmuth, R.D., Trautwein, A.W. and Jung, G., Bioorg. Med. Chem. Lett., 8 (1998) 2381.
165. Suto, M.J., Sullivan, R., Gayo-Fung, L.M. and Palanki, M.S.S., Tetrahedron, 54 (1998) 4141.
166. Szardening, A.K., Harris, D., Lam, S., Shi, L., Tien, D., Wang, Y., Patel, D.V., Navre, M. and Campbell, D.A., J. Med. Chem., 41 (1998) 2194.
167. Tadesse, S., Bhandari, A. and Gallop, M.A., J. Comb. Chem., 1 (1999) 184.
168. Thompson, A., Moore, F.L., Moon, Y.-C. and Ellman, J. A., J. Org. Chem., 63 (1998) 2066.
169. Tomasi, S., Corbel, J.-C., Renault, J., Roch, M.L., Carboni, B., Martin, B., Delcros, J.-G., Moncoq, D. and Uriac, P., Bioorg. Med. Chem. Lett., 8 (1998) 635.
170. Tumelty, D., Schwarz, M.K., Cao, K. and Needels, M.C., Tetrahedron Lett., 40 (1999) 6185.
171. Verner, E., Katz, B.A., Spencer, J.R., Allen, D., Hataye, J., Hruzewicz, W., Hui, H.C., Kolesnikov, A., Li, Y., Luong, C., Martelli, A., Radika, K., Rai, R., She, M., Shrader, W., Sprengeler, P.A., Trapp, S., Wang, J., Young, W.B. and Mackman, R.L., J. Med. Chem., 44 (2001) 2753.
172. Wahhab, A. and Leban, J., Tetrahedron Lett., 41 (2000) 1487.
173. Wang, Y. and Huang, T.-N., Tetrahedron Lett., 39 (1998) 9605.
174. Warmus, J.S., Ryder, T.R., Hodges, J.C., Kennedy, R.M. and Brady, K.D., Bioorg. Med. Chem. Lett., 8 (1998) 2309.
175. Washizuka, K.I., Nagai, K., Minakata, S., Ryu, I. and Komatsu, M., Tetrahedron Lett., 41 (2000) 691.
176. Watanabe, Y., Ishikawa, S., Takao, G. and Toru, T., Tetrahedron Lett., 40 (1999) 3411.
177. Wei, G.P. and Phillips, G.B., Tetrahedron Lett., 39 (1998) 179.
178. Wendeborn, S., De Mesmaeker, A. and Brill, W.K.-D., Synlett., (1998) 865.
179. Wilson, M.W., Hernandez, A.S., Calvet, A.P. and Hodges, J.C., Mol. Diversity, 3 (1998) 95.
180. Wong, C.-H., Hendrix, M., Manning, D.D., Rasenbohm, C. and Greenberg, W.A., J. Am. Chem. Soc., 120 (1998) 8319.
181. Wu, S. and Janusz, J.M., Tetrahedron Lett., 41 (2000) 1165.
182. Xie, Y.F., Whitten, J.P., Chen, T.Y., Liu, Z. and McCarthy, J.R., Tetrahedron, 54 (1998) 4077.
183. Yager, K.M., Johnson, C.R., Fantauzzi, P., Zhang, B. and Hocker, M., Tetrahedron, 54 (1998) 4097.
184. Yamamoto, Y., Ajito, K. and Othsuka, Y., Chem. Lett., (1998) 379.
185. Yang, R.Y. and Kaplan, A., Tetrahedron Lett., 41 (2000) 7005.
186. Yeh, C.-M. and Sun, C.-M., Tetrahedron Lett., 40 (1999) 7247.
187. Yokum, T.S., Alsina, J. and Barany, G., J. Comb. Chem., 2 (2000) 282.
188. Young, R.N., Huang, Z., Perrier, H., Giroux, A., Bayly, C.I., Lepine, C., Laliberte, F. and Han, Y., Tetrahedron, 55 (1999) 11669.

189. Zaragoza, F. and Stephensen, H., J. Org. Chem., 64 (1999) 2555.
190. Zhang, H.-C., Brumfield, K.K., Jaroskova, L. and Maryanoff, B.E., Tetrahedron Lett., 39 (1998) 4449.

References

1. Schneider, G. and Böhm, H.-J., Drug Discov. Today, 7 (2002) 64.
2. http://www.mdli.com/products/experiment/available_chem_dir/index.jsp
3. Nishibata, Y. and Itai, A., Tetrahedron, 41 (1991) 8985.
4. Gehlhaar, D.K., Moerder, K.E., Zichi, D., Sherman, C.J., Ogden, R.C. and Freer, S.T., J. Med. Chem., 38 (1995) 466.
5. Wang, R., Gao, Y. and Lai, L., J. Mol. Model, 6 (2000) 498.
6. Clark, D.E., Frenkel, D., Levy, S.A., Li, J., Murray, C.W., Robson, B., Waszkowycz, B. and Westhead, D.R., J. Comput.-Aided Mol. Design, 9 (1995) 13.
7. Murray, C.W., Clark, D.E., Auton, T.R., Firth, M.A., Li, J., Sykes, R.A., Waszkowycz, B., Westhead, D.R. and Young, S.C., J. Comput.-Aided Mol. Design, 11 (1997) 193.
8. Sun, Y., Ewing, T.J.A., Skillman, A.G. and Kuntz, I.D., J. Comput.-Aided Mol. Design, 12 (1998) 597.
9. Leach, A.R., Bradshaw, J., Green, D.V.S., Hann, M.M., Delany III, J.J., J. Chem. Inf. Comput. Sci., 39 (1999) 1161.
10. Martin, E.J. and Critchlow, R.E., J. Comb. Chem., 1 (1999) 32.
11. Shi, S., Peng, Z., Kostrowicki, J., Paderes, G. and Kuki, A., J. Mol. Graph. Model, 18 (2000) 478.
12. Sheridan, R.P., SanFeliciano, S.G. and Kearsley, S.K., J. Mol. Graph. Model, 18 (2000) 320.
13. Lobanov, V.S. and Agrafiotis, D.K., Comb. Chem. High Throughput Screen, 5 (2002) 167.
14. Cramer, R.D., Patterson, D.E., Clark, R.D., Soltanshahi, F. and Lawless, M.S., J. Chem. Inf. Comput. Sci., 38 (1998) 1010.
15. Kick, E.K., Roe, D.C., Skillman, A.G., Liu, G., Ewing, T.J.A., Sun, Y., Kuntz, I.D. and Ellman, J.A., Chem. Biol., 4 (1997) 297.
16. Bräse, S., Dahmen, S. and Pfefferkorn, M., J. Comb. Chem., 2 (2000) 710.
17. Dolle, R.E. and Nelson Jr., K.H., J. Comb. Chem., 1 (1999) 235.
18. Dolle, R.E., J. Comb. Chem., 2 (2000) 383.
19. Dolle, R.E., J. Comb. Chem., 3 (2001) 477.
20. Dolle, R.E., J. Comb. Chem., 4 (2002) 1.
21. Franzen, R.G., J. Comb. Chem., 2 (2000) 195.
22. Gustafson, G.R., Baldino, C.M., O'Donnell, M.-M.E., Sheldon, A., Tarsa R.J., Verni, C.J. and Coffen, D.L., Tetrahedron, 54 (1998) 4051.
23. Sigma-Aldrich Co., http://www.sigmaaldrich.com.
24. Acros Organics, http://www.acros.be.
25. Lancaster Synthesis, http://www.lancastersynthesis.com
26. TCI, Tokyo Kasei Kogyo Co., Ltd., http://www.tokyo-kasei.co.jp/catalogue/index.html.
27. Ullman, J.R., J. Associat. Comp. Machinery, 23 (1976) 31.
28. HyperChem(TM), Hypercube, Inc., Gainesville, Florida, USA, http://www.hyper.com.
29. Smith, A.L., Stevenson, G.I., Swain, C.J. and Castro, J.L., Tetrahedron Lett., 39 (1998) 8317.
30. Lipinski, C.A., Lombardo, F., Dominy, B.W. and Feeney, P.J., Adv. Drug Deliv. Rev., 23 (1997) 3.
31. Haque, T.S., Skillman, A.G., Lee, C.E., Habashita, H., Gluzman, I.Y., Ewing, T.J.A., Goldberg, D.E., Kuntz, I.D. and Ellman, J.A., J. Med. Chem., 42 (1999) 1428.
32. Kallus, C., Opatz, T., Wunberg, T., Schmidt, W., Henke, S. and Kunz, H., Tetrahedron Lett., 40 (1999) 7783.
33. MDL Drug Data Report (MDDR), http://www.mdli.com
34. Muryshev, A.E., Tarasov, D.N., Butygin, A.V., Butygina, O.Yu., Aleksandrov, A.B. and Nikitin, S.M., J. Comput.-Aided Mol. Design, 17 (2003) 597.
35. Willett, P., Barnard, J.M. and Downs, G.M., J. Chem. Inf. Comput. Sci., 38 (1998) 983.
36. Clark, R.D., In Ghose, A.K. and Viswanadhan, V.N. (Eds.), Combinatorial Library Design and Evaluation, Marcel Dekker, Inc., New York, 2001, pp. 337–362.
37. Gillet, V.J. and Willett, P., In Ghose, A.K. and Viswanadhan, V.N. (Eds.), Combinatorial Library Design and Evaluation, Marcel Dekker, Inc., New York, 2001, pp. 379–398
38. Sayasith, K., Sauve, G. and Yelle, J., Exp. Opin. Ther. Targets, 5 (2001) 444.
39. Zouhiri, F., Mouscadet, J-F., Mekouar, K., Desmaële, D., Savoure, D., Leh, H., Subra, F., Le Bret, M., Auclair, C. and d'Angelo J., J. Med. Chem., 43 (2000) 1533.
40. South, M.S., Dice, T.A. and Parlow, J.J., Biotechnol. Bioeng., 71 (2000) 51.
41. Verner, E., Katz, B.A., Spencer, J.R., Allen, D., Hataye, J., Hruzewicz, W., Hui, H.C., Kolesnikov, A., Li, Y., Luong, C., Martelli, A., Radika, K., Rai, R., She, M., Shrader, W., Sprengeler, P.A., Trapp, S., Wang, J., Young, W.B. and Mackman, R.L., J. Med. Chem., 44 (2001) 2753.
42. Silen, J.L., Lu, A.T., Solas, D.W., Gore, M.A., Maclean, D., Shah, N.H., Ciffin, J.M., Bhinderwala, N.S., Wang, Y., Tsutsui, K.T., Look, G.C., Campbell, D.A., Hale, R.L., Navre, M. and DeLuca-Flaherty, C.R., Antimicrob. Agents Chemother., 42 (1998) 1447.