A Fast Algorithm for Searching for Molecules Containing a Pharmacophore in Very Large Virtual Combinatorial Libraries

Roberto Olender* and Rakefet Rosenfeld†

Molecular Modeling Group, Peptor Ltd., Kiryat Weizmann 16, Rehovot 76326, Israel

Received October 18, 2000

We present a new algorithm for identifying molecules that display a pharmacophore, or in general a structural motif, by efficiently constructing and screening huge virtual combinatorial libraries of diverse compounds. The uniqueness of this algorithm is its ability to build and screen libraries of ca. 10^{18} 3D molecular conformations within a reasonable time scale, thereby increasing the chemical space that can be virtually screened by many orders of magnitude. The algorithm may be used to design new molecules that display a desired pharmacophore on predefined sets of chemical scaffolds. This is demonstrated herein by screening a library of backbone cyclic peptides to find candidate peptido- and proteinomimetics.

INTRODUCTION

Recent advances in the fields of combinatorial chemistry and high-throughput screening[1,2] enable the rapid synthesis and biological screening of huge numbers of diverse compounds. Nevertheless, even the largest combinatorial libraries that can be synthesized and screened constitute only a very small subset of the available chemical space of all possible molecules.[3] Thus, to create libraries focused around a subspace of interest, computational tools are often employed.

The major aim of these tools is to virtually screen large numbers of compounds in silico, enabling rejection of both molecules with undesirable properties and molecules that lack desired attributes.[4,5] Optimally, the virtual libraries screened should be as large as possible, yet in practice their size is limited by the time required to generate the molecules and, depending on the attributes used for screening, the time required for the actual screen. For example, using only 1D and 2D information, it is currently possible to virtually screen a library of 10^9 compounds, a relatively small subset of chemical space.[4] The limitation on the number of compounds that can be generated is commonly tackled by avoiding full enumeration of the library. This is done, for example, by using either similarity clustering methods[6] or filtration methods such as REOS[4] that avoid generating compounds with undesirable properties on the basis of chemical/medicinal knowledge; for example, molecules containing moieties that are known to be toxic are not generated. Andrews et al.[7] have screened a scaffold-based virtual library of 10^{13} different small molecules using topomer shape similarity, a 2D topological similarity metric.[8] Libraries of this type can be screened efficiently because fragment descriptors are analyzed independently, and then combined to obtain the overall descriptors of each molecule.

Following the initial 1D- and 2D-based filtering and screening, 3D information comes into play. Structural information on a binding site can be used to dock candidates and estimate their binding potential, or structural information on a set of active compounds can be used to deduce a pharmacophore, which can be used to virtually screen additional compounds. Both these approaches require the generation of 3D conformations of the virtual libraries, and to date are limited to libraries of $10^4 - 10^7$ conformers.[4,9−11]

The pharmacophore—the 3D arrangement of the chemical groups (also referred to as descriptors) required for biological activity—is a useful concept in library design. Pharmacophores can be obtained directly—either inferred from the experimentally determined structure of a receptor−ligand complex or obtained by a shape−function complementarity analysis of a receptor binding site[12]—or indirectly in cases where the structure of the receptor is unknown. In the latter cases the pharmacophore is inferred from a series of active compounds, usually by calculating all possible superpositions of predefined chemical groups (possible pharmacophore descriptors) on all active compounds.[13,14]

Once a pharmacophore is determined, a classical approach in the drug optimization process is to make small, often systematic changes in the atoms or chemical groups that comprise the pharmacophore descriptors in a set of lead compounds (the "lead explosion" approach). While in some cases this approach results in analogues with increased binding affinities to the desired target, its major drawback is that the resulting molecules are generally similar to the original leads. Thus, if a drug candidate that originated from a specific lead fails at a later stage of the drug development process, as often happens, for reasons that are not directly related to its target-binding capabilities (e.g., toxicity, bioavailability), there is a good chance that all of the other compounds originating from the same lead will fail likewise.

Virtual screening methods overcome this drawback by using the pharmacophore to screen in silico large libraries of diverse compounds.[15,16] The result of this search is a set of molecules that display the pharmacophore, and therefore

* To whom correspondence should be addressed. E-mail: roberto@peptor.co.il. Phone: +972-8-9387763.

† Current address: QBI Enterprises Ltd., Weizmann Science Park, P.O. Box 4071, Nes Ziona 70400, Israel.

10.1021/ci000463o CCC: $20.00
Published on Web 03/15/2001

have a good chance of being bioactive, yet differ in the other parts of the molecule, and therefore are expected to have different pharmacological properties. The larger the diversity of the molecules identified in the search, the larger the chance that one of these compounds later becomes a drug. Thus, the virtual libraries should span intelligently the entire chemical space that has the potential to be relevant to the target of interest. The virtually screened libraries may be truly virtual, existing solely on the computer or corporately/commercially/publicly available databases of existing compounds.

In this work we present a new computer algorithm capable of representing very large databases, comprising orders of magnitude of 10^{18} conformers, and efficiently screening these (within minutes) for molecules that display a given pharmacophore. The high level of efficiency is obtained by (a) generating and screening a much smaller number of abstract structures, each representing several orders of magnitude of possible molecules, thus avoiding the necessity to enumerate all conformers of all compounds, and (b) a set of filters that efficiently eliminate candidates that are incompatible with the pharmacophore. The strength of the algorithm is demonstrated herein using a library of backbone cyclic[17] peptidic scaffolds to identify potential peptidomimetic and proteinomimetic analogues.

DESCRIPTION OF THE ALGORITHM

The algorithm requires as input a library of chemical scaffolds and a library of substituents, which are combined to form a "virtual combinatorial library" (VCL), and a pharmacophore, against which the virtual combinatorial library is searched (Figure 1). The scaffolds and substituents are each represented by a set of discrete 3D conformations (Figure 1a,b), the size of which is dictated by the number of degrees of freedom of each compound and the available computational resources. We refer in the following to the substituent conformers as rotamers. For each scaffold a set of attachment points is predefined along with attachment rules that dictate which substituents can be attached at each point. The first step of the algorithm is the combination of the two libraries to create the VCL. This is accomplished by attaching to each attachment point of each scaffold all rotamers of all substituents concurrently. Each of the resulting supermolecules of the VCL, depicted schematically in Figure 2, represents a combinatorially huge number of 3D conformations of different molecules. The pharmacophore (Figure 1c)—the 3D configuration of chemical groups that are associated with the biological activity being targeted—may be comprised of atoms, pseudoatoms (e.g., center of aromatic rings), or atoms and vectors, e.g., potential hydrogen bond donors or acceptors and their corresponding binding directions.[18]

Following the construction of the VCL, each supermolecule is submitted to a series of filtration steps, in which scaffolds and substituent rotamers that are incompatible with the pharmacophore, i.e., cannot possibly form a molecule that presents the pharmacophore, are eliminated. The filters are a set of necessary conditions for the existence of the pharmacophore and are designed such that their evaluation on the supermolecules is straightforward and fast. The filters are applied to each of the rotamers attached to the supermolecule, and rotamers that do not meet the filter condition are eliminated. Filtration is performed in an efficient way by applying the filters in increasing order of complexity, such that the first filters are simple and involve small computational efforts while the subsequent ones are more computationally demanding. Thus, the initial filters are applied to rotamers on supermolecules very rich in substituents, while the subsequent ones are applied to supermolecules that have been depleted of large numbers of substituents. An additional level of efficiency is attained by initially applying filters that

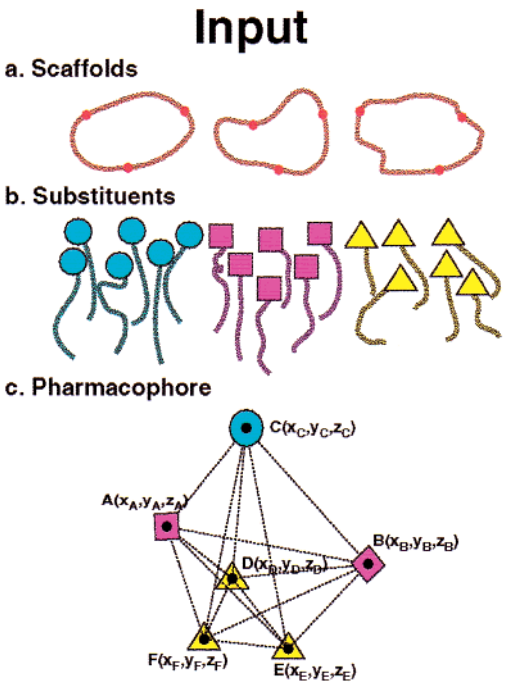

Figure 1. Input for the algorithm. (a) A library of scaffolds. The figure schematically depicts three conformations of a single scaffold having three attachment points (red dots) to which the substituents are connected. (b) A library of substituents. The figure schematically depicts three different substituents, each represented by six 3D conformations (rotamers). (c) A pharmacophore. The points A, B, C, D, E, and F are the descriptors of the pharmacophore, which is comprised of three different chemical groups (descriptor types), represented by a blue circle, yellow triangle, and magenta square.

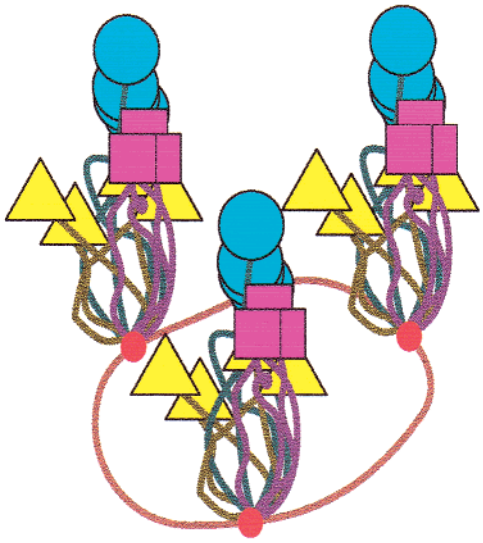

Figure 2. A supermolecule from the VCL. A supermolecule is created by connecting, at each of the attachment points on the scaffold, all possible rotamers of all substituents concurrently.

require information only on pairs of substituents, followed by filters that depend on information on triplets, and then higher order *n*-plets if necessary. Following the application of all filters, the scaffold with the few remaining substituents represents a relatively small set of physically realizable molecules that contain the pharmacophore.

The filtration concept may be translated to practice in numerous manners. An algorithmic description of the process used herein is described in the Appendix. In the following we present a conceptual description of the algorithm as applied to a cyclic peptide library. We note that the algorithm is general, and can be applied to any set of molecules that can be conceptually divided into scaffold and substituents; peptide libraries are chosen since they can provide a very large virtual combinatorial library, which effectively demonstrates the strength of the algorithm. For example, consider a library formed by 300 different cyclic peptide scaffolds[19] with 6 attachment points for each, and assume that 15 possible side chains (substituents) contain atoms or chemical groups that can be identified with the descriptors of the desired pharmacophore (in the following we will refer to these as the descriptors of the rotamer). If we consider an average of 5 conformations for each scaffold and 20 rotamers for each side chain, this library spans approximately $(300 \times 5)(15 \times 20)^6 \approx 1.1 \times 10^{18}$ possibilities that need to be checked for agreement with the pharmacophore. This library is represented in the algorithm by just 1500 supermolecules, corresponding to the different scaffold conformations. Note that in this particular example the 10^{18} possibilities, each explicitly treated by the algorithm, contain multiple conformers of $300 \times 15^6 \approx 3.4 \times 10^9$ different molecules. In general these could also be 10^{18} different molecules, since the algorithm treats each of the 300 rotamers attached at each position as distinct substituents.

I. Preparatory Steps. Initially, the pharmacophore is translated into the set of distances between all possible pairs of descriptors. Then, the supermolecules are generated by attaching, at each attachment point of each scaffold, all possible substituent rotamers. For each of the potential descriptors on the supermolecules, a descriptor type (in the example of Figure 1 the descriptor types are "blue", "yellow", or "magenta") and possible descriptor identities are assigned (for example, in the molecule of Figure 1 each chemical group of type yellow is assigned identities D, E, and F). In general, chemical groups on both the scaffold and the substituents can be descriptors, and are treated in separate manners by the algorithm.[20] Yet for clarity we will assume in the following that all descriptors are on the substituents (cf. the Appendix for a more general procedure). At this stage all rotamers of all substituents that have no descriptor type assignment associated with them or have a van der Waals clash with the scaffold are eliminated.

Then potential pharmacophore interactions are recorded for each supermolecule. A pair of descriptors on a supermolecule are defined as "interacting" if (a) the distance between them is equal to a corresponding distance in the pharmacophore, within an allowed tolerance, (b) there is no van der Waals clash between the corresponding rotamers, and (c) the rotamers are attached to different attachment points.[21] Then we scan all possible pairs of rotamer descriptors and record the ones that interact. The memory requirements of the algorithm are much smaller than implied by the huge number of possible pairs of rotamers, since the number of interacting ones is relatively small.

II. Filters. The filters are applied to each of the rotamers on the supermolecules in an iterative manner: Each rotamer is checked and eliminated if it does not meet the filter condition, and after scanning of all rotamers, the ones that are left are checked again. This is because for each rotamer one or more of its interacting partners may have been eliminated in the scan. Thus, the filtering is reiterated as long as rotamers are being eliminated. Additionally, supermolecules that have been depleted of all rotamers are eliminated, since their scaffolds cannot possibly display the desired pharmacophore.

The first filter scans all rotamer pairs regarding only their descriptor *type* assignment (blue, yellow, or magenta in the example of Figure 1). At this level two rotamers are considered to interact if any one of the pairs of descriptor identities associated with them interact. Each rotamer's list of interactions is checked, and rotamers which do not contain in their list at least one interaction for each of the descriptor type interactions in the pharmacophore are eliminated. In the example of Figure 1 each of the descriptors of type yellow is kept if it interacts with at least one descriptor of type blue, at least two descriptors of type magenta, and at least two descriptors of type yellow. Additionally the filter checks that these interactions are formed by rotamers on at least as many attachment points as there are pharmacophore descriptors (five in the example of Figure 1). This requirement reflects the fact that two or more substituents on the same attachment point cannot represent a physically realizable molecule.[21] This filter is simple, yet very effective, usually eliminating a large portion of the rotamers without considering the actual identities of the descriptors on them (for example, D, E, or F for the type yellow in Figure 1).

The subsequent filter refines the actual *identity* of the descriptors (A, B, C, D, E, and F in the example of Figure 1). This is done by checking, for each rotamer, all of its possible descriptor identities separately. An identity assignment is kept if it is compatible with at least one interaction for each of the relevant interactions defined in the pharmacophore. In the example of Figure 1 a descriptor that is assigned type yellow is now checked to determine whether it can be assigned the identity D, E, or F. Under the assignment D the rotamer must interact with rotamers containing assignments A, B, C, E, and F, under the assignment E it must interact with rotamers containing assignments A, B, C, D, and F, etc. As above, the filter also checks that these interactions are formed by rotamers on at least as many attachment points as there are pharmacophore descriptors. Descriptor assignments that do not meet the conditions of the filter are eliminated. Subsequently, rotamers that are left with no possible assignment are eliminated. We note that following the application of this filter a chemical group or atom can still have more than one potential descriptor assigned to it.

As stated above, the filters initially screen interacting pairs of descriptors. Thus, the remaining supermolecules may contain a rotamer that interacts with two rotamers that do not interact between themselves, a condition incompatible with the pharmacophore. Therefore, the subsequent filters are direct extensions of the previous ones, yet consider interacting triplets, quadruplets, and so on. An interacting

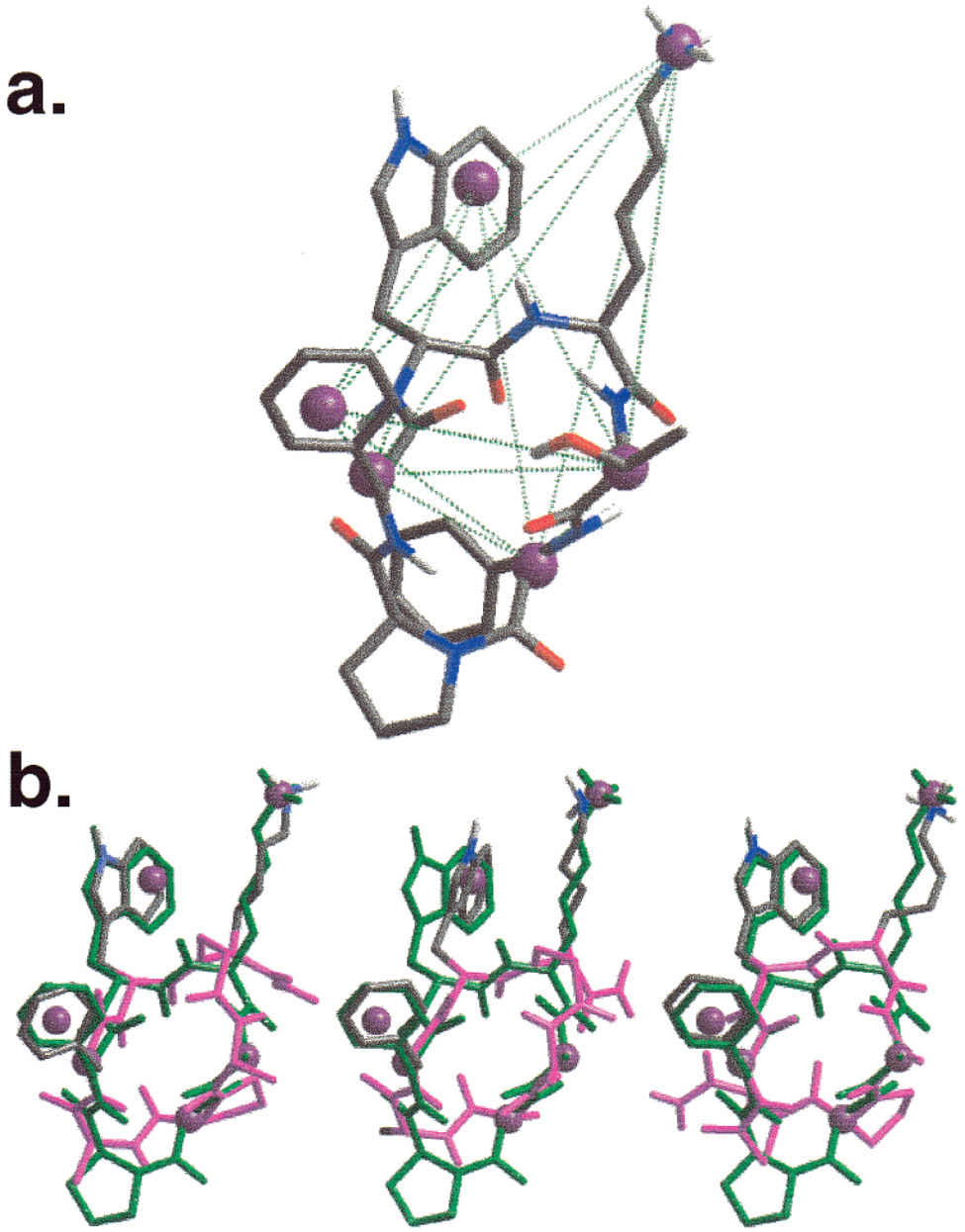

Figure 3. Novel peptidomimetic analogues containing a somatostatin pharmacophore. (a) Pharmacophore used to screen the VCL, as depicted on a well-studied analogue—the Veber compound.[23] The pharmacophore, depicted by magenta circles, consists of two centers of aromatic rings, a charged NH_3 group, and three C_α carbons of the peptidic backbone. (b) Three examples of novel backbone cyclic peptidomimetic analogues that can present the somatostatin pharmacophore, superposed on the Veber compound. The Veber compound is depicted in green, the descriptors of the pharmacophore are marked by the magenta circles, and the analogues identified by the algorithm are depicted by a magenta backbone and gray side chains with nitrogen atoms highlighted in blue. The descriptors of the pharmacophore in each pair overlap well (root-mean-square distance, RMSD, of 0.4 Å), despite the overall different conformations of the scaffold backbones.

triplet, and analogously for higher order *n*-plets, is formed if all possible pairs formed within it are interacting pairs. The filters applied to triplets and higher order *n*-plets are more computationally demanding than the corresponding ones applied to pairs, yet have a higher chance of detecting incompatible substituents. Since these filters are applied after the majority of substituents have been eliminated, the additional computational requirements are generally small. The level of filtration is increased stepwise: each rotamer is checked for triplets of descriptor types, followed by a check for triplets of descriptor identities, then quadruplets of descriptor types, and so on. The filtration stops either when there are no rotamers left or when *n*-plets are of the size of the number of pharmacophore descriptors.

III. Postprocessing. Following filtration the supermolecules contain a small set of substituents, and represent few physically realizable molecules. From each supermolecule all possible pharmacophore-containing molecules are extracted. The final outcome is a set of molecules that contain the pharmacophore within the tolerance specified in the input.

The reduction of the pharmacophore configuration into a set of distances results in the elimination of its chirality. The advantage of this approach is that only half of the scaffold library needs to be scanned, since both enantiomers are checked simultaneously. However, this requires that additional postprocessing be performed, checking for each of the resulting molecules which enantiomer has the correct pharmacophore configuration.

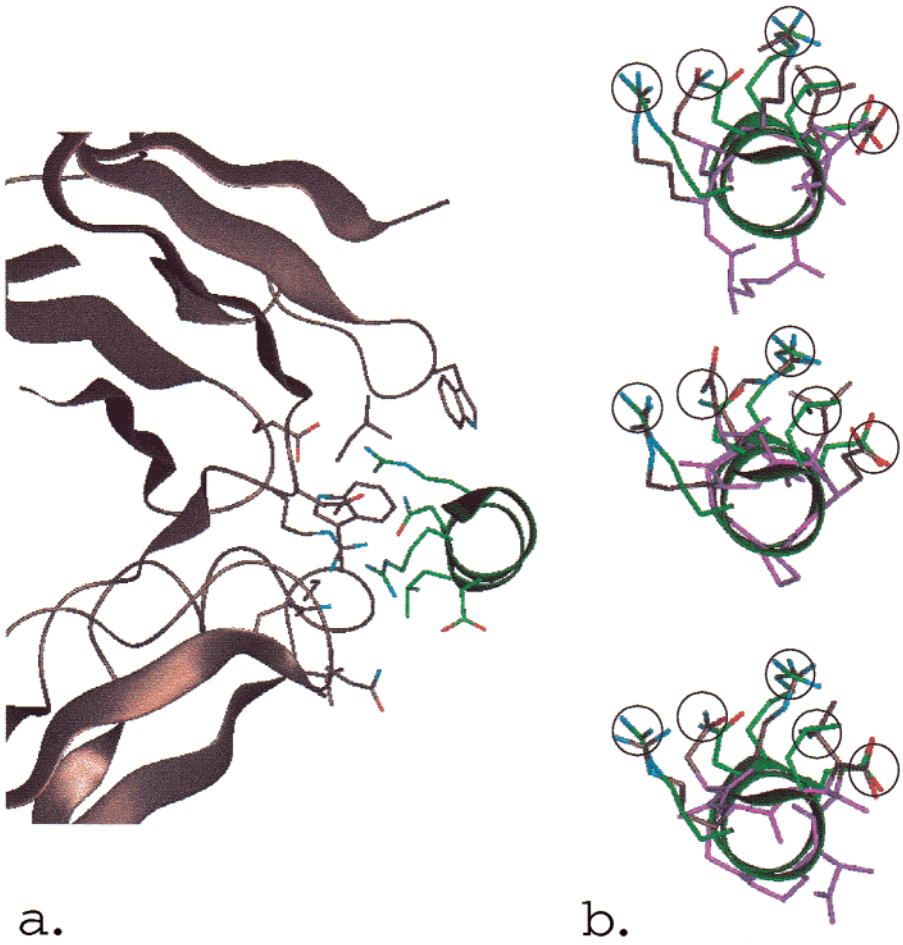

Figure 4. Novel proteinomimetic analogues mimicking the side chain configuration of residues protruding from a single face of a helix. (a) Helical fragment (green) contacting the receptor (silver), with the side chains that interact from both the helix and the receptor highlighted. (b) Three potential helix-mimetic backbone cyclic peptides. The original helical fragment, consisting of nine residues, is depicted in green, with side chain nitrogens and oxygens in blue and red, respectively. The six residue scaffolds of the mimetic analogues are purple, with gray side chains and side chain nitrogen and oxygen atoms as for the helix. The chemical groups of the pharmacophore used in the search are highlighted by circles. These groups overlap between the helix and the mimetics (RMSD of 0.3, 0.6, and 0.4 Å for the top, center, and bottom analogues, respectively), despite the clearly nonhelical nature of the scaffolds.

EXAMPLES

I. Somatostatin Pharmacophore. Figure 3a depicts a six-point pharmacophore for analogues of the hormone somatostatin,[22] superimposed on the NMR structure of an extensively studied somatostatin analogue, L363,301 (also termed the Veber compound).[23] Note that this pharmacophore has descriptors that can be assigned only to the scaffold and descriptors that can be assigned only to the substituents. As a first test of the algorithm we constructed a single supermolecule on the scaffold (backbone) of the Veber compound by attaching rotamer libraries[24] of Phe, Trp, and Lys side chains (the total size of the library is 143 rotamers, containing each of the rotamers of ref 24 enriched by adding and subtracting 1 standard deviation to/from rotamer χ values) to each of the 5 possible attachment points (all residues excluding the proline). The resulting supermolecule represents ca. 6×10^{10} possibilities. The algorithm ran on a Pentium computer (PII-400MHz) and, using a tolerance of 10%, filtered out all but the original side chain arrangement of the Veber compound within 3 s.

The above pharmacophore was then used to screen a VCL containing 6022 backbone cyclic[17] scaffolds of lengths of 3−6 residues. Each scaffold is represented by an average of 5 conformers,[25] and to each of the attachment points were added 143 rotamers. This VCL contained 3.3×10^6 different molecules represented by 1.4×10^{16} entries (numerous conformations for each molecule). Note that in practice the library spanned a much larger space, since substituents that did not include any descriptor of the pharmacophore were removed in a preparatory step. This library was screened in ca. 30 min on a cluster of 19 dual-Pentium computers using the same tolerance as above, and identified 247 novel molecules that can potentially present the somatostatin pharmacophore (Figure 3b).

II. Helix Mimetics. Next we tested the algorithm on a hypothetical pharmacophore that is defined by amino acid residues positioned on a single face of a helix in a protein. Examination of the crystal structure of the complex of growth hormone with its receptor (PDB code 3hhr) revealed helices of the hormone that contact the receptor. We arbitrarily defined a pharmacophore of five chemical groups from one of these helices (Figure 4a). The above library of scaffolds and a library of 8 substituents (298 rotamers) were combined to a VCL representing 1.1×10^{18} entities (conformations of 1.3×10^8 different molecules), which was screened in ca.

35 min. Examples of the resulting helix-mimetic peptides are depicted in Figure 4b.

DISCUSSION

We have presented an algorithm capable of scanning huge numbers of molecular entities in a highly efficient manner. The sets of conformers of all scaffolds and substituents are precalculated once. For each specific application only those scaffolds and substituents that contain the desired chemical moieties are used for the construction of the supermolecules. The efficiency of the algorithm stems from the fact that huge libraries are screened without actually generating the huge number of possible individual molecules, and from the set of filters which circumvent the need for checking full pharmacophores on large numbers of combinations. We have demonstrated that libraries on the order of 10^{18} compounds (either 10^{18} different molecules or numerous conformations of a smaller number of compounds) can readily be screened within minutes using reasonable computer resources. This number scales linearly with the number of scaffolds, while increasing the number of substituents increases the memory requirements of the algorithm and the time required for the generation of the VCL, but only slightly affects the time requirement of the filtration steps. Thus, the algorithm can deal with libraries larger than 10^{18}, yet we have not determined its upper bound. It is noteworthy to point out that libraries commonly employed in drug design are comprised of scaffolds with 3−4 attachment points, fewer than the examples herein, yet up to thousands of possible substituents, much larger than the hundreds in our examples. The supermolecules representing such libraries can be screened by the algorithm with the same level of efficiency as demonstrated for the peptidic libraries, since the major effect of increasing the number of substituents is on memory, and not speed.

The algorithm requires a pharmacophore defined by a spatial configuration of a set of chemical groups, and thus can be applied in the context of a more general structural motif, for example, a particular configuration of chemical groups associated with a metal binding site or an active site of a protein. Additionally, the user must also specify a tolerance for matching between interdescriptor distances in the pharmacophore and the corresponding distances in the VCL.

The tolerance is sometimes intrinsic to the pharmacophore, associated with the uncertainties in the definition of the spatial configuration of its descriptors. A tolerance is also necessary because the algorithm uses a discrete representation of the conformational space for both scaffolds and substituents. The actual value of the tolerance is problem dependent, and should consider the relative size of the pharmacophore and the density of the sampling of the conformational space of the scaffolds and substituents. An overly large tolerance compared to the discretization of the conformational space results in too many hits, many of them only approximately fitting the pharmacophore. An overly small tolerance (or an exceedingly coarse discretization) may miss possible solutions. In the examples that we present below, the tolerance was determined by trial and error, checking both the number of hits and their quality at each run.

APPENDIX

I. Input.

1. Library of 3D conformations of scaffolds.

2. Library of 3D conformations of substituents (and rules defining which substituent can be attached to which attachment point).

3. Pharmacophore.

3.1. Define (a) SCAF-DSC, number of descriptors that can be only on the scaffold (this parameter defines an on-scaffold subpharmacophore), (b) SUBS-DSC, number of descriptors that can be only on the substituents, (c) TOT-TYPES, number of different descriptor types on the pharmacophore, and (d) TOT-DSC, number of descriptors on the pharmacophore.

II. Preparatory steps.

1. Translate pharmacophore into the set of distances between all possible pairs of descriptors.

2. Construct VCL by attaching to each of the attachment points on each scaffold all possible rotamers of all possible substituents.

3. Eliminate rotamers that clash with the scaffold.

4. For each potential descriptor on both the scaffold and substituents of each supermolecule, assign a descriptor type and the corresponding descriptor identities.

5. Eliminate all rotamers that have no descriptor type assignment.

6. Record all interacting descriptor pairs (see the text for definition). In the current application these interactions are recorded in a matrix.

III. Filtration.

1. If an on-scaffold subpharmacophore exists, check whether the descriptor assignments on the scaffold satisfy its distance requirements in at least one combination. If such combinations exist, descriptor assignments that do not contribute to any of them are eliminated. If no such combination exists, the supermolecule is eliminated from the VCL.

2. Perform pairwise filtration.

2.1. Filter based on descriptor types.

2.1.1. Scaffold descriptors: For each on-scaffold descriptor count the total number of interacting pairs, considering only the respective descriptor types. For each descriptor type interaction, record the descriptor types involved, and whether the interaction is with an on-scaffold descriptor or with a substituent. Each on-scaffold descriptor type needs to form interacting pairs involving at least one of each pharmacophore descriptor type (a total of TOT-TYPES, including itself). The interactions must involve at least TOT-DSC different attachment positions and on-scaffold descriptor positions (including itself), at least SUBS-DSC different attachment positions, and at least SCAF-DSC different on-scaffold descriptor positions (including itself). If these conditions are not met, the assignments associated with the on-scaffold descriptor type are removed. This step is repeated iteratively until no more on-scaffold assignments are removed.

2.1.2. Substituents: For each substituent count the total number of interacting pairs, considering only the respective descriptor types. For each descriptor type interaction, record the descriptor types involved, and whether the interaction is with an on-scaffold descriptor or with a substituent. Each substituent needs to form interacting pairs

involving at least one of each pharmacophore descriptor type considering both on-scaffold descriptors and substituents (a total of TOT-TYPES, including itself). The interactions must involve substituents attached to at least TOT-DSC different attachment points (including itself) and on-scaffold descriptor positions, substituents attached to at least SUBS-DSC different attachment points (including itself), and at least SCAF-DSC different on-scaffold descriptor positions. If these conditions are not met, the substituent is removed. This step is repeated iteratively until no more substituents are removed.

2.1.3. Steps 2.1.1 and 2.1.2 are repeated iteratively until no more on-scaffold assignments are removed and no more substituents are removed. If on-scaffold assignments are removed, a full scaffold subpharmacophore systematic check is performed (step 1) before a new iteration is started.

2.2. Filter based on descriptor identities.

2.2.1. Scaffold descriptors: For each on-scaffold descriptor count the total number of interacting pairs, considering the respective descriptor identities. For each descriptor interaction, record the descriptors involved, and whether the interaction is with an on-scaffold descriptor or with a substituent. Each descriptor needs to form interacting pairs involving at least SUBS-DSC different descriptors from the substituents, at least SCAF-DSC different descriptors from the scaffold (including itself), and at least a total of TOT-DSC different descriptors, considering both scaffold and substituents (including itself). The interactions must involve at least TOT-DSC different attachment positions and on-scaffold descriptor positions (including itself), at least SUBS-DSC different attachment positions, and at least SCAF-DSC different scaffold descriptor positions (including itself). If these conditions are not met, the descriptor assignment is removed. This step is repeated iteratively until no more assignments are removed.

2.2.2. Substituents: For each descriptor identity assigned to a substituent count the total number of interacting pairs it forms. For each descriptor interaction, record the descriptors involved, and whether the interaction is with an on-scaffold descriptor or with a substituent. Each descriptor needs to form interacting pairs involving at least SUBS-DSC different descriptors from the substituents (including itself), at least SCAF-DSC different descriptors from the scaffold, and a total of at least TOT-DSC different descriptors, considering both scaffold and substituents (including itself). The interactions must involve at least TOT-DSC different attachment positions and on-scaffold descriptor positions (including itself), at least SUBS-DSC different attachment positions (including itself), and at least SCAF-DSC different scaffold descriptor positions. If these conditions are not met, the descriptor assignment is removed. This step is repeated iteratively until no more assignments are removed. A substituent with no descriptor assignments left is removed.

2.2.3. Steps 2.2.1 and 2.2.2 are repeated iteratively until no more on-scaffold assignments are removed and no more on-substituent assignments are removed. If on-scaffold assignments are removed, a full scaffold subpharmacophore systematic check (step 1) is performed before a new iteration is started.

3. Perform filtration on triplets. Repeat step 2 for triplets.

4. Perform filtration on higher order n-plets ($n = 4, 5, ...,$ size of the pharmacophore).

IV. Postprocessing.

1. Extract from the resulting supermolecules all possible pharmacophore-containing molecules.

2. Check the chirality of the pharmacophore, and generate an enantiomer if necessary.

ACKNOWLEDGMENT

We acknowledge stimulating discussions and fruitful comments of Drs. C. Keasar and H. Senderowitz. We thank C. Keasar for providing the programs for the generation of the scaffold libraries.

REFERENCES AND NOTES

(1) Rademann, J.; Jung, G. Integrating Combinatorial Synthesis and Bioassays. *Science* **2000**, *287*, 1947-1948.
(2) Wentworth, P., Jr.; Janda, K. D. Generating and analyzing combinatorial chemistry libraries. *Curr. Opin. Biotechnol.* **1998**, *9*, 109−115.
(3) Hann, M.; Green, R. Chemoinformatics—a new name for an old problem? *Curr. Opin. Chem. Biol.* **1999**, *3*, 379−383.
(4) Walters, W. P.; Stahl, M. T.; Murcko, M. A. Virtual Screening—an Overview. *Drug Discovery Today* **1998**, *3*, 160−178.
(5) Cramer, R. D.; Patterson, D. E.; Clark, R. D.; Soltanshahi, F.; Lawless, M. S. Virtual compound libraries: A new approach to decision making in molecular discovery research. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 1010−1023.
(6) Martin, E. J.; Blaney, J. M.; Siani, M. A.; Spellmeyer, D. C.; Wong, A. K.; Moos, W. H. Measuring diversity: experimental design of combinatorial libraries for drug discovery. *J Med. Chem.* **1995**, *38*, 1431−1436.
(7) Andrews, K. A.; Cramer, R. D. Toward general methods of targeted library design: Topomer shape similarity searching with diverse structures as queries. *J. Med. Chem.* **2000**, *43*, 1723−1740.
(8) Cramer, R. D.; Clark, R. D.; Patterson, D. E.; Ferguson, A. L. Bioisosterism as a molecular diversity descriptor: Steric fields of single "topomeric" conformers. *J. Med. Chem.* **1996**, *39*, 3060−3069.
(9) Ghose, A. K.; Viswanadhan, V. N.; Wendoloski, J. J. A Knowledge-based approach in designing combinatorial or medicinal chemistry libraries for drug discovery. 1. A qualitative and quantitative characterization of known drug databases. *J. Comb. Chem.* **1999**, *1*, 55−68.
(10) Clark, D. E.; Firth, M. A.; Murray, C. W. MOLMAKER: De novo generation of 3D databases for use in drug design. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 137−145.
(11) Webb, T. R.; Melman, N.; Lvovskiy, D.; Ji, X.-D.; Jacobson, K. A. The utilization of a united pharmacophore query in the discovery of new antagonists of the adenosine receptor family. *Bioorg. Med. Chem. Lett.* **2000**, *10*, 31−34.
(12) Murcko, M. A. Recent advances in ligand design methods. In *Reviews of Computational Chemistry*; Lipkowitz, K. B., Boyd, D. B., Eds.; Wiley-VCH, John Wiley and Sons: New York, 1997; Vol. 11, pp 1−67.
(13) Van Drie, J. H. Strategies for the determination of pharmacophoric 3D database queries. *J. Comput.-Aided Mol. Des.* **1997**, *11*, 39−52.
(14) Martin, Y. C.; Bures M. G.; Danaher, E. A.; DeLazzer J.; Lico I.; Pavlik, P. A. A fast new approach to pharmacophore mapping and its application to dopaminergic and benzodiazepine agonists. *J. Comput.-Aided Mol. Des.* **1993**, *7* , 83−102.
(15) Good A. C.; Mason J. S. Three-dimensional structure database searches. In *Reviews of Computational Chemistry*; Lipkowitz, K. B., Boyd, D. B., Eds.; Wiley-VCH, John Wiley and Sons: New York, 1997; Vol. 7, pp 67−117.
(16) Martin, Y. C. 3D Database Searching in Drug Design. *J. Med. Chem.* **1992**, *35*, 2145−2154.
(17) Gilon, H.; Halle, D.; Chorev, M.; Selinger, Z.; Byk, G. Backbone cyclization: a new method for conferring conformational constraint on peptides. *Biopolymers* **1991**, *31*, 745−750.
(18) Pharmacophores having vector descriptors are not discussed in this paper, yet a straightforward extension to the filters described herein can readily handle them. This extension should include the addition of filter conditions that, for example, consider two bending angles that uniquely determine the direction of the vector descriptors in space.
(19) Conceptually, a library of cyclic peptide scaffolds is composed solely of the peptidic backbone. In practice, such a library is composed of small cyclic polyalanines since lack of side chains would result in different backbone conformations (polyglycine is much more flexible than polyalanine). The C_α carbons of the alanine residues are defined as attachment points to which the substituents are attached along the $C_\alpha-C_\beta$ bond. A set of different scaffolds is obtained by interchanging

alanine positions with amino acids that affect the conformation of the backbone, for example, proline, glycine, N-alkylated alanine, N-alkylated glycine, or the corresponding D diastereomer of any of these residues. Additional diversity of scaffolds may be obtained by including molecules cyclized by different cyclization modes, such as head-to-tail cyclization, disulfide bonds, or backbone cyclization (see ref 17). The size of these scaffolds must be small enough so that the conformational space of each scaffold can be well represented by a relatively small set of conformers.

(20) In general there are descriptors that can reside only on the scaffold (e.g., a C_α atom in the example herein), descriptors that can reside only on substituents (e.g., an aromatic group), and descriptors that can reside on either the scaffold or the substituents (e.g., an amide group). If there is a subpharmacophore on the scaffold, it can be used to eliminate scaffolds from the VCL prior to the construction of their corresponding supermolecule. Additionally, descriptors on the scaffolds and those on the substituents require different handling by the algorithm: rotamers are eliminated due to lack of descriptor assignments, while in the case of scaffolds the filters merely eliminate descriptor assignments rather than eliminate the scaffold.

(21) In the current application we limit each substituent to contribute only one descriptor to the pharmacophore. During the filtration process there may be more than one assignment to each substituent, but for the resulting molecules each substituent contributes only one descriptor. The extension of the algorithm to override this limitation is straightforward, and only requires testing each rotamer for interacting pairs between the nonoverlapping descriptors assigned to them.

(22) Golender, L.; Rosenfeld, R.; Vorpagel, E. Small cyclic peptide SAR study using APEX-3D System: somatostatin receptor type-2 (sstr2) specific pharmacophores. In *Molecular modeling and prediction of bioactivity;* Gundertofte, K., Jorgensen, F. S., Eds.; Kluwer Academic/Plenum Publishers: New York, 1999; pp 336−337.

(23) Huang, Z.; He, Y-B.; Raynor, K.; Tallent, M.; Reisine, T.; Goodman, M. Main chain and side chain chiral methylated somatostatin analogs: syntheses and conformational analyses. *J. Am. Chem. Soc.* **1992**, *114*, 9390−9401.

(24) Ponder, J. W.; Richards, F. M. Tertiary templates for proteins. *J. Mol. Biol.* **1987**, *193*, 775−791.

(25) The backbone cyclic scaffold library was constructed by a modified MCMM conformational search protocol (see ref 26) using a modified AMBER/OPLS potential (see ref 27). The lowest energy conformations obtained were clustered, and up to five lowest energy conformations of each scaffold were used in the library provided that the energy difference between each conformation and the lowest energy one is below 4 kcal/mol.

(26) Chang, G.; Guida, W. C.; Still, W. C. An internal coordinate Monte Carlo method for searching conformational space. *J. Am. Chem. Soc.* **1989**, *111*, 4379−4386.

(27) Keasar, C.; Rosenfeld, R. Empirical modifications to the Amber/OPLS potential for predicting the solution conformations of cyclic peptides by vacuum calculations. *Folding Des.* **1998**, *3*, 379−388.

CI000463O