CHORTLES: A Method for Representing Oligomeric and Template-Based Mixtures

Michael A. Siani,*,†,§ David Weininger,‡ Craig A. James,‡ and Jeffrey M. Blaney†

Chiron Corporation, 4560 Horton Street, Emeryville, California 94608,
and Daylight Chemical Information Systems, 18500 Von Karman #450, Irvine, California 92715

Received May 30, 1995®

Screening mixtures of synthetic oligomers or fixed templates (e.g., rings) with varying substituents is increasingly the focus of drug discovery programs. CHORTLES is designed and implemented to facilitate representation, storage, and searching of oligomeric and template-based mixtures of any size. Building upon the CHUCKLES method of representing oligomers as both monomer-based sequences and all-atom structures, CHORTLES compactly represents a mixture without explicitly enumerating individual molecules. This method lends itself to a hierarchy relating mixtures to submixtures and individual compounds, as one finds when deconvoluting mixtures in drug lead discovery programs. In addition, we describe two methods of searching mixtures at the monomer level. We also present a simple pictorial representation for describing all components in a mixture, which becomes essential as the list of monomer names is expanded beyond common names (e.g., amino acids).

INTRODUCTION

The proliferation of libraries of natural (peptide)[1−7] and un-natural oligomers,[8−11] along with cyclic template libraries,[12] necessitates a new way to store and search vast compound libraries. Peptides, rapidly synthesized by standard techniques, are increasingly a part of chemical libraries. Peptoids[8] (N-substituted glycines (NSGs)) and carbamates[10] are also appropriate for generating molecular diversity. With automation of peptide and peptoid synthesis, one is able to make both pure compounds and large mixtures. Pure compounds lend themselves to straightforward screening against various receptors and enzymes. In a previous paper, we presented CHUCKLES,[13] a method for representing peptide and peptoid sequences on both the monomer and atomic levels. Mixtures permit rapid screening of large numbers of compounds at once. Zuckermann et al.[14] have shown that initial screening of complex mixtures and subsequent deconvolution and assaying of more specific compounds is a fast and powerful means of drug lead discovery.

Representation of mixtures presents a more complex problem in database storage and searching than individual molecules. A mixture can consist of a set of oligomers which differ at certain sequence positions. For example, the mixture sequence Ala[Arg;Lys;His]Thr contains three peptides: AlaArgThr, AlaLysThr, and AlaHisThr. The peptides in a mixture contain fixed positions (e.g., positions one and three) which contain a single monomer and mixture positions (e.g., position two) which contain multiple monomers.

Template-based mixtures consist of a substrate (e.g., a ring) with varied substituents at multiple attachment points. These templates may be used in mixture synthesis by competitive coupling of mixed substrates[15] or by controlled deprotection and coupling at specific sites.[16]

Our previous work with peptide/peptoid oligomers stressed the ability to convert between sequence and atom-level

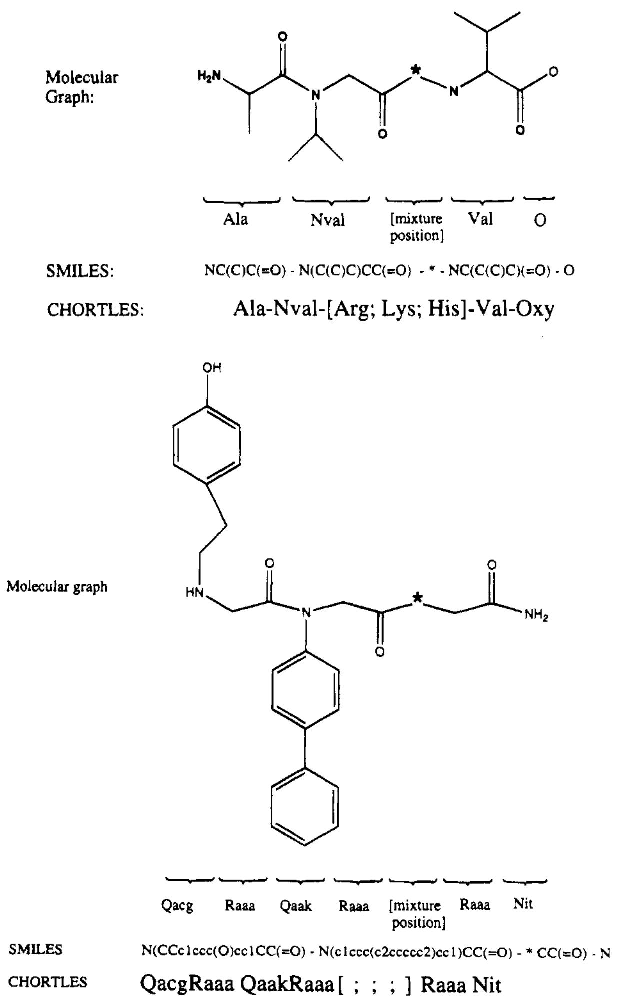

Molecular Graph:
H₂N
O
N
*
N
O
O
O
Ala
Nval
[mixture position]
Val
O
SMILES:
NC(C)C(=O) - N(C(C)C)CC(=O) - * - NC(C(C)C)(=O) - O
CHORTLES:
Ala-Nval-[Arg; Lys; His]-Val-Oxy
OH
Molecular graph
HN
O
N
O
*
O
NH₂
Qacg
Raaa
Qaak
Raaa
[mixture position]
Raaa
Nit
SMILES
N(CCc1ccc(O)cc1CC(=O) - N(c1ccc(c2ccccc2)cc1)CC(=O) - * CC(=O) - N
CHORTLES
QacgRaaa QaakRaaa[; ; ;] Raaa Nit

Figure 1. (a) Simple oligomeric mixture: molecular graph with "*" corresponding to mixture positions; SMILES representation of graph, and CHORTLES mixture sequence. Mixture sequence has fixed positions 1, 2, and 4, and mixture position 3 with basis set of Arg, Lys, and His. (b) The submonomer approach to peptoid synthesis lends itself to representing each traditional monomer as two submonomers, the amine (Qaaa, Qaab, Qaac, ...) and the backbone (Raaa, Raab, Raac). In position 1, Qacg represents the tyramine (side chain), and Raaa represents the peptoid backbone atoms CC(=O). In position 2, Qaak represents the biphenylamine, and Raaa represents the standard peptoid backbone. Position 3 is a mixture position with four monomers in the basis set.

[†] Chiron Corporation.

[‡] Daylight Chemical Information Systems.

[§] Permanent address: Gryphon Sciences, 250 E. Grand Avenue, Suite 90, South San Francisco, CA 94080. Phone: 415-952-7714. Fax: 415-952-3055. E-mail: siani@gryphonsci.com.

[®] Abstract published in *Advance ACS Abstracts,* August 15, 1995.

0095-2338/95/1635-1026$09.00/0

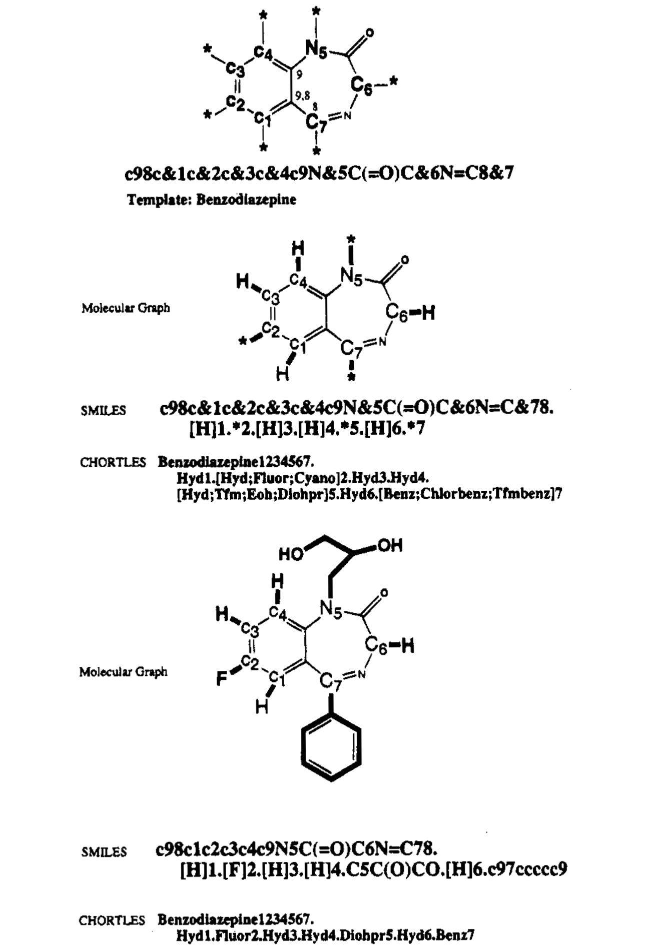

C4
N5
C3
C2
C1
C7
C6
9
9,8
8
c98c&1c&2c&3c&4c9N&5C(=O)C&6N=C8&7
Template: Benzodiazepine
Molecular Graph
SMILES c98c&1c&2c&3c&4c9N&5C(=O)C&6N=C&78.
[H]1.*2.[H]3.[H]4.*5.[H]6.*7
CHORTLES Benzodiazepine1234567.
Hyd1.[Hyd;Fluor;Cyano]2.Hyd3.Hyd4.
[Hyd;Tfm;Eoh;Diohpr]5.Hyd6.[Benz;Chlorbenz;Tfmbenz]7
HO
OH
F
Molecular Graph
SMILES c98c1c2c3c4c9N5C(=O)C6N=C78.
[H]1.[F]2.[H]3.[H]4.C5C(O)CO.[H]6.c97ccccc9
CHORTLES Benzodiazepine1234567.
Hyd1.Fluor2.Hyd3.Hyd4.Diohpr5.Hyd6.Benz7

Figure 2. (a) Benzodiazepine template with potential substitution (mixture) points; substituents attach at the atoms with indices which are not satisfied. The indices 8 and 9 in the SMILES indicate bonds which cyclize within the template. Indices 1−7 are ordered connection points for mixture substitution; these unsatisfied bonds are preceeded by "&" to indicate they correspond to other indicies in the CHORTLES. (b) The benzodiazepine template in the context of a mixture. Positions 1, 3, 4, and 6 are fixed with [H]1, a hydrogen attached to the aromatic nitrogen and carbons. The second position is a mixture (indicated by "*2") where each substituent is in the basis set of the example mixture sequence. Position 5 is a mixture (indicated by "*5") with a basis set indicated in the example mixture sequence. Position 7 is a mixture (indicated by "*7") with a basis set indicated in the example mixture sequence. The substituents are Hyd for hydrogen (H), Fluor for fluorine (F), Cyano for cyano (C#N) where the "#" indicates a triple bond, Tfm for trifluoromethyl (C(F)(F)(F)), Diohpr for dihydroxy homo propane (CC(O)CO), Benz for benzyl (c1ccccc1), Chlorbenz for chlorobenzene (c1c(Cl)cccc1), and Tfmbenz for trifluoromethylbenzene (c1c(C(F)(F)F)cccc1). (c) An actual single molecule component of the mixture shown in b.

representations in order to allow both sequence and substructure searching. However, mixture pools, consisting of different—but related—compounds do not lend themselves to a single atomic representation.

We developed CHORTLES, a method for representing mixtures as simple strings of characters. As with the previously described CHUCKLES method, this new method uses a monomer reference table which pairs a monomer name with an all-atom chemical SMILES[17] or SMARTS.[18] The CHORTLES language is a superset of CHUCKLES which allows mixtures to be specified as variability at monomer positions.

Conventional sequence searching[19,20] assumes explicit representation of individual oligomers. Performing a substructure search with mixtures using currently available packages requires enumeration and storage of all component sequences within a mixture. Such explicit representation, with fixed positions repeated in each oligomer, would require prohibitive amounts of space. In addition, searching a mixture of **n** components would require **n** sequence comparisons per query. We touch on some search strategies here; we will describe a more comprehensive approach for searching CHORTLES in a subsequent paper.

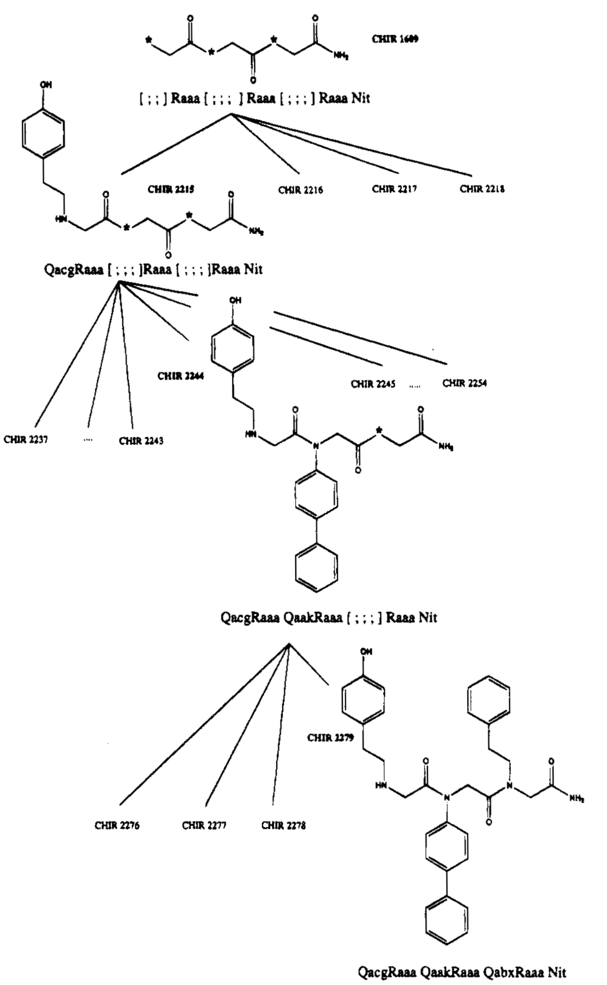
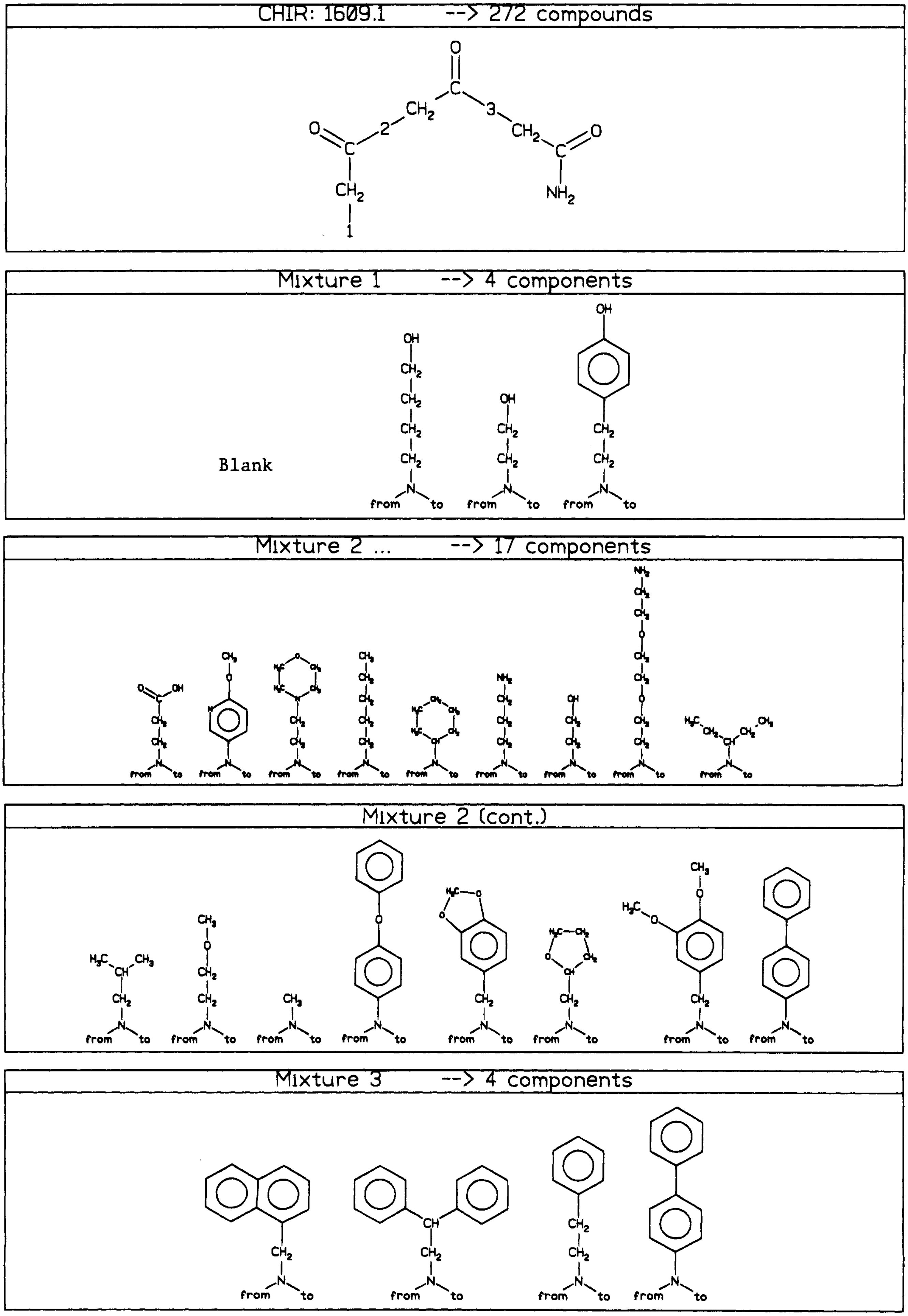
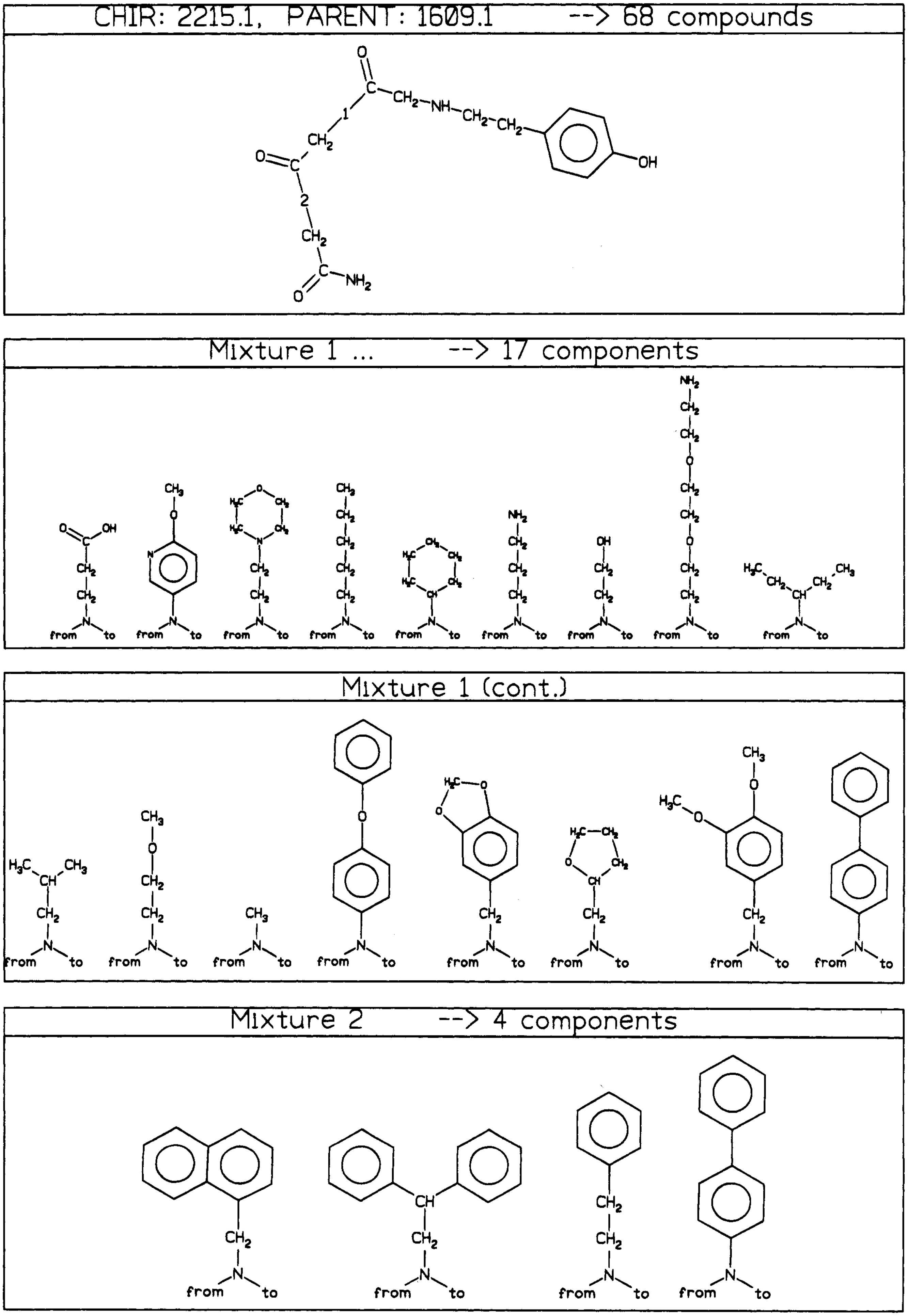

CHIR 1609
[; ;] Raaa [; ; ;] Raaa [; ; ;] Raaa Nit
CHIR 2215
CHIR 2216
CHIR 2217
CHIR 2218
QacgRaaa [; ; ;]Raaa [; ; ;]Raaa Nit
CHIR 2244
CHIR 2245 ..... CHIR 2254
CHIR 2237 .... CHIR 2243
QacgRaaa QaakRaaa [; ; ;] Raaa Nit
CHIR 2279
CHIR 2276
CHIR 2277
CHIR 2278
QacgRaaa QaakRaaa QabxRaaa Nit

Figure 3. Deconvolution family hierarchy for a$_1$-adrenergic receptor hit, CHIR 1609. The most general mixture with three mixture positions is deconvoluted into four children, each with the first position fixed. CHIR 2215, the active child, is further deconvoluted into its 17 children, each with the first two positions fixed. Of the 17 siblings, CHIR 2237−2254, CHIR 2244 is active and is further deconvoluted to four children which are individual compounds (all three positions fixed). CHIR 2279 is the compound responsible for most of the inhibition of the α_1-adrenergic receptor.

Synthetic combinatorial libraries contain hundreds to 10^6 molecules; a single phage-displayed peptide library may contain 10^{12} individual sequences. Enumeration of individual molecules in libraries would exceed the capability of even the largest database systems very rapidly. Furthermore, experiments (e.g., receptor binding assays) are performed on the mixture, not on the individual components of the mixture. Since experimental data only pertains to the mixture as a whole, we store and represent the mixture as a single entity. As a mixture is deconvoluted,[14] one or more mixture positions is expanded into separate submixtures or individual compounds; these components can then be added to the database.


CHIR: 1609.1 --> 272 compounds
1
2
3
Mixture 1 --> 4 components
Blank
Mixture 2 ... --> 17 components
Mixture 2 (cont.)
Mixture 3 --> 4 components
from
to


CHIR: 2215.1, PARENT: 1609.1 --> 68 compounds
1
2
Mixture 1 ... --> 17 components
Mixture 1 (cont.)
Mixture 2 --> 4 components
from
to

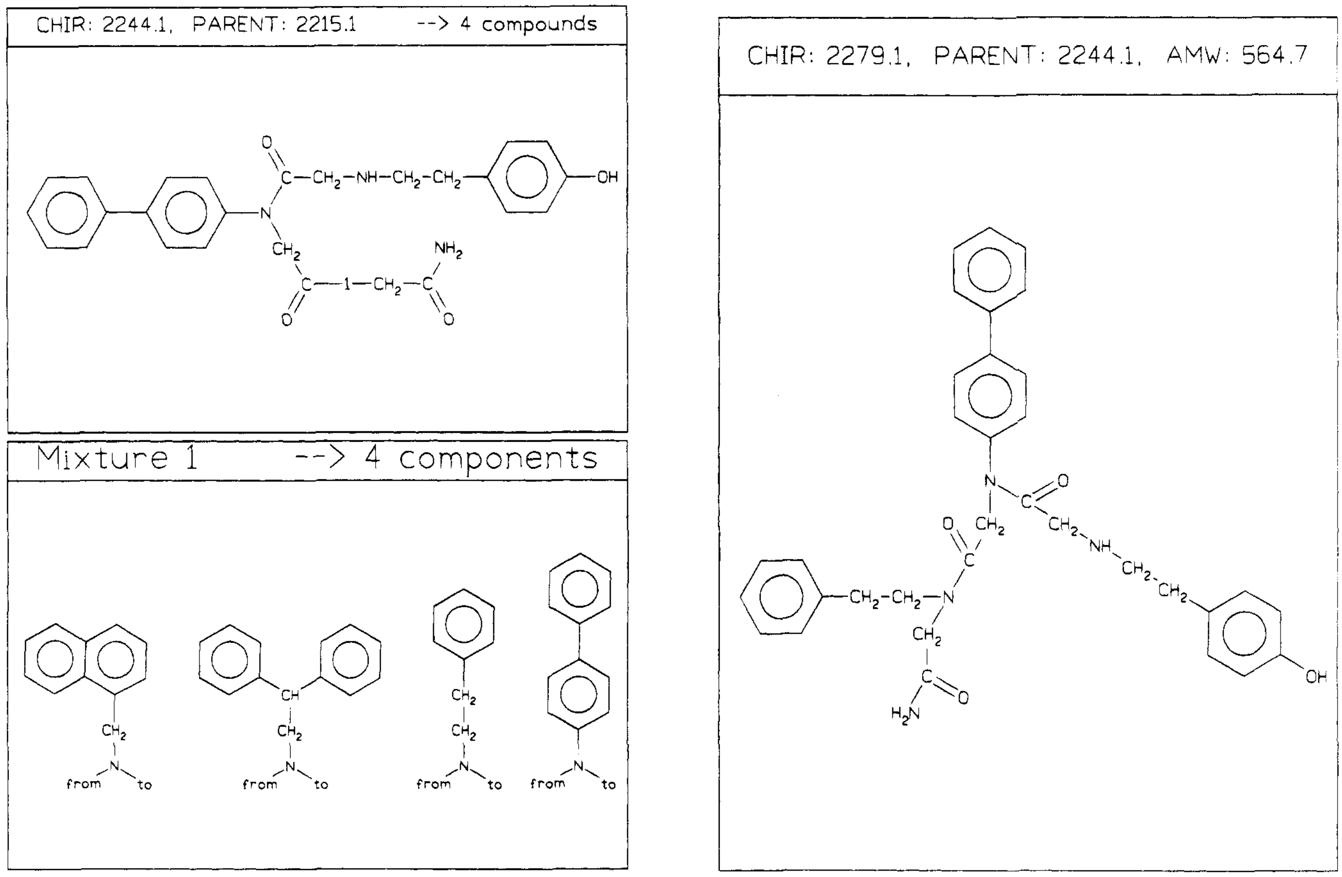

CHIR: 2244.1, PARENT: 2215.1 --> 4 compounds
Mixture 1 --> 4 components
from
to
CHIR: 2279.1, PARENT: 2244.1, AMW: 564.7

Figure 4. The combinatorial libraries which show successive deconvolution; each library is less complex than its predecessor by one position. The overall relationship between the libraries is shown in Figure 3. Full CHORTLES notation is described in the text. (a) CHIR 1609 mixture basis sets at all three positions in the trimer. Component amines are depicted in separate boxes. (b) CHIR 2215 mixture, child of 1609, represents a single level deconvolution of its parent. The first position is fixed; it has one fixed position and two mixture positions. This less complex library has three siblings; together they comprise all the components of the full library of the parent, 1609.1. (c) CHIR 2244 mixture represents a further deconvolution of the parent 2215. The first and second positions are now fixed and only one mixture position remains. (d) CHIR 2279, an individual compound, child of 2244, contains no mixture positions, and is the active member of siblings 2276–2279.

METHOD

Representation of Mixtures: Language Hierarchy. A mixture sequence is represented as a sequence of monomer names. Monomer names start with an upper-case letter followed by zero or more lower-case letters. Fixed positions are just the monomer name. Mixture positions are delineated by a pair of matching square brackets ("[" and "]"), containing the basis set monomer names separated by semicolons. A mixture sequence is canonicalized by alphabetizing all members of a basis set. This representation is extremely thrifty with space. For example, AlaNval[Arg;Lys;His]Val is a pool of three compounds: AlaNvalArgVal, AlaNvalLysVal, and AlaNvalHisVal, where positions one, two, and four are fixed positions, and position three is a mixture position. The mixture [Arg; Glu; Ile; Ser; Tyr][Arg; Glu; Ile; Ser; Tyr][Arg; Glu; Ile; Ser; Tyr] contains 125 compounds.

In our database, the mixture pool is stored under a canonical SMILES for the "parent" sequence where each mixture position is represented by the atom, "*". We treat "*" as a valid atom. Using the CHUCKLES method, the sequence is converted into an all-atom graph and subsequently into a SMILES. See Figure 1a. This SMILES is the key in our database. Associated with this identifying SMILES, we have CHORTLES mixture sequences. All CHORTLES that are subordinate to a SMILES have the same fixed positions but may have different basis sets at the mixture positions.

Because of the large number of monomers used in our libraries, we have automated naming of monomers. And with the submonomer approach to generating peptoid mixtures,[9] we separate the name for the side-chain-containing amine from the back-bone segment. We name amines Qaaa, Qaab, Qaac, ... Backbone segments are represented by Raaa, Raab, Raac, ... See Figure 1b for a sample sequence.

Template-based mixtures differ from oligomeric mixtures because they contain multiple attachment points on a single template. The attachment points are built into the template. Benzodiazepine, represented by c98c&1c&2c&3c&4c9N&5C(=O)C&6N=C8&7, has seven attachment points, represented by indices 1–7. The "&" preceding an index indicates that the bond will be satisfied outside the template monomer in the CHORTLES. The index 9 indicates a ring closure between the two atoms associated with it, likewise for index 8. See Figure 2a. In the mixture sequence, the substituents connected to the attachment points are disconnected monomers (separated by a "." character) attached via an index matching an index on the template. The mixture Benzodiazepine1234567.Hyd1.[Hyd;Fluor;Cyano]2.Hyd3.Hyd4.[Hyd;Tfm;Eoh;Diohpr]5.Hyd6.[Benz;Chlorbenz;Tfmbenz]7 has fixed positions 1, 3, 4, and 6, and mixture positions 2, 5, and 7. See Figure 2b for illustration and description of monomers at mixture positions. An individual compound for a template-based system might look like Benzodiazepine1234567.Hyd1.Fluor2.Hyd3.Hyd4.Diohpr5. Hyd6.Benz7, where the indices 1−7 refer to groups attaching at seven different positions on the benzodiazepine. See Figure 2c. Note that the template is simply another monomer and need not be more complex than any other component of the CHORTLES.

Representation of Mixtures: Combinatorial Library Relation Hierarchy. We link related mixtures in a hierarchy with more complex mixtures at the root. We call a mixture the parent when it is deconvoluted into individual compounds or less complex mixtures (with fewer mixture positions), known as the children. The children of a parent, which all contain the same mixture positions, are known as siblings. To permit traversal of such a deconvolution, we introduce the Mixture_Parent, Mixture_Child, and Mixture_Sibling pointer datatypes (simply directional pointers between related molecules/mixtures) into our database.

When a mixture is found to be a "hit" during screening against a target molecule, its components (children) are synthesized as submixtures and then screened against the same target. Figure 3 depicts the familial relationship between successive deconvolutions for an α_1-adrenergic receptor hit.[14] Consider the mixture CHIR 1609 *-Raaa-*-Raaa-*-Raaa-Nit (see Figure 4a) which consists of three mixture positions of peptoids. CHIR 1609 inhibited $[^3H]$-prazosin binding to the α_1-adrenergic receptor. To discover which component(s) of the mixture were inhibiting, the first deconvolution of the parent mixture was performed.

Deconvolution of the first position yields three child submixtures: CHIR 2215 QaaeRaaa-*-Raaa-*-Raaa-Nit, CHIR 2216 QabjRaaa-*-Raaa-*-Raaa-Nit, and CHIR 2217 QacgRaaa-*-Raaa-*-Raaa-Nit. See Figure 4b. These three submixtures are known as siblings. Subsequent synthesis and screening of these submixtures revealed that submixture CHIR 2215 inhibited prazosin binding to the α_1-adrenergic receptor.

The new parent pool, CHIR 2215, was then deconvoluted into its 17 component children (CHIR 2237−2254), with both positions one and two fixed. See Figure 4c. Assaying these pools showed CHIR 2244 (QacgRaaa-QaakRaaa-*-Raaa-Nit) to be the most active.

The final mixture position in CHIR 2244 is then deconvoluted by the synthesis of its four children, CHIR 2276 (QacgRaaa-QaakRaaa-QaakRaaa-Nit), CHIR 2277 (QacgRaaa-QaakRaaa-QaatRaaa-Nit), CHIR 2278 (QacgRaaa-QaakRaaa-QabiRaaa-Nit), and CHIR 2279 (QacgRaaa-QaakRaaa-QabxRaaa-Nit). See Figure 4d. CHIR 2279 was the most potent inhibitor of the α_1-adrenergic receptor ($K_i = 5 \pm 3$ nM).

Figures 4a−d illustrate a general, compact pictorial display of combinatorial libraries. These displays are generated by translating the CHORTLES code into its associated monomers at each position and sending the resulting SMILES file to the Daylight program, PRADO, which produces PostScript output.

Fast Mixture Sequence Searching. A mixture pool may represent any number of component oligomers or template-based molecules. One can expand the mixture sequence into a list of its components and then perform a substructure search on the molecular graphs of the individual components. However, enumerating a mixture is costly in time and space.

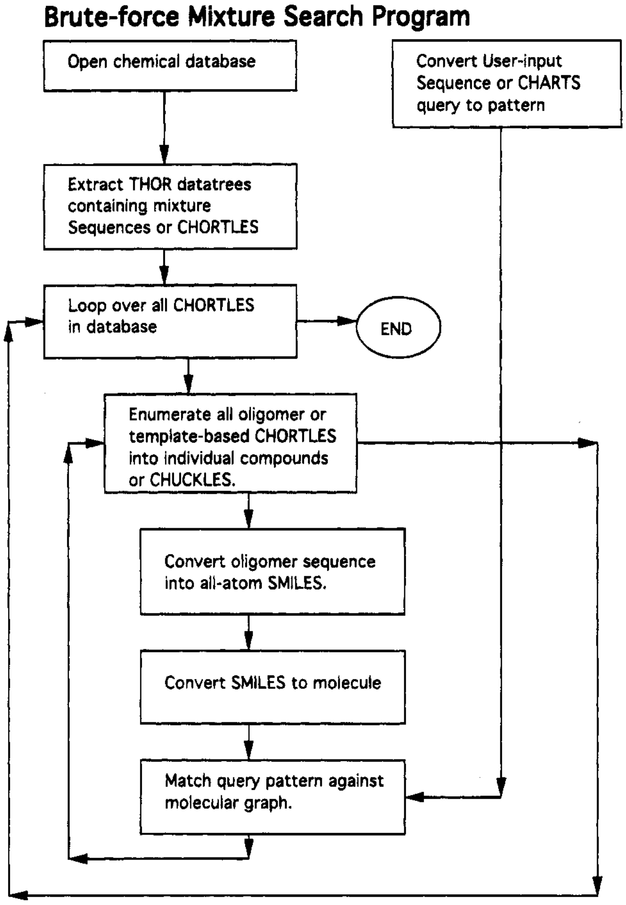

Brute-force Mixture Search Program
Open chemical database
Convert User-input Sequence or CHARTS query to pattern
Extract THOR datatrees containing mixture Sequences or CHORTLES
Loop over all CHORTLES in database
END
Enumerate all oligomer or template-based CHORTLES into individual compounds or CHUCKLES.
Convert oligomer sequence into all-atom SMILES.
Convert SMILES to molecule
Match query pattern against molecular graph.

Figure 5. Brute-force mixture search flowchart. The query is turned into a pattern to be matched against the molecular graph of each enumerated member of the mixture. Each member is enumerated on the fly.

We have implemented a simple brute-force search algorithm which successively expands each mixture sequence into its component sequences and then converts these to all-atom structures for substructure searching.

We implemented this brute-force search method using toolkits from Daylight Chemical Information Systems, Inc. The first step is to convert the query CHORTLES or SMILES to a Daylight pattern. The database containing the CHORTLES mixtures is then opened, and all mixtures are extracted. For each CHORTLES (oligomeric or template-based), all components are enumerated, yielding individual molecules or CHUCKLES. Each CHUCKLES is then converted into the all-atom SMILES or molecular graph. This molecular graph is then converted into Daylight's molecule representation against which the pattern is matched. This approach makes use of the standard Daylight matching routines without any special search capabilities. When a match occurs, the search is terminated. See Figure 5 for a flow chart description of this search method.

Consider the benzodiazepine library in Figure 2. If one is looking for molecules which contain aromatic rings substituted with a fluorine (query: c1c(F)cccc1), nowhere in the database is such a ring explicitly represented. The aromatic ring is contained in the benzodiazepine template and the fluorine substituent is a monomer at position 2. However, the molecule represented by the query is not explicitly represented in the molecular graph in Figure 2b; a straightforward substructure search of the database will not yield a match. A brute-force expansion of the benzo-

	Ala	Nser	[Arg;Lys;His]	Val
database mixture 1	1000000000000000000000000	0000000000000000000001000	0100000100010000000000000	0000000000000000000010000
	Ala	Nser	[Ala;Gly;Arg;Lys;His]	Val
database mixture 2	1000000000000000000000000	0000000000000000000001000	0100000100010000000000000	0000000000000000000010000
	Ala	Nser	[Ala;Gly;Ile;Leu;Phe;Val; Nala;Nphe;Nval]	Val
database mixture 3	1000000000000000000000000	0000000000000000000001000	1000010001100100000110011	0000000000000000000010000

sample query: AlaNserArgVal	1000000000000000000000000	0000000000000000000001000	0100000000000000000000000	0000000000000000000010000
database mixture 1	1000000000000000000000000	0000000000000000000001000	0100000100010000000000000	0000000000000000000010000
intersection of two strings	1000000000000000000000000	0000000000000000000001000	0100000000000000000000000	0000000000000000000010000

sample query: Ala[*]ArgVal	1000000000000000000000000	1111111111111111111111111	0100000000000000000000000	0000000000000000000010000
database mixture 1	1000000000000000000000000	0000000000000000000001000	0100000100010000000000000	0000000000000000000010000
intersection of two strings	1000000000000000000000000	0000000000000000000001000	0100000000000000000000000	0000000000000000000010000

Figure 6. (a) Bit-based CHORTLES search method. Each database mixture entry may be represented as a bit-string. Each mixture sequences has *n* bits per position where *n* is the number of monomers in the monomer table. For fixed positions, only one bit is set, the bit corresponding to the monomer at that position. For mixture positions, all the bits corresponding to the members of the basis set are set. (b) The query bit-string is set the same way. Wild cards may represent all the monomers at a particular position by setting all the bits at that position in the bit-string. (c) The intersection of the mixture and the query strings is used to determine whether the query sequence is contained in the database mixture sequence.

diazepine library CHORTLES and subsequent pattern matching of the query will yield a match. When enumerating the components of the benzodiazepine library CHORTLES, the compound in Figure 2c, Benzodiazepine1234567.Hyd1.Fluor2.-Hyd3.Hyd4.Diohpr5.Hyd6.Benz7 will come up. When it is converted to the molecular graph, it does contain the fluorine substituted aromatic ring, and the query pattern will yield a positive match. This method is compatible with any substructure searching method since each molecular graph is generated on the fly.

Standard sequence searches, via substring or regular expression,[21] can be performed directly on the mixture sequence or CHORTLES itself; no enumeration of the sequence is necessary. This is done by extending standard regular expressions to includes monomers as part of the character set. This approach is most appropriate for oligomeric libraries which might be represented in peptide-like sequences. There are severe limitations to this approach when searching template-based libraries since monomer definitions, and thus library representations, can vary for even the same sets of molecules.

Alternatively, CHORTLES lets us create a bit-based sequence pattern which is concise and easily compared to a potential query target. Given a monomer reference table with **n** entries, where **n** is a reasonable number (e.g., less than 10 000), one can create a checklist for each position in a mixture/peptide sequence of length **s**. For an arbitrary sequence we allocate a bit-string of length **n** monomers times **s** positions. With the monomer reference table order fixed, we assign each monomer entry bit **j** for each sequence position **i** to a monomer. Initially all bits are set to zero, and then the sequence is mapped onto the bit string. For fixed positions in a mixture, the bit corresponding to the given monomer is set. For mixture positions, all the bits corresponding to the members of the basis set are set. See Figure 6a for an example of the relationship between bit string and sequence. There is a one-to-one correspondence between the sequence bit string and the mixture sequence.

The query is converted to a sequence bit string. For defined fixed positions in the string, the bit corresponding to the specified monomer is set. For general and single-position wild cards, all bits at the position are set. That is, a wild card may match any fixed or mixture position. Interpretation of the query may be extended to match one or more positions against a wild card. See Figure 6b for the bit-string representation of a query.

The search is performed by bit-based intersection of the query bit string with each mixture bit string in the library/database. The resulting intersection bit string is decoded on sequence position boundaries such that a match at sequence position **i** is indicated by having at least one of the **n** bits at position **i** set. Having at least one bit set at every position indicates a match; that is, the query is contained in the mixture pool. See example in Figure 6c. This method may be extended so that exact mixtures are detected; that is, the query might contain mixture positions which must match the database sequences exactly.

This search method is fast because the bit strings for all entries in a database may be precomputed. They need only be updated when the monomer reference table is changed. Each query/comparison takes only as long as the bit-string intersection of two strings of length **n** times the maximum sequence length. In addition, this representation is compact because many related oligomers are represented by a single mixture sequence.

The comparison method may be extended to handle gaps in matches as in current sequence analysis methods. Also, wild cards may be interpreted to match more than one position; for example, a general wild card at a particular position may be represented by setting all the bits for that position in the query to 1.

Limitations. Brute-force mixture searches are time-intensive since all sequences contained within a mixture must be enumerated; however, there are various approaches to optimizing the search which could be used.

CHORTLES bit-based searches can be space-intensive. For example, consider a monomer table of 3000 entries. For sequences of 16 monomers, the bit string representation for each sequence requires 16*3000 = 48 000 bits. Precomputing the bit strings for 100 000 mixture pools would require

600 Mega-bytes of storage space. However, one is not likely to have 100 000 mixture pools from the whole monomer set of 3000, so one can subdivide the monomer reference table for different classes of mixtures. Also, one can generate the bit strings on the fly.

Both the regular expression and bit-based searching cannot cross monomer boundaries as a full, atomic-level substructure search can, and, thus, monomer definitions can effect search success.

CONCLUSIONS

The use of oligomeric and template-based mixtures in drug-lead discovery presents us with new data storage problems. The CHORTLES method provides a straightforward, table-driven approach to accurate and compact representation of mixtures. The method lends itself to easy traversal of the mixture deconvolution hierarchy. Both the brute-force enumeration search and the bit-based search illustrate that all the data about the individual components can be derived from the CHORTLES mixture sequences.

REFERENCES AND NOTES

(1) Geysen, H.; Meloen, R.; Barteling, S. Use of Peptide Synthesis to Probe Viral Antigens for Epitopes to a Resolution of a Single Amino Acid. *Proc. Natl. Acad. Sci. U.S.A.* **1984**, *81*, 3998–4002.

(2) Houghten, R. A.; Pinilla, C.; Blondelle, S. E.; Appel, J. R.; Dooley, C. T.; Cuervo, J. H. Generation and Use of Synthetic Peptide Combinatorial Libraries for Basic Research and Drug Discovery. *Nature* **1991**, *354*, 84–86.

(3) Lam, K.; Salmon, S.; Hersh, E.; Hruby, V.; Kazmiersky, W.; Knapp, R. A New Type of Synthetic Peptide Library for Identifying Ligand-Binding Activity. *Nature* **1991**, *354*, 82–84.

(4) Furka, A.; Sebestyen, M.; Asgedom, M.; Dibo, G. General Method for Rapid Synthesis of Multicomponent Peptide Mixtures. *Int. J. Pept. Protein Res.* **1991**, *37*, 487–493.

(5) Scott, J. K.; Smith, G. P. Searching for Peptide Ligands with an Epitope Library. *Science* **1990**, *249*, 386–390.

(6) Devlin, J. J.; Panganiban, L. C.; Devlin, P. E. Random Peptide Libraries. *Science* **1990**, *249*, 404–406.

(7) Cwirla, S. E.; Peters, E. A.; Barrett, R. W.; Dower, W. J. Peptides on Phage: A Vast Library of Peptides for Identifying Ligands,. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 6378–6382.

(8) Simon, R. J.; Kania, R. S.; Zuckermann, R. N.; Huebner, V. D.; Jewell, D. A.; Banville, S. C.; Ng, S.; Wang, L.; Rosenberg, S.; Marlowe, C. K.; Spellmeyer, D.; Tan, R.; Frankel, A. D.; Santi, D. V.; Cohen, F. E.; Bartlett, P. A. Peptoids: A modular approach to drug discovery. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 9367–9371.

(9) Zuckermann, R. N.; Kerr, J. M.; Kent, S. B. H.; Moos, W. H. Efficient Method for the Preparation of Peptoids [Oligo(N-substituted glycines)] by Submonomer Solid Phase Synthesis. *J. Am. Chem. Soc.* **1992**, *114*, 10646–10647.

(10) Ng, S. C.; Siani, M. A.; Bradley, E. K.; Moos, W. H.; Simon, R. J. Novel Backbones for Chemical Diversity: Oligo N-substituted Carbamates via a Submonomer Approach. *J. Am. Chem. Soc.* **1995**, in press.

(11) Cho, Y. C.; Moran, E. J.; Cherry, S. R.; Stephans, J. C.; Fodor, S. P. A.; Adams, C. L.; Sundaram, A.; Jacobs, J. W.; Schultz, P. G. An Unnatural Biopolymer. *Science* **1993**, *261*, 1303–1305.

(12) Bunin, B. A.; Ellman, J. A. A General and Expedient Method for the Solid-Phase Synthesis of 1,4-Benzodiazepine Derivatives. *J. Am. Chem. Soc.* **1992**, *114*, 10997–10998.

(13) Siani, M. A.; Weininger, D.; Blaney, J. M. CHUCKLES: A Method for Representing and Searching Peptide and Peptoid Sequences on both Monomer and Atomic Levels. *J. Chem. Inf. Comput. Sci.* **1993**,

(14) Zuckermann, R. N.; Martin, E. J.; Spellmeyer, D. C.; Stauber, G. B.; Shoemaker, K. R.; Kerr, J. M.; Figliozzi, G. M.; Siani, M. A.; Simon, R. J.; Banville, S. C.; Brown, E. G.; Wang, L.; Moos, W. H. Discovery of Nanomolar Ligands for 7-Transmembrane G-Protein Coupled Receptors from a Diverse (N-Substituted) Glycine Peptoid Library. *J. Med. Chem.* **1994**, *37*, 2678–2685.

(15) Carell, T. In *Exploiting Molecular Diversity*; Cambridge Healthtech Institute: San Diego, CA, 1994.

(16) Wickham, B.; Yazhong, P.; Marlowe, C.; Moos, W. New Chemical Diversity on the Solid Phase: Novel and Efficient Syntheses of 1,3,4,5-tetrasubstituted-2,5,-Diketo-1,4-piperazines (DKPs) on the Solid Phase. Manuscript in preparation.

(17) Weininger, D. SMILES, a Chemical Language and Information System 1. Introduction to Methodology and Encoding Rules. *J. Chem. Inf. Comput. Sci.* **1988**, *28*, 31.

(18) Daylight Chemical Information Systems, Daylight Software Manual: Theory; Santa Fe, NM, 1993.

(19) Genetics Computer Group, Wisconsin Sequence Analysis Package; Madison, WI, 1993.

(20) IntelliGenetics Corporation; IntelliGenetics: Suite 5.4, Mountain View, CA 94040, 1992.

(21) Muster, J.; Birns, P. Lurnix Unix Power Utilities for Power Users; Management Information Source, Inc.: Portland, OR, 1989.

CI9500501