Article

Beyond Mere Diversity: Tailoring Combinatorial Libraries for Drug Discovery

Eric J. Martin, and Roger E. Critchlow

J. Comb. Chem., **1999**, 1 (1), 32-45• DOI: 10.1021/cc9800024 • Publication Date (Web): 18 December 1998

Beyond Mere Diversity: Tailoring Combinatorial Libraries for Drug Discovery

Eric J. Martin*,† and Roger E. Critchlow‡

Chiron Corporation, 4560 Horton Street, Emeryville, California 94608, and Daylight Chemical Information Systems, Inc., 419 E. Palace Avenue, Santa Fe, New Mexico 87501

Received June 30, 1998

Combinatorial library design attempts to choose the best set of substituents for a combinatorial synthetic scheme to maximize the chances of finding a useful compound, such as a drug lead. Initial efforts were focused primarily on maximizing diversity, perhaps allowing some bias by the inclusion of a small, fixed set of pharmacophoric substituents. However, many factors besides diversity impact good library design for drug discovery. A library can be better "tailored" by assigning the candidate substituents to categories such as polar, pharmacophoric, rigid, low molecular weight, and expensive. Stratified sampling by successive steps of D-optimal design generates diverse designs which are also consistent with desirable profiles of these properties. Comparing the diversity scores among design profiles reveals the tradeoffs between diversity, physical property distributions, synthetic difficulty, expense, and pharmacophoric bias. The diversity scores can be calibrated by scoring the best designs from subsets of the candidates made either from specific classes of substituents or by randomly eliminating candidates. This procedure shows how poor random designs are compared even to highly biased optimal designs. Library design requires a synergistic effort between computational and synthetic medicinal chemists, so specialized interactive software has been developed to integrate substructure searching, display, and statistical experimental design to facilitate this interaction for the effective design of well-tailored libraries.

Introduction

Combinatorial library design is the attempt to choose the "best" set of substituents for a combinatorial synthetic scheme to maximize the chances of finding a useful compound, such as a drug lead. Initial efforts in combinatorial library design focused primarily on maximizing information content and minimizing redundancy by maximizing "diversity", allowing some bias by forcing the inclusion of a small, fixed set of pharmacophoric substituents.[1] Diverse libraries were designed, synthesized, and screened, and potent ligands were identified.[2,3] Inspection of the hits, however, revealed that many were more flexible, insoluble, lipophilic, or higher molecular weight than would be preferred in a drug lead. This outcome underscored the many factors beyond diversity which impact good combinatorial library design for drug discovery. Molecular weight range, lipophilicity, ease of synthesis, phamacophore focus, rigidity, reagent costs, solubility, incorporation of common drug fragments, complementarity to other libraries, and medicinal−chemical intuition should all be taken into account. Merely maximizing diversity has been shown to systematically bias the library away from the desired ranges for many of these properties. The goal of library design should be to provide high structural diversity while constraining pertinent physicochemical properties to suitable ranges for small-molecule drugs. We have developed a protocol, called "tailoring" the library, which achieves high diversity while emphasizing desirable attributes and identifying the tradeoffs which are inherent in the chemistry. One can quickly see how much diversity is sacrificed by using fewer groups that require protection or how many and which pharmacophoric fragments might best be included in a target-biased, yet diverse library. The entire process is illustrated in Figure 1. Suitable reagents are identified from a database of commercially available compounds. Structural properties are calculated for each candidate substituent. From these a "property space" is computed in which proximity reflects similarity. The substituents are also divided into "bins" based on additional properties, besides diversity, which are important in small-molecule drugs. Finally, a small set of substituents are selected from the candidates that maximize diversity while at the same time satisfying a specified profile of these additional properties. The following example will

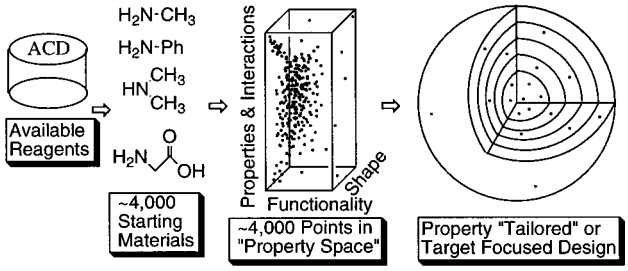

Figure 1. Schematic illustration of designing a "tailored" combinatorial library. Candidate reagents are selected, properties are computed, candidates are assigned to bins, and a small number of substituents are selected which maximize diversity while matching a desired profile of key pharmaceutical properties.

† Chiron Corporation.
‡ Daylight Chemical Information Systems, Inc.

10.1021/cc9800024 CCC: $18.00
Published on Web 12/18/1998

illustrate in detail all of the steps in generating a typical carefully tailored design and will present an analysis of the results.

Background Theory

Similarity, Property Space, and Multidimensional Scaling. Selection of substituents for a combinatorial library entails two steps: the calculation of a "property space" in which proximity between substituents reflects structural similarity, and the subsequent selection of points that are well distributed throughout that space. If the structural properties are correlated with activity, substituents tightly clustered in property space are potentially redundant, whereas a widely and evenly dispersed set is "diverse".

A vector of properties for each substituent can be regarded as coordinates in a property space. Often, however, similarities (i.e., distances) between compounds can be directly calculated, but the coordinates cannot. Given a table of properties (coordinates), the Pythagorean theorem can easily be applied to compute a matrix of all pairwise dissimilarities (distances). The reverse operation, calculating "latent" properties for each substituent from a dissimilarity matrix, is known as multidimensional scaling (MDS). It is computationally expensive, and if the similarity coefficient is not a "metric"—and the Tanimoto similarity coefficient used widely in drug discovery problems is not a metric[4]—it can only be achieved approximately.[5] It is currently practical for sets of up to about 10 000 substituents, adequate for most substituent selection problems.[6]

An algorithmic definition of "diversity" should incorporate two key concepts: "redundancy" and "coverage". A nonredundant set of points is widely separated in space. A set of points covers space if all regions of space, and in particular, all the dimensions of the space, are sampled. While there are methods for measuring the "diversity" of a set of points that only require a matrix of the distances between pairs of candidates, coordinates are required for the powerful covariance-based methods, such as D-optimal or A-optimal design (see Discussion). As an illustration, consider trying to select a diverse geographical sampling of the country. A map (i.e., the coordinates) would be more helpful than just a table of intercity distances. Methods based only on distances typically can select points spread out in space, i.e., nonredundant sets, but generally cannot identify collinearities, unrepresented holes in the space, determine whether all dimensions have been sampled, or recognize whether a point lies near the center of the space or near the edges. In short, pure distance-based methods assess redundancy only, but coordinate-based methods assess coverage as well. The extra effort to apply MDS to get coordinates is therefore justified, since a well-tailored library greatly benefits from actual coordinates rather than just a distance-based selection.

Experimental Design

Selecting a subset of experiments to best represent a much larger potential candidate set falls under the discipline of "experimental design". Experimental design has been applied to many pharmaceutical problems, including the design of structure−activity relationship (SAR) compound sets,[7−11] the optimization of synthetic processes,[12,13] the design of calibration standards in analytical chemistry,[14] and the selection of screening subsets from corporate chemical archives.[15] D-optimal design has recently been applied to selecting small numbers of substituents from larger sets of suitable reagents to use in synthetic combinatorial libraries for drug discovery.[16]

In cases where any combination of property values can be achieved, such as time, temperature, and reagent concentrations in a synthesis, precomputed classical designs are available. For choosing among chemical substituents, which offer only discrete, poorly distributed combinations of properties, algorithmic designs such as D-optimal or A-optimal design are preferred. In algorithmic designs, a "candidate set" is identified, and a much smaller "design set" is chosen from the candidates, which optimize a mathematical criterion for design quality, i.e., "diversity". An initial set of substituents can optionally be preselected for inclusion in the design, and that set can be "augmented" to the desired size, choosing the remaining members so as to optimize the diversity criterion. This set of substituents can then be used as the initial set in a subsequent design augmentation. This capacity to build up a complex design from successive augmentations is the basis for our approach to tailoring library designs.

Methods

Property Calculations. The property space of the substituents was calculated as previously described.[1] The space included the calculated octanol/water partition coefficient, "shape" descriptors derived from principal components analysis of 81 topological indices, "chemical functionality" descriptors derived from MDS of Tanimoto similarities based on Daylight 2-D substructure fingerprints, and "receptor recognition" descriptors derived from MDS of similarities from "atom layer tables", which give the distribution along the substituents of atoms, hydrogen-bonding groups, charged groups, and aromatic groups. The numbers of chemical functionality and receptor recognition descriptors were determined as the fewest MDS dimensions required to reproduce the dissimilarity matrix with a relative standard deviation of 10%. These properties were chosen to characterize similarity and diversity with respect to lipophilicity, shape, chemical functionality, and distribution of key receptor binding features. The properties were automatically calculated using the program MAKESPACE, which comprises a collection of commercial programs, C and FORTRAN code, and Tcl scripts, all coordinated through the UNIX "make" utility. The program takes lists of reagents or substituents as input. It normalizes the structures by removing counterions, standardizing resonance forms, etc. It then determines the best commercial source for each reagent, computes the similarities and properties, performs MDS to create the property space, and loads the structures and properties into a THOR database[17] for searching with MERLIN.[18]

Substructure Searching. Tailored library design requires the ability to rapidly and interactively sort or subset a collection of reagents or candidate substituents by structural criteria or physical properties. This is used both to identify

feasible candidate reagents and to categorize substituents into "bins" (see Results). Searching and sorting were performed with MERLIN, and the subsets were stored as THOR data tree files (TDT). Designs were displayed by TKPRADO, an interactive, window driven, Tcl/Tk script that displays structures using the Daylight PRADO utility.

Optimal Designs. The thousands of optimal designs computed to choose the substituents are set up by an interactive, window driven, Tcl/Tk program TAILOR, which executes a slightly modified version of the public domain FORTRAN program DOPT to perform the D-optimal designs.[19,20] TAILOR also creates and manipulates files and directories to organize the input and output data. It reads and writes TDT files to interface with the Daylight software for searching and displaying the chemical structures. In addition, TAILOR performs automatic calibration designs to aid interpretation of the results and also automatically finds the best replacements for each compound in the design (see below). MAKESPACE and TAILOR are not commercially available programs, but most of the individual modules that comprise them are.[1] MAKESPACE and TAILOR primarily automate their application.

To circumvent the D-optimal algorithm's tendency to select some points twice (see Discussion), TAILOR can "saturate" the designs, i.e., it generates models where the number of terms are equal to the desired number of substituents. Saturated models are built up systematically by first including an intercept and linear terms, in order, starting with the largest principal component (PC). Squared terms can optionally be added, followed by enough cross terms starting with PC1 × PC2, PC1 × PC3, PC2 × PC3, etc. to saturate the model. For an N-dimensional property space, this protocol covers up to $(N^2 + 3N)/2$ points, e.g., up to 170 substituents for the 17 dimensions in the current example. Higher terms can also be added if needed.

Results

Candidate Reagent Identification. Library design begins with identifying the candidate pool of potential substituents. This example study began with 7812 aliphatic amines and 9077 phenols in the available chemicals directory[21] (ACD). Most of these were not suitable, being too large, too expensive, too hard to obtain, insufficiently reactive, containing toxic or labile moieties, or requiring too much protection chemistry to be practical. The unsuitable compounds were removed by a series of substructure and property range searches using MERLIN. Table 1 shows the searches that were used to remove undesirable amines. The specifics are particular to each reaction scheme, and filters are applied interactively while continuously reevaluating the reagent list. For example, Br and I are passed by the very general rule in step 6, but selected for removal later in step 19. Some searches even included a hand inspection of the results in which only some of the matching compounds were eliminated. An extensive catalog of previously compiled queries are available to streamline the filtering process. Even so, an experienced library designer—working closely with the chemist who developed the synthetic scheme—should expect to spend no less than 2−4 h on this step. This is time well spent. Culling out only the suitable reagents at the outset saves far more time later on. The final list yielded 1083 amines and 825 phenols.

A complete set of useful reagents is not yet a substituent candidate set. For instance, the same amine might be available with two different counterions or be listed in the ACD under several tautomeric or resonance forms. Likewise, an unsymmetrical diamine could generate two different substituents. MAKESPACE uses an extensive set of 144 rules to "standardize" the structures by eliminating counterions and choosing standard tautomers and resonance forms. This set of rules has been developed and tested over many years and correctly handles virtually every small molecule in the current ACD database. The original reagent structure is kept as a link back into the ACD for later ordering the reagents. Another set of rules identifies the best vendor and price.

Property Calculation. The amine and phenol sets were combined for a total of 1693 compounds for the property calculations. Thus, all substituents for both positions were embedded into a single property space, allowing diversity to be maximized between sites as well as within sites. Substituent properties were calculated by MAKESPACE as described above, and a THOR database was generated containing the standardized substituent structures, vendors, prices, preferred names, and some computed properties including log P, pK_a, MW, number of rotatable bonds, and distance from the centroid of property space. The property space required five shape descriptors, seven chemical functionality descriptors, and six receptor interaction descriptors, as well as log P, for a total of 19 dimensions. Principal components (PCs) analysis showed that 17 PCs explained 99.5% of the variance, so the last two PCs were disregarded. In this example, the final substituent database contained 943 amines and 750 phenols.

Creation of Bins. The next task was the creation of bins. To control the distribution of properties present in the library and to evaluate the tradeoffs between high diversity and bias toward drug-like properties, one first assigns the candidate substituents to subsets representing desired (or undesired) properties. These subsets of substituents to be emphasized or deemphasized in the design are called "bins". Each substituent has many properties, and the bin catagories overlap, so most are assigned to several bins. Table 2 shows the bins used for the amine-derived position. The library was to emphasize rigid, polar, validated, drug-like, and pharmacophoric substituents. The exact number of compounds to be used from each bin was to be determined later as part of the design process. The "validated" bin contained 76 amines for which the yield and purity of the reaction had been confirmed. The "seed" bin contained 4-methoxybenzylamine, which was the validated point nearest the centroid of the property space, as well as 4-hydroxyphenethylamine and benzhydrylamine, whose corresponding side chains were previously found in potent ligands for the α_1-adrenergic and μ-opiate receptors.[2] The "extreme" category held 27 of the 35 amines farthest from the centroid of property space. These compounds were mostly complex hydrocarbons or sugar analogues. Typically, many intuitively undesirable com-

Table 1. Criteria Used To Cull Acceptable Candidate Amines[a]

description	SMARTS
keep aliphatic primary	[CX4][NH2]
keep aliphatic secondary & take union	[CX4][NH][CX4]
remove MW > 250	N. A.
remove obscure vendors	N. A.
remove cost > $500/g	N. A.
remove weird elements	[!C!c!O!o!N!n!S!s!F!Cl!Br!I!Na!Ca!K!P]
long unbranched chains	[D2R0]~[D2R0]~[D2R0]~[D2R0]~[D2R0]~[D2R0]
aromatic tricyclicbridgehead	[aR3]
aromatic many cycles (kept a few)	[aR2]a[aR2]
bridgehead (kept if substituted)	[R3]
fluorines	F. F. F. F. F.F
linear F's	FC(F)C(F)F
acids and enols	*=*[OH]
long skinny (kept some)	[D2R0]~[D2R0]~[D2R0]~[D2R0]~[D2R0]
thiophenes, furans	[s,o]1cccc1
big rings	[r8,r9,r10,r11,r12,r13,r14,r15]
epoxides	C1OC1
alpha eliminators	[O,o,N,n][CX4][O,o,N,n]
Br or I	[Br,I]
enol-ethers	C=CO
disulfides	SS
benzofurans	o1ccc2c1cccc2
benzoquinones	O=C1CCC(=O)c2c1cccc2
N−O	N−O
aldehydes	O=[CH]
alkyl halides	[CX4][Cl,Br,I]
isocyanates	N=C=O
sulfides, disulfides	[SD2]
mixed, unsymmetrical diamines	[CX4][NH][CX4].[NH2][CX4]
secondary diamines (keep symmetrical)	[CX4][NH][CX4].[CX4][NH][CX4]
primary diamines (keep symmetrical)	[NH2][CX4].[NH2][CX4]
cyclopropylamines	[NH]1CC1
elimination problems	[NH,NH2]CC[$(C#N),$(N=O),$(S=O),$(C=O),[N+]]
elimination problems	[$(C#N),$(N=O),$(S=O),$(C=O),[N+]][CH2][CH2][$(OC=O),$(NC=O),Br,Cl,I, [N+],[n+],$(OS=O),$(ON=O)]
anilines	[$([NH][CX4,c]),$([NH2])]c
hindered primary amines	[$([NH]C(*)(*)*),$([NH]C*(*)(*)*)
hindered secondary amines	*[CX4](*)([NH])[CX4](*)*
acidic C−H	[$(C#N),$(N=O),$(S=O),$(C=O),$(cn),$(cccn),[N+]][CH2,CH][$(C#N),$(N=O), $(S=O),$(C=O),$(cn),$(cccn),[N+]]

[a] The SMARTS queries in column 2 can be used in Merlin to eliminate reagents with the features described in column 1.

pounds are concentrated at the extremes of property space. If they were not isolated in a special bin, they would be highly sampled in a diversity design. We wanted to limit the number of highly fluorinated compounds, so the 30 amines with three or more fluorines were isolated in the FF bin. The "good" bin held 699 amines that excluded the extreme and FF groups, as well as some anticipated synthetic and metabolic problems. "Rigid" held the 346 good amines with two or fewer rotatable bonds. Table 3 shows a list of significant fragments found in the 100 top-selling drugs. The "Drugish" bin held 151 good amines that contained these fragments. The "polar" bin held 378 substituents with at least one H-bond acceptor atom and an estimated octanol/water partition coefficient (log *P*) < 1.5 as a neutral amine. The "tbu" bin held carboxylic acids available protected as the *tert*-butyl ester. The "boc" group held diamines that can be purchased boc-protected, and "Okamine" held diamines expected to work without protection. The remaining bins are self-explanatory combinations of the sets above.

Calibrations. To evaluate the loss of diversity incurred by forcing the design to conform to various property profiles, we first calculated some benchmark designs to identify the boundaries of the diversity scale. The D-optimal diversity scores have relative significance for designs of a given size and model, but are difficult to interpret on an absolute scale. The purpose of this particular design was to select a set of 50 substituents for a diversity position in a broad screening combinatorial library, i.e., it was not aimed at any particular biological target. Calibration designs were made by selecting both D-optimal and random sets of 50 substituents from the larger of the bins. These benchmark designs, shown in Table 4, establish a ladder of "D-scores" to calibrate the diversity of the tailored designs as described below, as well as demonstrating the diversity inherent in the individual bins. For comparison between D-optimal design and a purely distance-based method (see Discussion), some S-optimal designs were also computed using the SAS QC software.[22] The "S-scores" are listed in column 6. All D-optimal designs were made for a quadratic model saturated with cross terms using 17 of 19 PCs. The "methods" column indicates how the design was created. "100%" simply indicates finding the D-optimal (or S-optimal) design of 50 from the full set mentioned in the second column. Other percentages indicate that unbiased subsets were created from the "all" set by first randomly choosing 50%, 20%, or 10% of the all set and then computing the optimal design of 50 substituents from

Table 2. Bin Profiles Used in Amine Design[a]

name	description	no.	range	tries
center	validated point nearest centroid	1	0	0
pharma	from previous assay hits	5	0	0
rigid	two or fewer rotatable bonds	346	0	0
FF	more than three fluorines	30	0	0
extreme	more than three std dev from centroid	27	0	0
seed	center + two pharmas	3	3	1
LoRgPlV	low MW, rigid, polar, validated	13	-10−14	3
boc	available with boc-protected amine	6	1	1
tbu	available with *t*-butyl-protected acid	9	2	2
LoPlrV	low MW, polar, validated	21	-15−19	3
LoRgPlr	low MW, rigid, polar	77	5−6	2
LoPlr	low MW, polar	126	4−5	2
DrgPlr	drug-like and polar	88	-30−34	3
OKamine	diamines that need no protection	46	1	2
DrugV	drug-like and validated	11	2−4	2
valid	reaction has been validated	76	2	2
Drugish	contain pieces from 100 top drugs	151	1	1
LowMW	molecular weight < 130	229	2	2
polar	H-bond acceptor and log P < 1.5	378	-46	3
FF_xtrm	union of FF and extreme bins	57	2	2
good	chemistry is expected to work well	699	2	2
HC	hydrocarbons only	177	0	0
tyr_lik	closest analogues to tyramine	50	0	0
all	union of all bins above	762	0	0

[a] Column "no." is the number of substituents that fit that category. Column "range" gives the acceptable numbers of substituents to be taken from that bin. A range starting with a hyphen (-) indicates taking enough substituents from that bin to bring the total to that number. Column "tries" shows the number of numerical attempts to find the D-optimal compound set.

Table 3. Fragments Found in 100 Top-Selling Drugs

fragment	count	fragment	count
5-aromatic, [1,3] heteroatoms	21	pyridine	6
beta lactam	19	furan	5
5-aromatic, 1 heteroatom	16	quinazoline	5
6-nonaromatic, [1,4] heteroatoms	16	imidazole	4
pyrrolidine	16	indole	4
5-nonaromatic, [1,3] heteroatoms	14	naphthalene	4
6-aromatic, 1 heteroatom	13	purine	4
6-aromatic, [1,3] heteroatoms	13	benzimidazole	3
piperazine	12	pyrimidine	3
guanidine	11	quinoline	3
nitro	10	1,3-dioxolane	2
piperidine	7	morpholine	2
thiazole	7		

those random subsets. The "random" method refers simply to choosing 50 substituents from a set at random and then determining its score with no optimal design step. All random selections and subsequent optimizations were repeated five times, so an average score and standard deviation are reported.

Since the "all" set is the union of all bins, it establishes the maximal possible D-score as 136. The set of 50 closest tyramine analogues establishes a very nondiverse benchmark, and its score of −148 can be considered a practical lower bound. While there is some correspondence between the size of a set and the resulting score, the correlation is low. Except for the very restrictive set of low molecular weight−rigid−polar compounds, the worst of the optimized designs from restricted bins is still much better than selecting 50 substituents at random from the "all" set. This illustrates how very poor random designs are. The penalty for eliminating 50% of all synthetically suitable compounds is not great, and even eliminating 80% of the compounds at random retains diversity comparable to that of the rigid or polar compounds alone. The low molecular weight restriction limits diversity drastically, almost as much as a calibration bin that only includes the pure hydrocarbons.

Table 4. Results of Calibration Designs[a]

method	bin	size	D-score	D-σ	S-score	S-σ
100%	all	762	136		3.23	
50%	all	381	124	2.6	3.01	0.02
20%	all	152	105	4.5	2.65	0.05
100%	rigid	346	105		2.63	
100%	polar	378	101		2.58	
10%	all	76	87	2.9	2.13	0.09
100%	RgPlr	146	81		2.32	
100%	valid	76	80		2.11	
100%	Drugish	151	76		1.94	
100%	LowMW	229	66		2.01	
100%	DrgPlr	88	60		1.78	
100%	HC	177	61		2.00	
100%	LoRg	154	55		1.92	
random	all	50	50	3.6	1.42	0.16
random	valid	50	47	7.5		
random	polar	50	39	5.8		
random	RgPlr	50	38	5.6		
random	rigid	50	37	8.2		
random	DrgPlr	50	21	4.6		
100%	LoRgPlr	77	16			
random	LowMW	50	5	7.8		
random	LoRg	50	3	10.0		
random	Drugish	50	−1	7.0		
random	LoRgPlr	50	−9	6.9		
random	HC	50	−16	5.4		
100%	tyr_lik	50	−148		0.5	

[a] Column "method" refers to a random fraction of the "bin", in column 2, used for the design. ("Random" sets of 50 are 6.6% of the "all" bin). Bin descriptions are in Table 2. Column "size" is the number of candidates in the subset. D-score and S-score are scores for D-optimal and S-optimal designs, respectively. Random subsets were each sampled five times, with "D-σ" and "S-σ" showing the standard deviations for the five attempts.

D-Optimal "Tailored" Designs. Anticipated ranges of bin membership for the tailored design were selected as shown

Table 5. Scores and Profiles of Selected "Tailored" Designs from Table 2

rank	score (83−101)	LoRgPlV (7−11)	LoPlrV (0−6)	LoRgPlr (5−6)	LoPlr (4−5)	DrgPlr (0−10)	DrugV (2−4)	polar (2−8)	drug (2−14)	valid (11−17)	polar (30−32)	low (18−24)	rigid (12−17)
1	**101.6**	8	1	5	4	9	3	4	12	12	31	18	13
2	**100.6**	9	1	6	4	4	2	8	6	12	32	20	15
35	***97.1***	***10***	***0***	***6***	***5***	***7***	***4***	***2***	***11***	***14***	***30***	***21***	***16***
67	96.1	11	0	6	5	2	4	6	6	15	30	22	**17**
74	95.9	9	0	5	4	10	4	2	**14**	13	30	18	14
178	94.7	8	5	6	4	5	4	2	9	**17**	30	23	14
470	92.4	8	5	6	5	0	4	6	4	17	30	**24**	14
1320	82.8	8	4	6	4	6	3	3	9	15	31	22	14

[a] Column 1 is the rank order. Column 2 is the D-score. Columns 3−9 are the number of points drawn from each bin described in Table 2 for which a range was specified. The range is indicated by the numbers in parentheses. Columns 10−14 are total numbers drawn from bins that require each property. E.g., the "drug" column is the sum of the "Drug-polar" and "Drug-valid" columns.

in Table 2. Some categories, like small, rigid, polar groups, were emphasized. Others, like large, hydrophobic, and extreme groups, were deemphasized. The D-optimal designs were all made for a quadratic model saturated with cross terms, using the largest 17 of 19 PCs only. Separate designs were made for each of the 1320 profiles consistent with these ranges and a total design size of 50 amines, i.e., each design of 50 was generated from a profile starting with the "seed" set and then augmenting by D-optimal design with the specified number of substituents from bin "LoRgPlV". This combined set was further augmented with the "boc" set, and so on, until all bins had been appropriately sampled. The order was chosen to sample the most restrictive sets early, so the optimization algorithm has broad latitude to fill remaining holes in property space from the most general sets toward the end. Since the sets can overlap, the specified ranges actually determine a minimum bias, e.g., the selections from the final "good" bin might also be LowMW and/or polar and/or rigid as well. Likewise, sets to be deemphasized, like FF_xtrm, must not overlap a general set (such as good), unless the set is never actually sampled (such as ALL, which has a range of 0). Each D-optimal step was performed several times, as shown in the "tries" column, to avoid local optima. With 1320 profiles times 34 D-optimal steps per profile, this required performing 44 880 separate D-optimal designs. The entire calculation took about 1 day elapsed time on an SGI Indigo 2 with a 150 MHz R4400 CPU running IRIX 5.2.

Some notable designs are presented in Table 5, in order of decreasing diversity. The most diverse of the 1320 tailored designs had a score of 102. Comparison to the benchmarks in Table 4 reveals that this score is comparable to eliminating 80% of the candidates at random or to the best designs from the rigid or polar subsets. This significant, but acceptable, reduction in diversity is the penalty paid to achieve a profile of properties suitable for bioavailable drugs. The worst tailored design had a score of 83, slightly worse than an optimal design made after eliminating 90% of all feasible candidates at random, but still much better than a simple random selection of 50 compounds from all feasible candidates, with an average score of 50. It is usually possible to generate carefully tailored designs that maintain much of the possible diversity.

Table 5 also lists the bin membership from those bins for which a range was specified. Some bins combine several overlapping categories, so Table 5 also shows the total number of drug-like, valid, polar, low MW, and rigid compounds chosen in each design, e.g., the "drug" column is the sum of the "drug-polar" and "drug-valid" columns. Surprisingly, the most diverse design, no. 1, has an unusually high drug-like bias, having 12 of a possible 14 members from the drug-like bins. However, its members tend to be high molecular weight and flexible, and very few were validated. Design 2 has only 6 drug-like substituents, but has better low MW and rigid bias. Design 67, with a score of 96, is the most diverse of the designs with the maximum 17 rigid groups. The least effort would be required for those libraries with the full 17 validated substituents. The most diverse among these is no. 178 with a score of 94.7. This design is fairly good in every category and would be a good candidate for synthesis, except that its diversity is disappointingly low. Design 470, the best design with the maximum 24 low MW substituents, scores only 92.4.

Since this example library was intended primarily for broad screening, a high diversity score was desired. Design 35, with a score of 97, was chosen as a good compromise between diversity, property bias, and synthetic ease. Even a very thorough initial review of the candidate reagents rarely eliminates all undesirable substituents. In this case, examination of the 50 chosen substituents revealed that one was a dipeptide, which might cause formulation, delivery, and metabolism problems. A second, more general complaint was that the library contained 10 substituents with amide hydrogens, which are believed to carry similar liabilities.[23] A new design rectified these deficiencies. The six dipeptide reagents were eliminated from all bins by a substructure search. The number of substituents with amide protons was reduced by allowing them in the most elite bins, but removing them from the more general LoPlr, LowMW, polar, FF_xtrm, good, Drugish, and valid bins. The new bin sizes and design profile are shown in Table 6. This profile, which was focused around the previously favored design 35, contained only 18 possible designs, so the calculation took only 20 min. The results are shown in Table 7. The best design had a D-score of 96, showing only a small diversity penalty for this improvement, and only 6 of 50 substituents now had amide hydrogens. For this library, the design was deemed acceptable. More typically, however, further analysis would lead to additional cycles of evaluation and refinement, each taking about 10−20 min, before achieving a final acceptable design.

Having completed the amine design, a 50-member phenol design was created by using the amine design as a 50-

Table 6. Bin Profiles Based on Design 35 for the Refinement Design To Remove Dipeptides and Reduce the Number of Amide Protons

name	description	no.	range	tries
seed	center + two pharmas	3	3	1
LoRgPlV	low MW, rigid, polar, validated	13	9−11	3
boc	available with boc-protected amine	6	1	1
tbu	available with *t*-butyl-protected acid	9	2	2
LoPlrV	low MW, polar, validated	21	0−1	1
LoRgPlr	low MW, rigid, polar	77	5	2
LoPlr	low MW, polar	112	5	2
DrgPlr	drug like and polar	88	6−8	3
OKamine	diamines that need no protection	46	1	1
DrugV	drug like and validated	11	4	2
valid	reaction has been validated	76	2	2
Drugish	contain pieces from 100 top drugs	139	1	1
LowMW	molecular weight < 130	215	2	2
polar	H-bond acceptor and log P < 1.5	331	-46	3
FF_xtrm	union of FF and extreme bins	54	2	2
good	chemistry is expected to work well	649	2	2

Table 7. Scores and Profiles of the 18 Refinement Designs from Table 6, Similar in Profile to Design 35 in Table 5, but with Dipeptides Deleted and with Amide Hydrogens Allowed Only in the Most Elite Groups

rank	score (86−96)	LoRgPlV (9−11)	LoPlrV (0−1)	DrgPlr (6−8)	polar (1−4)	drug (10−12)	valid (13−16)	polar (30)	low (19−21)	rigid (14−16)
1	96	10	0	7	3	11	14	30	20	15
2	95	9	0	6	5	10	13	30	19	14
3	93	9	0	7	4	11	13	30	19	14
4	93	11	0	6	3	10	15	30	21	16
5	92	9	0	8	3	12	13	30	19	14
6	92	9	1	6	4	10	14	30	20	14
7	92	10	1	8	1	12	15	30	21	15
8	91	9	1	8	2	12	14	30	20	14
9	91	9	1	7	3	11	14	30	20	14
10	91	10	1	7	2	11	15	30	21	15
11	90	10	0	8	2	12	14	30	20	15
12	89	11	1	7	1	11	16	30	22	16
13	88	10	0	6	4	10	14	30	20	15
14	88	11	0	8	1	12	15	30	21	16
15	87	11	0	7	2	11	15	30	21	16
16	87	10	1	6	3	10	15	30	21	15
17	87	11	1	8	0	12	16	30	22	16
18	86	11	1	6	2	10	16	30	22	16

member "seed" bin and augmenting it from phenol candidate bins to yield a tailored design of 100 members. The tailoring of the phenol design was completely analogous to the amine design just described. Including the amine derived substituents in the phenol diversity design ensures high diversity between positions in the final library as well as within positions.

At this point the reagents could be ordered and validated. Inevitably, some reagents will be out of stock, and others will fail in validation. To save time later, Tailor automatically generates the best D-optimal alternates for each member of the design, so that they can be ordered in advance and validated in parallel. To keep the property profile unchanged, each alternate is chosen from the same bin from which the original substituent had been selected. Figure 2 shows some examples. In the first two cases, the best replacement was an obvious analogue of the original substituent, but in the other two examples, the D-optimal algorithm completed the design by selecting a radically different structure, presumably to fill another large hole in diversity space. The latter examples are more useful, because if the original selection fails in validation, that does not suggest that the replacement will likewise fail. Since the same reagent was often the best alternate for more than one substituent and it is helpful to find a dissimilar replacement, second, third, and higher alternates are often generated. Order sheets were automatically generated for all reagents, sorted by preferred vendor, including price, structure, name, and catalog number.

Discussion

Whole Molecule vs Fragment Based Properties. The design of the tailored library was based on the calculated properties of the fragments from the variable positions in the library, rather than on the assembled final molecules. This substituent approach takes advantage of the inherent structural similarities between all of the members of a combinatorial library. It assumes that diverse substituents generate diverse libraries. The obvious risk of working with substituents is that a design based on fragment properties does not explicitly account for interactions between the fragments in the assembled molecules, so assembling diverse substituents might not result in diverse molecules. However, since a *combinatorial* library includes every combination of substituents, it does form a full factorial design to characterize any interactions implicitly. The whole molecule alternative is limited by computer resources. If any of 1000 substituents

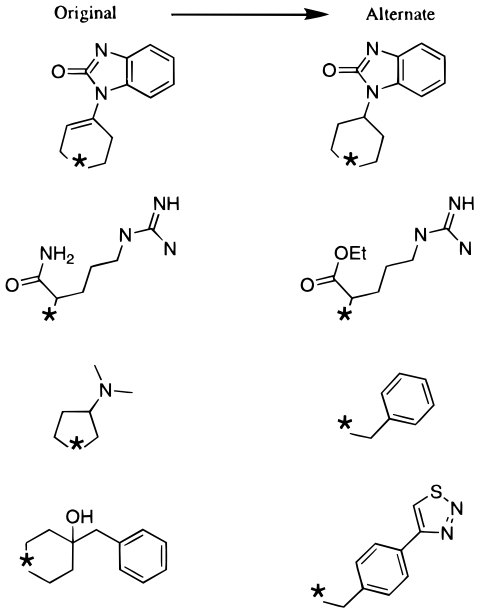

Figure 2. Sample alternates chosen by D-optimal design. In the first two cases, the best replacement was an obvious analogue of the original substituent. In the other two examples, the D-optimal algorithm completed the design by selecting a radically different structure, presumably to fill another large hole in diversity space.

could be put at each of 4 positions on a scaffold, the enumerated virtual library would contain 10^{12} assembled compounds. The method just described can be performed routinely on 4000 substituents in a few days, which has been sufficient for every library design problem we have encountered. Any approach which could be applied to even 10^9 assembled molecules could only use the most primitive of property spaces and selection methods. We believe the assumption that diverse substituents yield diverse libraries is less dangerous than the assumption that very simplistic property calculations and selection procedures would yield effective libraries.

Furthermore, property space formed from whole molecule calculations may not describe the diversity of the final combinatorial library as well as the corresponding substituent calculations. For example, the 2-D Daylight fingerprint descriptors only include paths of up to seven bonds. However, diversity libraries based on side chain descriptors for three positions incorporate information on up to 21 bonds per compound in the final library. Furthermore, including the scaffold in the calculation disguises information about the substituents. Any fingerprint bits set by the scaffold are set in every single molecule, so the presence or absence of those substructures cannot be distinguished in the side-chains. Patterson et al. described this phenomenon when they found that 2-D fingerprint similarity for substituents correlated with biological activity consistently better than the corresponding whole molecule descriptors across 20 quantitative SAR data sets from the literature.[24] We have tried designing libraries where each substituent was attached to the scaffold before calculation of the property space, and we likewise found that the computed similarities between members increased and many important structural distinctions were lost. Again, these arguments apply specifically to the design of combinatorial libraries, such as those made by split and mix resin synthesis or parallel "array" synthesis. Other diversity design problems, such as selecting subsets of corporate archives or purchasing compounds from collections of arbitrary structures, require methods that can deal with whole molecule descriptors.

D-Optimal Design. D-optimal design works by choosing a subset of substituents from a large candidate set, maximizing the determinant of the "information matrix", $|\mathbf{X}'\mathbf{X}|$, for a design matrix $\mathbf{X}$.[25] The rows of $\mathbf{X}$ are the substituents, and the columns are the "model terms", i.e., the property space dimensions, and/or higher order terms such as their squares or cross terms. This minimizes the determinant of the inverse and, thus, the prediction error for a regression model. Equivalently, information theory shows that this same criterion maximizes the expected entropy change, i.e., it selects the set of substituents that together carry the most information for estimating the model.[8] Roughly speaking, maximizing the determinant requires large diagonal elements and small off-diagonals. This implies large variances, so the selected points are well spread out, and small covariances, so collinearities are minimized and all the dimensions are sampled.

In substituent-based, tailored, combinatorial library designs the total number of substituents at each stage of augmentation is small, and the number of dimensions in property space is comparatively large. In fact, the dimensionality must often be reduced by principal components analysis, so that the number of dimensions (degrees of freedom) does not exceed the number of substituents. Since covariance-based methods, like D-optimal design, minimize collinearities as well as maximizing spread, they should still produce "balanced" designs that optimally sample all dimensions, even when there are few or no extra degrees of freedom. Since simple distance-based methods ignore collinearity, they are likely to select sets of points that do not sample the full dimensionality of the property space.

A simple 3-D problem was devised to graphically illustrate the advantages of covariance-based methods such as D-optimal design, over a standard distance-based method, at sampling the full dimensionality of property space. Although property space is often presented as a hypercube, examining our property spaces computed from many actual substituent sets (including this study) showed that the distribution of points was always roughly elliptical; so for this example, a random set of test points was generated inside an ellipsoid from the equation $x^2/1^2 + y^2/0.9^2 + z^2/0.8^2 = 1$. (Using the first three principal components of real data sets gave results that were essentially the same as this test case.) The simplest distance-based method that has been recommended for substituent selection is "MaxMin", which selects points to maximize the smallest near-neighbor distance in the design set.[26] While this criterion may work for pure diversity designs, it is inappropriate for tailored designs, since one often purposely includes some pharmacophoric analogues that lie in close proximity in property space, and these would dominate the MinMax score. Other distance-based methods spread points out in space by maximizing various averages of all of the near-neighbor distances or of all pairwise distances. The "S-optimal" method ("S" for "spread") maximizes the harmonic mean of distances between each

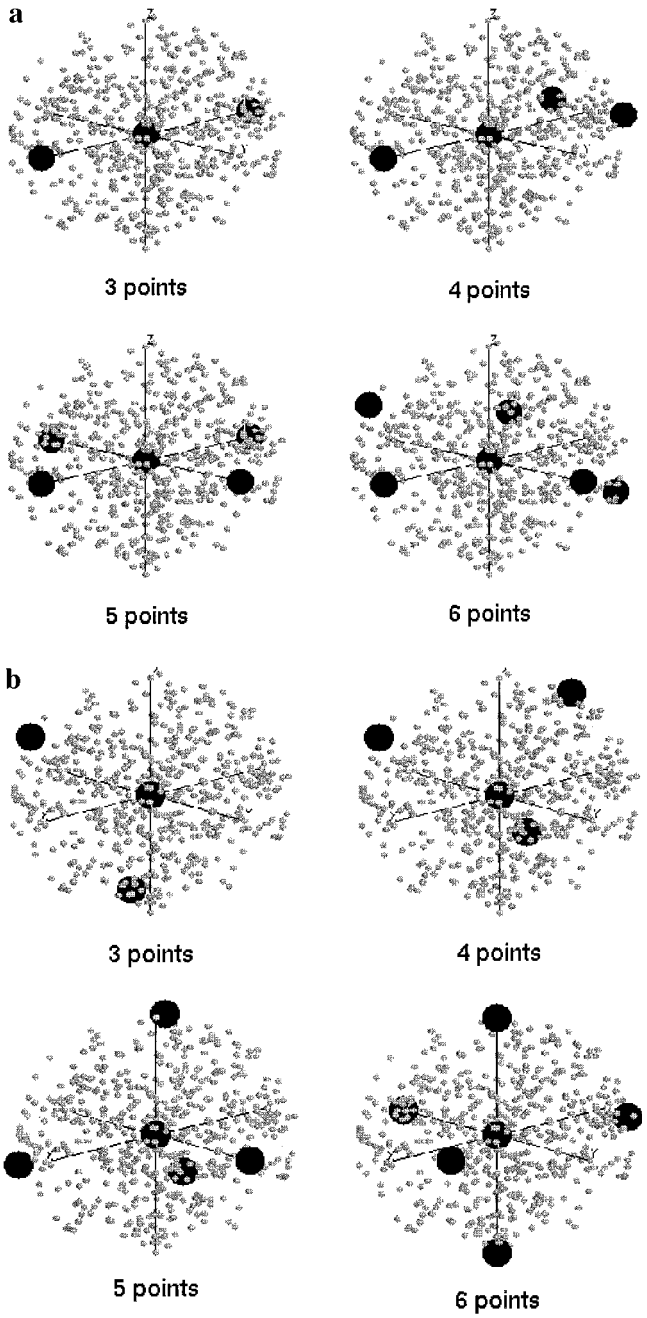

Figure 3. (a) Sample 3-D S-optimal designs for three, four, five, and six points including a center point. Random candidate points were generated inside an ellipsoid from the equation $x^2/1^2 + y^2/0.9^2 + z^2/0.8^2 = 1$. (b) Sample 3-D D-optimal designs, using a linear model, for three, four, five, and six points including a center point. Random candidate points were generated inside an ellipsoid from the equation $x^2/1^2 + y^2/0.9^2 + z^2/0.8^2 = 1$.

chosen point and its nearest neighbor in the design.[22] Since it includes all near-neighbor distances but the harmonic mean gives extra weight to the shorter distances, it is well suited for diversity design. Criteria that maximize all pairwise distances, rather than just near-neighbor distances, tend to "flatten" the designs into the largest few dimensions.

The test results are shown in Figures 3a,b. In each example, inclusion of the origin was forced as a required bin. Figure 3a shows the most "diverse" sets of three, four, five, or six points, including the fixed center, selected by S-optimal design. The three-point design is collinear. Four- and five-point designs are coplanar, so the "*z*" property is never varied. The six-point design is a slightly puckered pentagon so it finally gives some (small) variation to the *z* dimension. The second set of plots shows the points selected by the corresponding D-optimal designs using a linear model. The three-point design is an obtuse triangle, and thus varies two dimensions. The four-point design is a flattened tetrahedron and samples all three dimensions. The five-point design is a beautifully balanced design consisting of a large tetrahedron plus the origin at its center. Six points yield a triangular bipyramid plus center point. Seven points (not shown) give an octahedron plus the center. Evidently, because D-optimality sacrifices some spread in order to minimize multicollinearities, it samples all of the dimensions of property space even with very few extra degrees of freedom. Hence, for tailored designs, where the number of points is frequently close to the number of dimensions, D-optimality is a good criterion for "diversity". In these cases, D-optimal designs generally have reasonably high S-scores, but S-optimal designs often have poor D-scores.

Table 8. Results from Designs of 18 Points in the 17-Dimensional Property Space[a]

method	size	D-score	D-σ	S-score	S-σ
100%	762	37.0	0.38	3.94	0.04
D-opt. w/ center pt	762	36.3		3.53	
50%	381	35.0	0.85	3.68	0.03
20%	152	30.2	0.94	3.41	0.10
S-opt. w/ center pt	762	27.9		3.92	
10%	76	26.7	1.11	3.19	0.11
5%	146	21.4	1.40	2.77	0.12
random (2.4%)	76	13.83	1.34	1.79	0.35

[a] Methods listed as a percent indicate a fraction of the candidate set chosen at random. D-σ and S-σ are standard deviations for five such random subsettings. "D-opt. w/ center pt" indicates the point nearest the centroid of property space was augmented with 17 additional points by D-optimal design—likewise for S-optimal design.

As an analogous test for the 17-dimensional space of the current study, D-optimal and S-optimal libraries of 18 points were generated, insisting on the point nearest the centroid as the sole fixed requirement in each design. There is no way to visualize how well 18 points fill 17 dimensions, so D-optimal and S-optimal calibration designs were also run. The results are compiled in Table 8. To visually compare the D-scores and S-scores in Figure 4a, they were both linearly scaled to set the score for a random design at the bottom of a diversity "yardstick" and the pure maximal diversity design at the top. The calibration points for randomly reduced candidate sets are indicated and have been connected with dotted lines to help visually align the two scales. The D-optimal design forced to include the center point had a D-score of 36.3, which calibration showed was only slightly below the maximum possible value for an 18-point design of 37.0 (with no center point requirement). The S-optimal design requiring the center point had an S-score of 3.92, again very comparable to the S-optimal maximum possible score of 3.94. The D-optimal design with the fixed center point had a respectable S-score of 3.5, roughly comparable to an S-optimal design run on 33% of the candidates. The S-optimal design with the fixed centroid, however, had a relatively poor D-score of 27.9, roughly equivalent to throwing away 88% of the data at random and

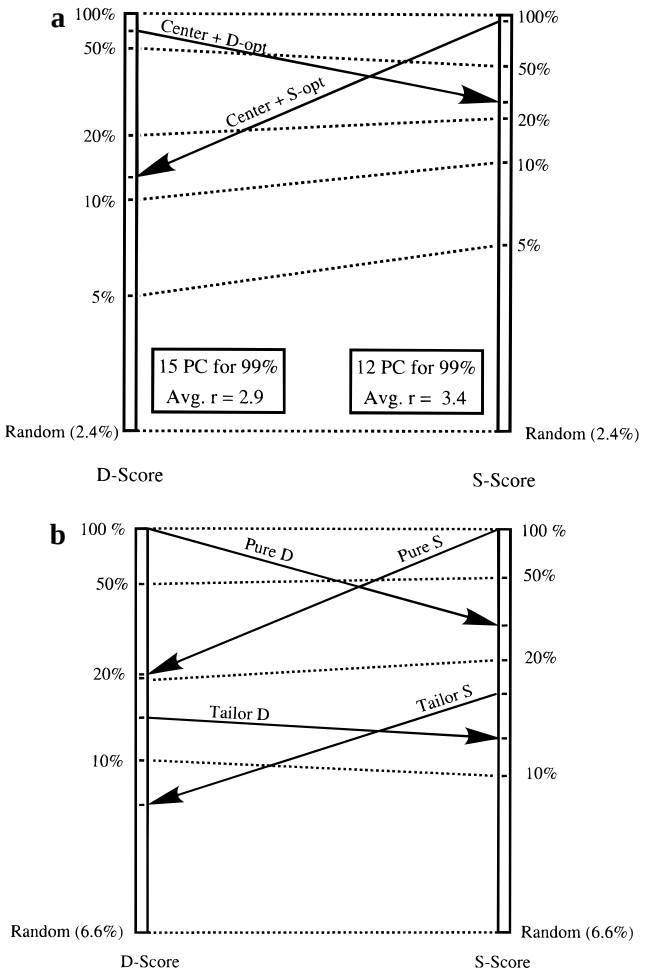

Figure 4. (a) 18-Point "yardsticks" of D-scores and S-scores. 100% refers to the maximally diverse designs. Other percentages refer to randomly removing all but that fraction of the candidates and determining the maximally diverse 18-point designs. The lowest value is a random selection of 18 points. The arrows compare the S-score of a D-optimal design and the D-score of an S-optimal design. (b) 50-Point "yardsticks" of D-scores and S-scores. 100% refers to the maximally diverse designs. Other percentages refer to randomly removing all but that fraction of the candidates and determining the maximally diverse 50 point designs. The lowest value is a random selection of 50 points. The upper arrows compare the S-score of a D-optimal design and the D-score of an S-optimal design. The lower arrows are similar comparisons for tailored designs.

performing D-optimal design on the remaining 12%. Nevertheless, it is still much better than random selection (keeping 2.4%), with a D-score of only 14.

Principal components analysis was performed to test the approximate dimensionality of the designs. For the D-optimal design, 15 PCs were required to cover 99% of the variance, so the 18 points could be said to sample about 15 of 17 possible dimensions. For the S-optimal design only 12 of the 17 PCs were required to cover 99% of the variance, so by this measure, it captured three fewer dimensions than D-optimal design. Finally, the average correlation coefficient between variables for D-optimal design was 0.29, but was 0.34 for the S-optimal design, showing greater collinearities in the S-optimal design. Hence, D-optimal design sacrificed a small amount of spread (redundancy) but did a better job of covering property space.

Figure 4b uses the calibration data from Table 4, scaled as above, to visually compare S-optimal and D-optimal based tailored designs of 50 points. The S-score for the D-optimal based tailored design was 2.3, very comparable to the score of 2.5 for the actual S-optimal based tailored design from the same bin profile. These values are comparable, respectively, to selecting 17% or 13% of the "all" set at random before performing a pure S-optimal design. However, the D-score for the S-optimal based tailored design was only 77.5, compared to 96 for the D-optimal based tailored library. As Table 4 shows, this is equivalent to eliminating all but 9% of the candidates vs 15%, respectively. A random design, with a score of 50, corresponds to 6.6%, so the D-score for the S-optimal based tailored design is approaching random. Hence, the D-optimal based tailored design has a decent S-score, but the S-optimal based tailored design has a poor D-score, even in this tailored library with almost three times as many points as the number of dimensions.

As the number of points exceeds the number of dimensions, D-optimal design will eventually suggest resampling some points. In the three-dimensional example above this happened at nine points. This indicates that more points were requested than are required to estimate a linear model, so higher order model terms could be added to the model. A useful rule of thumb is to add cross terms from the higher principal components until the S-score indicates that resampling has been prevented. An alternative approach is to use "Bayesian optimal design", which automatically adds some weight to all of the cross terms.[27] In practice, adding enough cross terms to saturate the model generally works well for tailored designs (see Methods above).

It should be reiterated that this analysis was specifically aimed at problems where part of the design is preselected and the number of points is not much larger than the number of dimensions. For other problems, such as selecting a subset from a corporate archive of hundreds of thousands of compounds in a space of only five or six dimensions, other methods would be preferred, such as sampling from cells in close packed lattices. An additional limitation of D-optimal design is that the scores can only be compared between libraries of the same size. S-optimal scores can be compared between libraries of different sizes so they are useful for initial studies of the appropriate number of substituents.

Evaluation of the Library. The goal of this library design example was to provide high structural diversity while constraining pertinent physicochemical properties to suitable ranges for small molecule drugs. To examine this, Figure 5a−d presents histograms and quantile box plots for three sets of substituents: the full set of 756 useful amines, the final tailored design of 50 compounds with a diversity score of 96 (see above), and the maximally diverse D-optimal design of 50 compounds with no bias at all which had a diversity score of 136 (see Table 4). Distributions are presented for four properties: molecular weight, calculated log *P*, number of rotatable bonds, and distance from the center of property space. Tables 9−12 give the corresponding quantiles, means, and numbers of observations. The boxes in the "box plots" indicate the 25, 50, and 75 percentiles. The diamonds depict the means and standard deviations.

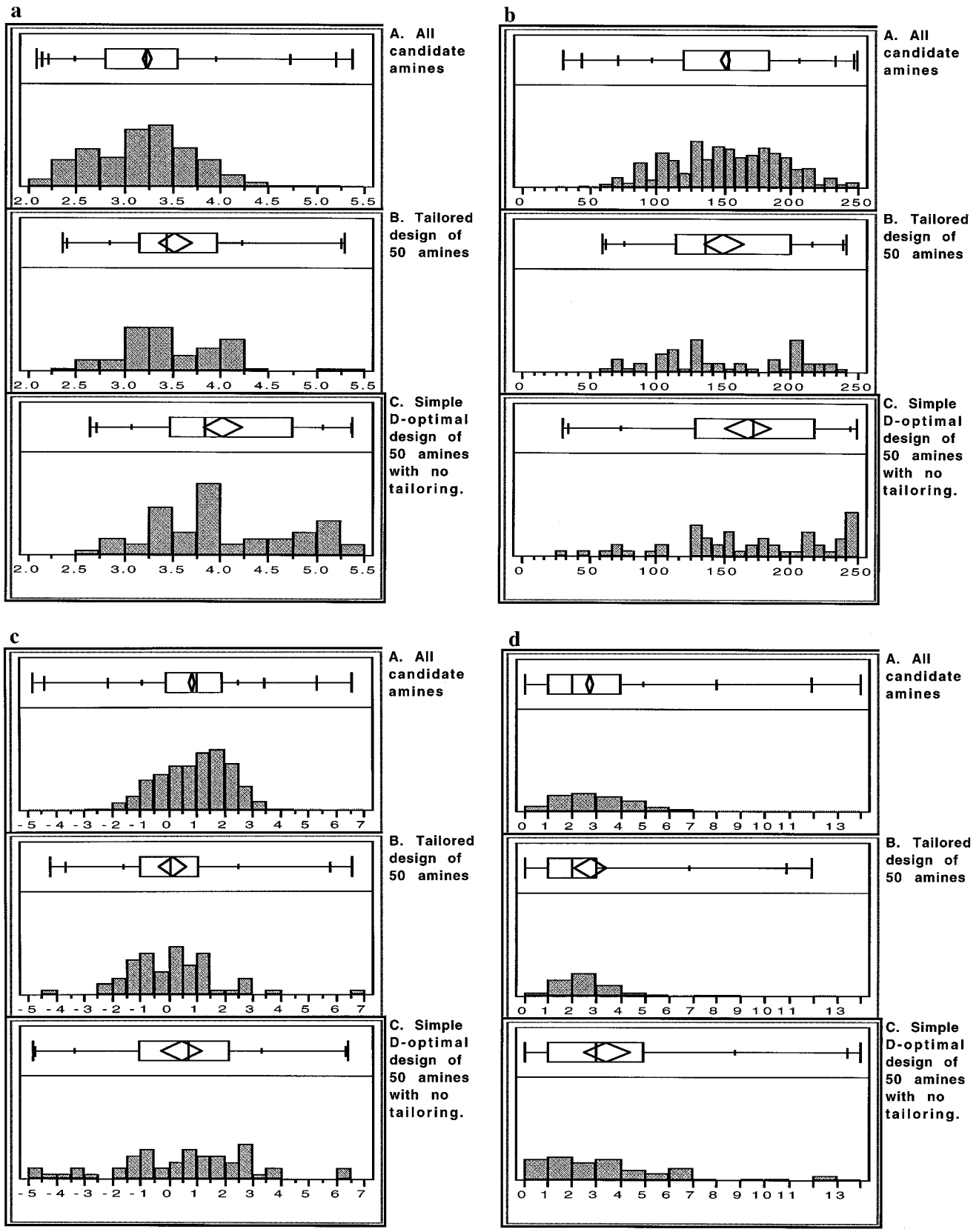

Figure 5. (a) Distances from center of property space for three designs of 50 amines each. Histograms and quantile box plots show that the tailored design does not favor the extremes of property space as does a simple D-optimal "diversity" design. (Quantiles are in Table 9.) (b) Molecular weights for three designs of 50 amines each. Histograms and quantile box plots show that the tailored design does not favor the extremes of molecular size as does a simple D-optimal "diversity" design. (Quantiles are in Table 10.) (c) Values of log Kow for three designs of 50 amines each. Histograms and quantile box plots show that the tailored design does not emphasize the extremes of high and low lipophilicity as does the simple D-optimal "diversity" design. (Quantiles are in Table 11.) (d) Number of rotatable bonds for three designs of 50 amines each. Histograms and quantile box plots show that the tailored design does not emphasize flexible substituents as does the simple D-optimal "diversity" design. (Quantiles are in Table 12.)

Table 9. Frequency Distributions of Distances from the Centroid of Property Space[a]

quantile, %	all	tailored	max div
100.0	5.38		
99.5	5.20	5.30	5.38
97.5	4.73	5.27	5.36
90.0	3.96	4.23	5.07
75.0	3.57	3.97	4.75
50.0	3.23	3.44	3.86
25.0	2.80	3.15	3.47
10.0	2.48	2.86	3.08
2.5	2.22	2.42	2.72
0.5	2.15	2.36	2.64
0.0	2.09		
mean	3.24	3.54	4.03
N	756	50	50

[a] Column "all" is the distribution for the original candidate set. Column "tailored" is the tailored design from this study. Column "max div" is the maximally diverse D-optimal design of 50 substituents from the "all" set.

Table 10. Molecular Weight Frequency Distributions

quantile, %	all	tailored	max div
100.0	249.36		
99.5	246.71	241.46	249.36
97.5	232.44	238.96	249.32
90.0	206.29	216.87	244.22
75.0	183.57	200.29	217.82
50.0	153.18	137.18	172.77
25.0	121.18	114.19	128.95
10.0	97.15	76.31	74.24
2.5	71.12	63.13	34.36
0.5	45.08	60.10	31.06
0.0	31.06		
mean	152	150	168
N	756	50	50

Table 11. CLOGP Frequency Distributions

quantile, %	all	tailored	max div
100.0	6.63		
99.5	5.36	6.63	6.45
97.5	3.45	5.79	6.38
90.0	2.54	2.52	3.33
75.0	1.92	1.08	2.20
50.0	1.04	0.10	0.73
25.0	−0.06	−0.99	−1.06
10.0	−0.92	−1.58	−3.31
2.5	−2.12	−3.63	−4.74
0.5	−4.38	−4.19	−4.76
0.0	−4.76		
mean	0.89	0.18	0.48
N	756	50	50

Additional tick marks are the other quantiles listed in Tables 8−11. The distributions of the full candidate set represent the expected distributions of the random sets, which had an average diversity score of only 50 (see Table 4).

Concern has been raised that pure D-optimal designs (as well as other pure diversity designs) sample only the "outer edges" of property space.[26] The radial distributions in Figure 5a and Table 9 show that, for better or worse, the pure D-optimal diversity set does indeed oversample the extremes of property space relative to the original distribution. The tailored design shows only a modest outward shift relative to the candidates. Apparently, the constraints of sampling from property bins counteracts the D-optimal algorithm's propensity to sample mainly remote regions of property space.

Table 12. Frequency Distributions for Counts of Rotatable Bonds

quantile, %	all	tailored	max div
100.0	14		
99.5	12	12	14
97.5	8	11	13.5
90.0	5	7	9
75.0	4	3	5
50.0	2	2	3
25.0	1	1	1
10.0	1	1	0
2.5	0	0	0
0.5	0	0	0
0.0	0		
mean	2.74	2.76	3.44
N	756	50	50

The histograms and quantile boxes of the three properties in Figure 5b−d show that the pure diversity set distributions are relatively broad and flat and include most of the highest and lowest values of each property as shown in Tables 10−12. Recall that CLOGP is actually one of the 19 dimensions of the property space. Molecular weight and number of rotatable bonds were not specifically included in the property space calculations, but they are indirectly included through correlations with topological indices, so extremes of property space might well imply extremes of these properties as well.

Examining the property histograms shows that the pure diversity design emphasized large flexible groups with either extremely high or extremely low lipophilicity. Orally available drugs tend to be small, rigid compounds with intermediate lipophilicity, so pure diversity designs bias libraries away from ideal drug properties. The tailored library's property distributions have wide tails but are not as flat and extreme as the pure diversity designs. This is understandable, since only a few members from the extreme bins were permitted in this design. It is more hydrophilic than the original distribution: including a few extreme values, but concentrating most of the members in the desirable moderately hydrophilic region. About 75% of the substituents in the tailored set have three or fewer rotatable bonds versus four in the original distribution and five in the pure diversity set, showing that tailoring has limited the fraction of flexible substituents. The median (50%) molecular weight in the tailored design is lower than the original distribution and much lower than the pure diversity design, but there is a curious bimodal distribution with peaks at about 130 and 200. The pure diversity design has an extremely top heavy molecular weight distribution, with the most frequent value in the histogram actually being the highest MW slice, which had a very low original frequency. Since the low molecular weight bin was strongly emphasized in the tailored profile, this suggests that structural diversity requires complexity, and complexity requires mass. Recall that the molecular weight cutoff for the low MW bin was 130. The diversity algorithm emphasizes the heaviest members available in each bin. The bimodal distriution is aliasing from the two discrete MW cutoffs: the peak near 130 from sampling the low MW bins and the other at the highest values from sampling the

bins with no specific MW restrictions. Due to the lack of appropriate controls, it is difficult to say whether tailored designs are producing more or fewer hits than the earlier maximum diversity libraries; it may well be fewer. It is qualitatively apparent, however, that the average hits from tailored libraries are more attractive leads in terms of polarity, rigidity, and MW.

Conclusion

The intended purpose of this tailored library was for broad screening. Such libraries are often referred to as "random libraries". This designation might appropriately indicate that the library is equally suited for any arbitrary target. However, this exercise showed that random selection of compounds for "random screening" was poor both in structural diversity and in distribution of physicochemical properties. More recently, such libraries have been referred to as "diversity libraries". Pure diversity designs, however, were found to be systematically biased toward heavy, flexible compounds with very high or very low lipophilicity. These properties are poorly suited to yield bioavailable drugs. Contrary to what these names suggest, designing optimal libraries for broad screening requires a combination of property calculations, structural diversity calculations, experience, and good medicinal chemical intuition. Stratified D-optimal sampling from bins provides machinery to combine all of these requirements.

If this much tailoring is useful even in broad screening designs, how much more can it be used in target-focused problems? Two principal appeals of this approach to tailoring property distributions are generality and simplicity. As long as a medicinal chemist can identify which compounds or fragments are likely to share a property, a bin can be added to influence its contribution to the design. The fragments that dock into a receptor model, problem substituents indicated by a metabolism model, groups with desirable physicochemical properties, or just intuitively favored fragments, can all be simultaneously managed by this method. Other, more algorithmic methods are possible. Borth et al. presented a technique to simultaneously optimize additional criteria, such as cost or synthetic difficulty along with the D-criterion for information content, such as for finding the most diverse design possible for a given price.[8] Although this approach is rigorous and highly automated, the weights given to various criteria are still subjective, and the intuitive "what if" interaction with the design is absent. Using multiple bin profiles, one works directly with the tradeoffs between diversity and other design criteria. Balancing these tradeoffs benefits from art, experience, and the clarity of "hands on" exploration more than from a complex objective function. It is close to the way practicing medicinal chemists and biologists think about drug discovery and thus facilitates interaction within a project team.

While this work criticizes pure diversity designs, it does not marginalize the need for good diversity computation methods. High diversity is essential for efficient screening libraries, and it is the privileged property that is optimized within the constraints of the bin profiles. However, it is only one of many important factors, some of which are difficult to quantify, and can only be recognized by the practiced eye of experienced medicinal chemists. Stratified sampling from bin profiles allows experienced drug discovery teams to design well-tailored libraries that are diverse but satisfy these additional, often more nebulous, factors as well.

Acknowledgment. The authors thank David Jan for very useful Daylight tool kit programs that count the number of occurrences of a large set of standard fragments and the number of rotatable bonds in a molecule. Thanks to Jeff Blaney and David Spellmeyer for many helpful discussions.

References and Notes

(1) 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.

(2) 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.

(3) Ng, S.; Warne, R. L.; Zuckermann, R. N.; Martin, E. J.; Simon, R. J. Peptoids as opioid receptor ligands. Patent No. US 5481020 A 960102, USA, 1996.

(4) Willett, P. In *Similarity and Clustering in Chemical Information Systems*; John Wiley & Sons: New York, 1987; p 54.

(5) Dillon, W. R.; Goldstein, M. *Multivariate Analysis: Methods and Applications*; John Wiley & Sons: New York, 1984.

(6) Spellmeyer, D. *Unpublished results*, 1996.

(7) Unger, S. H. Consequences of the Hansch Paradigm for the Pharmaceutical Industry. In *Drug Design*; Ariens, E. J., Ed.; Academic Press: New York, 1980; Vol. 9, pp 47−119.

(8) Borth, D. M.; McKay, R. J.; Elliott, J. R. A Difficulty Information Approach to Substituent Selection in QSAR Studies. *Technometrics* **1985**, *27*, 25−35.

(9) Austel, V. Experimental Design in Synthesis Planning and Structure−Property Correlations. *Methods Princ. Med. Chem.* **1995**, *2*, 49−62.

(10) Sjoestroem, M.; Eriksson, L. Applications of Statistical Experimental Design and PLS Modeling in QSAR. *Methods Princ. Med. Chem.* **1995**, *2*, 63−90.

(11) Brannigan, L. H.; Grieshaber, M. V.; Schnur, D. M. Use Experimental Design in Organic Synthesis. *CHEMTECH* **1995**, *25*, 225−35.

(12) Scott, P. J.; Penlidis, A.; Rempel, G. L. Ethylene-vinyl acetate semi-batch emulsion copolymerization: experimental design and preliminary screening experiments. *J. Polym. Sci.* **1993**, *31*, 403−26.

(13) Weiser, M. W.; Fong, K. B. Experimental Design for Improved Ceramic Processing, Emphasizing the Taguchi Method. *Am. Ceram. Soc. Bull.* **1993**, *72*, 87−92.

(14) Kettaneh-Wold, N. Use of Experimental Design in the Pharmaceutical Industry. *J. Pharm. Biomed. Anal.* **1991**, *9*, 605−10.

(15) Shemetulskis, N. E.; Dunbar, J. B.; Dunbar, B. W.; Moreland, D. W.; Humblet, C. Enhancing the Diversity of a Corporate Database Using Chemical Database Clustering and Analysis. *J-CAMD* **1995**, *9*, 407−16.

(16) Martin, E.; Blaney, J.; Siani, M.; Spellmeyer, D. *Measuring Diversity: Experimental Design of Combinatorial Libraries for Drug Discovery*. 209th National Meeting of the American Chemical Society, Anaheim, CA, 1995.

(17) Weininger, D. *Thor*; Daylight Chemical Information Systems: Irvine, CA, 1993.

(18) Weininger, D. *Merlin*; Daylight Chemical Information Systems: Irvine, CA, 1993.

(19) Miller, A.; Nguyen, N.-K. A Fedorov Exchange Algorithm of D-optimal Design. *Appl. Stat.* **1994**, *43*, 669−678.

(20) Miller, A.; Nguyen, N.-K. *DOPT*, http://lib.stat.cmu.edu/, 1996.

(21) *Available Chemicals Directory*; Molecular Design Limited: San Leandro, CA, 1993.

(22) *SAS/QC Software: Usage and Reference, Version 6*; SAS Institute Inc.: Cary, NC, 1995.

(23) Conradi, R. A.; Hilgers, A. R.; Ho, N. F. H.; Burton, P. S. The influence of peptide structure on transport across Caco-2 cells. II. Peptide bond modification which results in improved permeability. *Pharm. Res.* **1992**, *9*, 435−9.
(24) Patterson, D. E.; Cramer, R. D.; Ferguson, A. M.; Clark, R. D.; Weinberger, L. E. Neighborhood Behavior: A Useful Concept for Validation of "Molecular Diversity" Descriptors. *J. Med. Chem.* **1996**, *39*, 3049−3059.
(25) Federov, V. V. *Theory of Optimal Experiments*; Academic Press: New York, 1972.
(26) Hassan, M.; Bielawski, J. P.; Hempel, J. C.; Waldman, M. Optimization and visualization of molecular diversity of combinatorial libraries. *Mol. Diversity* **1996**, *2*, 64−74.
(27) DuMouchel, W.; Jones, B. A Simple Bayesian Modification of D-Optimal Designs to Reduce Dependence on an Assumed Model. *Technometrics* **1994**, *36*, 37−47.

CC9800024