

Balancing focused combinatorial libraries based on multiple GPCR ligands

Farhad Soltanshahi · Tamsin E. Mansley ·
Sun Choi · Robert D. Clark

Received: 26 June 2006 / Accepted: 29 August 2006 / Published online: 13 October 2006
© Springer Science+Business Media B.V. 2006

Abstract G-Protein coupled receptors (GPCRs) are important targets for drug discovery, and combinatorial chemistry is an important tool for pharmaceutical development. The absence of detailed structural information, however, limits the kinds of combinatorial design techniques that can be applied to GPCR targets. This is particularly problematic given the current emphasis on focused combinatorial libraries. By linking an incremental construction method (OptDesign) to the very fast shape-matching capability of ChemSpace, we have created an efficient method for designing targeted sublibraries that are topomerically similar to known actives. Multi-objective scoring allows consideration of multiple queries (actives) simultaneously. This can lead to a distribution of products skewed towards one particular query structure, however, particularly when the ligands of interest are quite dissimilar to one another. A novel pivoting technique is described which makes it possible to generate promising designs even under those circumstances. The approach is illustrated by application to some serotonergic agonists and chemokine antagonists.

Keywords ChemSpace · Combinatorial library · Library design · Multi-objective scoring · OptDesign · Pareto ranking

F. Soltanshahi · T. E. Mansley · S. Choi · R. D. Clark (✉)
Informatics Research Center, Tripos, Inc., 1699 S. Hanley
Road, St Louis, MO 63144, USA
e-mail: bclark@tripos.com

Present Address:

S. Choi
College of Pharmacy and Division of Molecular Life
Sciences, Ewha Womans University, Seoul 120-750,
South Korea

Abbreviations

5HT _{1F}	5-Hydroxytryptamine (serotonin) receptor, subtype 1F
Boc	<i>tert</i> -Butoxycarbonyl
CCR1	Chemotactic cytokine receptor 1
GPCR	G-protein coupled receptor

Introduction

The G-protein coupled receptor (GPCR) superfamily is the largest class of transmembrane receptors. They are characteristically comprised of seven membrane-spanning α -helices, which are joined and stabilized by intra- and extra-cellular loop regions. Many natural ligands contain a basic amine, which is often involved in binding and agonistic effect. GPCRs are typically involved with signaling pathways, converting an extracellular signal represented by ligand binding into an intracellular signal in the form of G-protein activation. Such signaling is known to be involved in many critical physiological pathways, and these receptors are targeted by 40–50% of the new drugs developed in recent years [1].

There is considerable structural variation among the receptor families, with the transmembrane helices being more conserved within and between families than are the intervening loop regions. Few crystal structures of GPCRs are available, so scientists interested in applying structure-based techniques must resort to constructing homology models. Such models were originally constructed based on the structure of bacteriorhodopsin, which had been characterized by electron microscopy and, later, by X-ray

crystallography. The mammalian bovine rhodopsin has been the preferred template since its crystal structure was determined in 2000 [2]; it is the only mammalian GPCR whose crystal structure has been published to date.

Unfortunately, the structural homology of rhodopsin to most GPCRs is still less than ideal for development of homology models with the degree of confidence needed for docking studies. Hence focused combinatorial design methods that depend upon docking and scoring for product selection [3] are not applicable in this case. This makes ligand-based drug design particularly important for GPCR targets, using known agonists and antagonists to characterize the binding pocket and identify features involved in ligand binding and receptor activation.

We used an incremental construction method (OptDesign [4, 5]) to design focused combinatorial libraries targeted to two GPCRs—the serotonergic receptor 5HT_{1F} and the chemokine receptor CCR1. The program is an extension of optimizable *K*-dissimilarity selection (OptiSim [6–8]) that generates product-based designs, in that reagents are chosen at each step based on the properties of their virtual products. A small random sample of qualified candidate reagents is considered at each step, and the one that yields the best virtual products is selected for inclusion in the library. Only a fraction of the possible products need be considered, which makes the method very efficient as well as making it practical to calculate quite complex product properties without consuming the large amounts of CPU time that would be required to carry out such calculations on all possible products.

For diverse libraries, the goal is to sample the population space so as to produce libraries that are representative of the full combinatorial population as well as being structurally diverse. This is achieved by defining the “best” candidate as the one whose products are most dissimilar to those products selected for inclusion in the library in previous steps [9]. For the focused design work described here, we have introduced a multi-objective scoring function based on Pareto ranking [10] that allows simultaneous consideration of similarity to queries based on multiple ligands. “Similarity” in this case is based on topomer distances [11], a measure of shape similarity that has proven useful for identifying individual products of interest in very large virtual libraries in prospective studies [12].

Both targets considered here are pharmaceutically relevant and timely. 5HT_{1F} is located in the CNS where it is thought to play a role as a serotonin autoreceptor. Agonists of 5HT_{1F} (e.g., LY334370 (**1**)) are effective

against migraine [13–15]. CCR1, on the other hand, is involved in immune responses and has been implicated in inflammatory diseases, such as asthma and allergy, psoriasis, multiple sclerosis, rheumatism, arthritis and inflammatory bowel disease [16]. Antagonists of CCR1 are therefore of pharmacological interest for their anti-inflammatory activity. Structures of numerous drug candidates targeting it have been published and patented, and several are in some phase of discovery and clinical research, although none are yet approved drugs.

Though we only describe applications to these two GPCRs here, the method should be equally applicable to other targets where crystal structures or extremely robust homology models are not available.

Methods

The combinatorial constraint involved in full-matrix library designs can unduly restrict the range of accessible products and force the designer to sharply limit the range of reagents used. Hence OptDesign supports generation of either full or sparse matrix designs in either single or multi-block modes [5, 9]. A sparse design permits products to be skipped, allowing “holes” in the design. If reagents X_i and Y_j each contribute 25 heavy atoms to a product, for example, both can be included in a sparse design without having to violate a constraint that no product consist of more than 40 heavy atoms; that *particular* product simply doesn’t appear in the library. Allowing design densities to fall below 1.0 increases the structural diversity of the design produced.

OptDesign ordinarily uses a *one-by-one* pivoting scheme wherein reagent selection alternates between reagent classes at each step, adding one of *K* qualified X_i reagents, then one Y_i , then X_{i+1} (or Z_i , and then X_{i+1} , etc). A block is complete once the number of products in the (sub)design meets or exceeds the number requested, or all reagent quotas have been met, or no viable candidate reagents remain to choose from, whichever comes first [9].

Slice pivoting strategy

Sublibraries of practical interest are usually not square, however, so one must allow for the fact that more reagents may be desired from population **X** than from population **Y**. For combinatorial designs that involve a di-substituted scaffold, for example, intermediates generated in the first reaction step are often synthesized and purified in bulk (i.e., on a multi-gram scale),

then parallel synthesis and high-throughput micro-scale chromatography are used to synthesize and purify products obtained from secondary reactions [17]. Hence cost considerations lead to unsymmetrical designs in which m (the number of primary reagents required) is considerably smaller than m' (the number of secondary reagents). Which primary reagent is selected at each step is influenced by the selection of secondary reagents from previous iterations, since those dictate which products are considered.

The simplest way for OptDesign to handle this situation is to pivot evenly between X and Y until a square $m \times m$ sublibrary is in hand, then stop pivoting and simply add Y_{m+1} , Y_{m+2} and so forth until the desired $m \times m'$ design is complete [9]. This strategy generally works well, but it is not usually optimal. Early on in the design process, it is often wiser to select reagents more frequently from Y so that each subsequent candidate X is judged against reaction with a bigger set of Y s. Consider, for example, a 3×9 matrix design. If a one-by-one pivoting strategy is used, only Y_1 , Y_2 and Y_3 affect the selection of the three X s, thereby disproportionately influencing the design as a whole. It is more reasonable to balance the frequencies of selection, picking three Y s for every X chosen. Doing so spreads influence across more secondary reagents, thereby reducing the chance that picking a “bad” Y will unduly restrict the scope of the design. This alternative *slice pivoting* scheme is illustrated in Fig. 1.

Library construction and product filtering

Virtual libraries were generated by entering a scaffold and then defining the types of reagents “reacting” at each variation site. The core structure (scaffold) for

each target library was created separately, but the initial lists of commercially available reagents defining the extent of the full combinatorial libraries were shared. These lists were drawn from ChemSpace [18], which is a discovery research platform developed at Tripos for building, managing, filtering and searching sets of large combinatorial libraries [12, 19, 20].

Virtual libraries built in ChemSpace are searched as combinatorials—i.e., without enumeration. In particular, 3D searches are carried out based on topomeric distances [11]. These are obtained by cleaving the query structure at all combinations of exocyclic single bonds that yield two or three fragments, depending on the complexity of the library being searched. The various alternative core and substituent substructures constitute subqueries that are standardized, put into a characteristic conformation and aligned to a reference lattice. The molecular field for each core and substituent generated from the query is then calculated and compared to the core and synthon fields for libraries stored in ChemSpace [21] (Fig. 2). Distances are computed from the squared field differences across the lattice, summed across the cores and all substituents. The piecemeal distances are relatively large in most cases, so great swaths of the product space can quickly be excluded from further consideration: if the difference between a core subquery and a library core is larger than the designated search radius r_p , there is no reason to consider any product from the corresponding library for that particular query fragmentation pattern [19, 20].

In practice, topomer searching is so fast that it is usually applied as the first step in an analysis. Products that “hit” are then filtered for properties on the basis of physical properties such as ClogP [22];

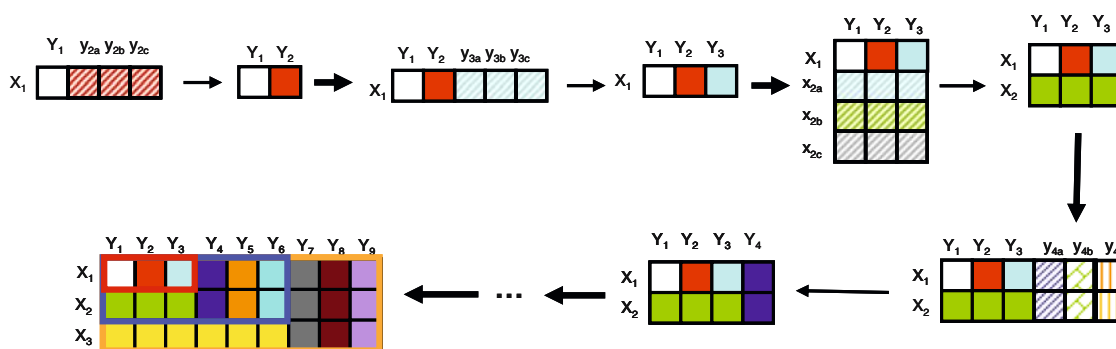
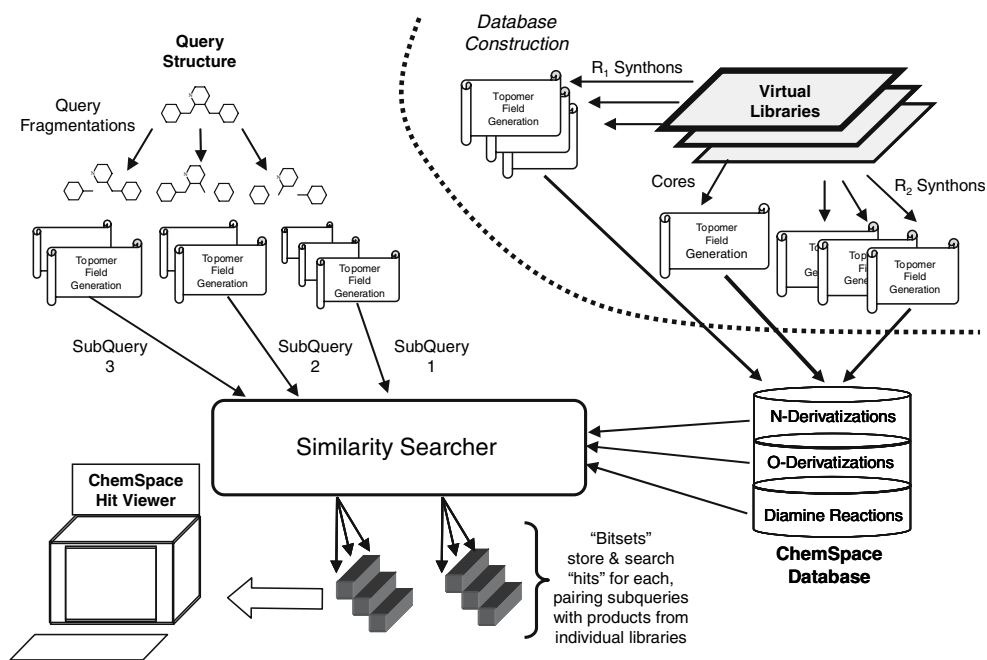


Fig. 1 Slice pivoting scheme for designing a 3×9 library with a reagent subsample size of $K = 3$. Uppercase letters indicate selected reagents, whereas lowercase letters indicate candidate reagents. The 1×3 block outlined in red is generated first. This is then expanded first into a 2×6 sublibrary outlined in blue and

finally into the full 3×9 pattern outlined in orange. The particular design is characterized in terms of the block dimensions at the end of each stage. Hence the procedure shown here corresponds to a 1×3 ; 2×6 ; 3×9 design

Fig. 2 How ChemSpace handles and searches virtual libraries [19]



hydrogen bond donor and acceptor atom counts; and molecular weight. For the GPCR studies described here, we searched the full virtual libraries using topomer queries derived from sets of ligands known to be active against 5HT_{1F} or CCR1. The products identified were then filtered for drug-likeness and “Rule of Five” compliance [23]. Products having more than eight rotatable bonds were also removed, as were products with more than one chiral center; the latter produce diastereomeric mixtures that may be difficult to purify

Reagent filtering

There is no point in considering any reagent that yields no product satisfying all constraints. Hence combinatorial sublibraries were defined by extracting the relevant synthons from the filtered topomer “hits” for the target libraries. These lists of reagents were further trimmed by applying substructural filters to remove reagents that would introduce alkylating or other potentially toxicophoric groups into the products. Reagents containing a nitro group, for example, were dropped. The substitution reactions involved alkylation and acylation reactions, so compatibility filters were applied to remove reagents containing nucleophilic centers (e.g., $-\text{NH}_2$, $-\text{OH}$, $-\text{SH}$) and extraneous electrophilic centers (e.g., $-\text{CO}_2\text{H}$, $-\text{COCl}$, $-\text{SO}_2\text{Cl}$, $-\text{N}=\text{C}=\text{O}$, $-\text{N}=\text{C}=\text{S}$, halides). These filters operate through substructure searches based on the SYBYL line notation (SLN [24]).

Multi-objective scoring

OptDesign operates iteratively, selecting the best candidate from a qualified subsample of K reagents at each step. It has two competing objectives when used to generate diverse libraries. *Representativeness* is conferred by drawing the subsamples randomly at each iteration. *Diversity*, on the other hand, is conferred by the scoring function, which defines the “best” candidate reagent in each step’s subsample as being the one that yields products most distinct from those included in the design during previous iterations [19]. The experiments described here make use of a new Pareto ranking scheme designed to favor candidates that will yield products similar in shape to several query ligands. Similarity was maximized by choosing reagents that minimized the topomer distances between queries and products.

Figure 3 illustrates how Pareto ranking guides reagent selection when the goal is to optimize similarity to multiple queries. A sparse matrix design is shown with a minimum stepwise product density of 0.50 and a subsample size $K = 3$. X_1 , X_2 and X_3 have already been selected, as have Y_1 and Y_2 . Four of the six possible combinations have been included in the subdesign. The figure illustrates how the program picks Y_3 from among three candidates— y_{3a} , y_{3b} and y_{3c} . Each candidate y can produce up to three products, but the minimum density of 0.50 means that each candidate needs to contribute two valid products to the design [25]. The goal is to maximize the shape similarity to the queries

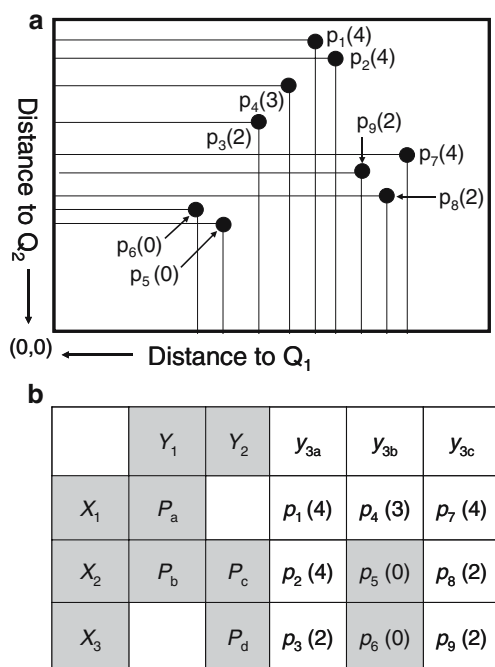


Fig. 3 Scoring based on Pareto ranking. **(a)** Plot relating the Pareto ranks (shown in parentheses) to the degree of dominance—i.e., the number of products better by both criteria than the one being ranked. **(b)** Sparse block subdesign resulting from incorporation of the new products having the lowest maximum Pareto rank. Pareto ranks are shown in parentheses. Selected products are shaded in gray

Q_1 and Q_2 by minimizing the corresponding topomer distances, which are shown in Table 1.

The Pareto rank of each candidate product is defined as the number of products that dominate it [10]. In Fig. 3, the non-dominated products p_5 and p_6 are given a rank of zero, because there are no products that are more similar (closer) to *both* queries. Product p_7 , on the other hand, receives a rank of four because it is dominated by p_5 , p_6 , p_8 and p_9 . Given the target density

of 0.50, each candidate reagent y needs to contribute only two products. Hence to rank each reagent candidate y we take the *best* two product Pareto rankings and use the *worst* of that pair to rank the reagent. In this example y_{3b} is the best candidate since its two best products both have a Pareto rank of zero. Since p_5 and p_6 have lower Pareto ranks than p_4 , they are the products that get incorporated into the growing library.

Ties in Pareto rank were resolved firstly by favoring reagents whose products “hit” more targets and, secondarily, by favoring those reagents whose products lie closer to the target molecules in terms of topomer distance.

Note that here, each candidate reagent is represented by an *ensemble* of points in the Pareto space, one for each expected product. This is fundamentally different from methods in which there is a one-to-one correspondence between candidates and points in the Pareto space [10].

Analysis details

OptDesign was run with several additional constraints:

- The target design consisted of a single block with a density greater than 0.50.
- No reagent x_{ij} (or y_{ij}) was included in the respective reagent subsample if it was too similar in terms of its substructural fingerprint to any reagent already included in the design—i.e., for any X_j (or Y_j) for $j < i$. The exclusion radii r_x and r_y were expressed as a maximum allowed Tanimoto similarity.
- Only products that “hit” at least two of the four queries were considered valid. The product inclusion radius specified in each case r_p took both steric and feature differences into account.

Table 1 Product correspondences and topomer distances for the Pareto rankings used in Fig. 3

Secondary candidate	Primary reagent	Candidate product	Distance ^a to Q_1	Distance ^a to Q_2	Pareto rank	MiniMax ^b Pareto
y_{3a}	X_1	p_1	251	272	4^b	4
	X_2	p_2	254	269	4	
	X_3	p_3	239	260	2	
y_{3b}	X_1	p_4	246	266	3	0
	X_2	p_5	230	251	0	
	X_3	p_6	225	253	0	
y_{3c}	X_1	p_7	261	256	4	2
	X_2	p_8	259	254	2	
	X_3	p_9	256	255	2	

^a Topomer distances [11] were calculated on a 2 Å lattice

^b Smallest maximum Pareto score across all possible *pairs* of products from each reagent y at the given density (0.50). The corresponding best product pairs are highlighted in boldface type

Results

Focused 5HT_{1F} designs

The virtual reaction scheme behind the 5HT_{1F} library is shown in Fig. 4. Pyrrole carboxylates are subject to elaboration into a protected pyrroloazepinone by reaction with benzylamine-*N,N*-diacetate. The intermediate diester can be hydrolyzed and decarboxylated in acid [26], followed by catalytic hydrogenation to remove the benzylic blocking group. Addition of a Boc protecting group allows alkylation to proceed selectively at the pyrrole nitrogen, and subsequent acid-catalyzed deprotection clears the way for acylation of the nitrogen in the azepinone ring.

The full virtual library was created in ChemSpace and searched against a set of known 5HT_{1F} receptor agonists [27, 28] to identify topomerically similar structures. Due to the lack of structural diversity among known potent agonists, those chosen as queries are quite structurally similar to one another, and the “hit” lists obtained overlapped to a large degree. Note, however, that the queries are very different from the pyrroloazepinone scaffold itself. Figure 5 lays out the structures of the queries, and Table 2 indicates the number of “hits” found for each query and how many survived subsequent filtering steps.

Fig. 4 Synthetic scheme upon which the 5HT_{1F} library design was based

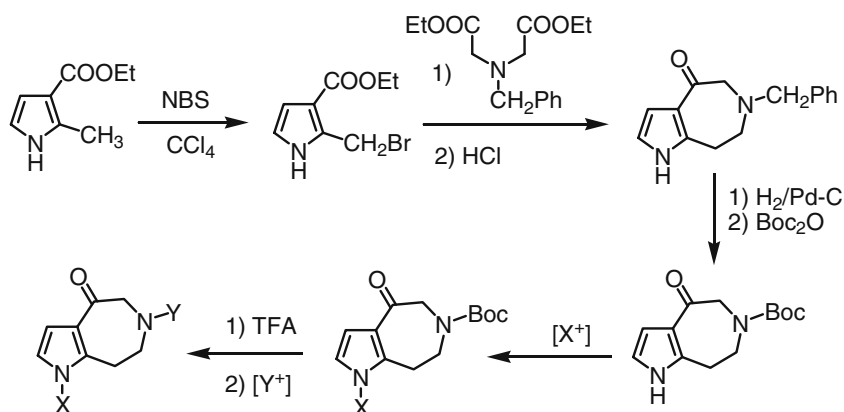
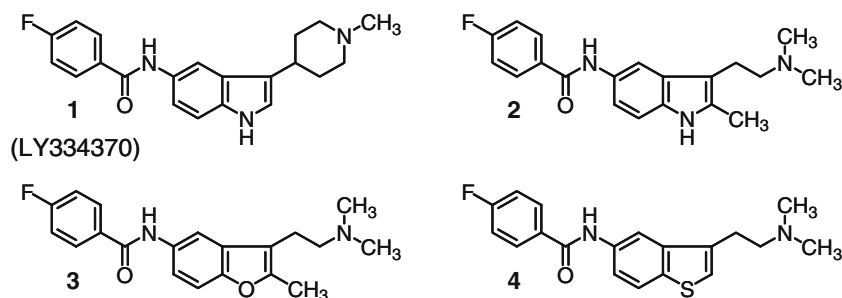


Fig. 5 5HT_{1F} agonists used as queries in ChemSpace searches and sublibrary design



The reagent exclusion thresholds were set to $r_x = 0.90$ and $r_y = 0.90$, and the maximum topomer distance r_p was set to 270. Two different 30×100 designs were run: one using one-by-one pivoting and a second using a slice pivoting scheme— 1×10 ; 2×20 ; 3×30 ; 4×40 ; 5×50 ; 6×60 ; 7×70 ; 8×80 ; 9×90 ; 30×100 . The goal was to try to design sublibraries made up predominantly of products similar to all four queries. Table 3 shows the pairwise and all-way overlaps among the hitlists obtained.

Both approaches produced a sublibrary where the products selected had good search scores across all four queries. This is mainly due to the fact that the topomer hitlists overlap so substantially. Changing the pivoting scheme did not have any appreciable effect on the outcome. This can also be seen by examination of Fig. 6, which shows the distribution of product similarities to each query in the form of a bar chart.

Focused CCR1 designs

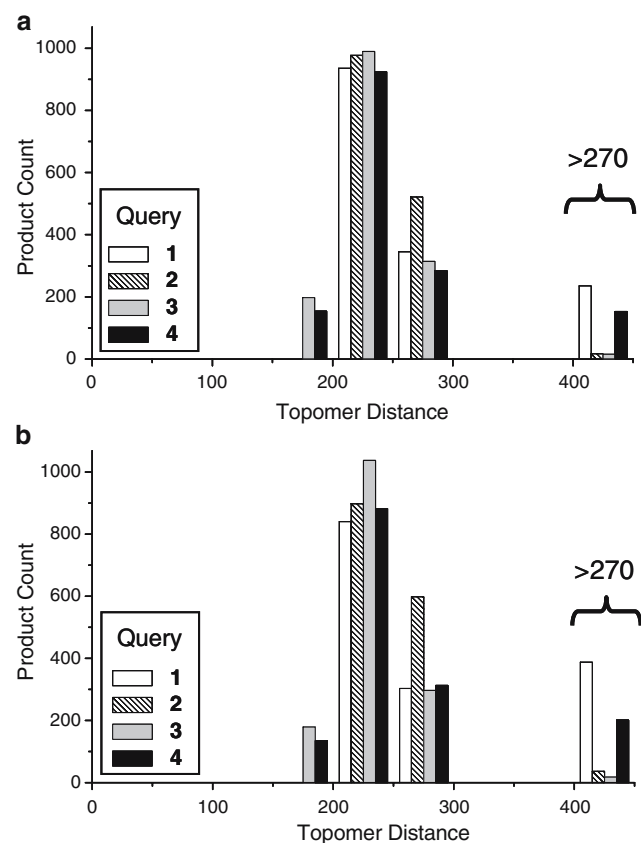
The virtual reaction scheme used to define the full CCR1 library was based on the commercially available 4-aminopiperazine (Fig. 7). Note that the order in which the two classes of electrophile are applied is switched from that described above for the 5HT_{1F} library. The Boc protected starting material is acylated

Table 2 Topomeric product similarity to individual queries across the full 5HT_{1F} virtual library and effect of filtering steps

Query structure	Filtering step			
	Similarity to query ($r_p = 270$)	Drug-likeness	Reaction compatibility	Physical properties
1	1,733,068	1,389,218 (80%)	585,515 (34%)	264,101 (15%)
2	3,493,541	2,361,822 (68%)	804,864 (23%)	544,487 (16%)
3	3,359,834	2,048,031 (61%)	675,289 (20%)	499,994 (15%)
4	1,892,252	1,271,493 (67%)	470,729 (25%)	349,996 (18%)

Table 3 Relationships among individual 5HT_{1F} topomer hitlists

	FullLibrary ^a	One-by-one pivot design ^b	Slice pivot design ^b
Full Design	554,579 (4382 × 8186)	1515 (30 × 100)	1531 (30 × 100)
1 ∩ 2	109,519	1,265	1,125
1 ∩ 3	97,617	1,131	978
1 ∩ 4	96,351	1,268	1,133
2 ∩ 3	412,932	1,362	1,330
2 ∩ 4	291,399	1,483	1,477
3 ∩ 4	274,103	1,348	1,299
1 ∩ 2 ∩ 3 ∩ 4	86,554	1,122	973

^a Number of products that pass all filters^b Number of products included in the design at a density of 50%**Fig. 6** Distribution of topomer distances to query structures for products included in focused 5HT_{1F} sublibraries. (a) One-by-one pivot design. (b) Slice pivot design

at the secondary nitrogen, then the distal primary amino group is deprotected and alkylated.

The full virtual library was created in ChemSpace and searched against the set of antagonists [29–33] shown in Fig. 8. These query molecules are much more structurally diverse than were the 5HT_{1F} agonists, and the intersections of their topomer hitlists were much more sparsely populated as a result. The searches were carried out taking both shape and feature similarity into consideration. The results from the individual searches were filtered for drug likeness, reaction compatibility and physical properties. The results are shown in Table 4.

Searching was done at an r_p of 300 topomer units and the maximum pairwise similarity allowed between reagents were set to $r_x = 0.95$ and $r_y = 0.95$.

The pooled filtered results were submitted to Opt-Design, specifying creation of 20 × 50 sublibraries at a density in excess of 50%. Two different designs were generated. Reagent pivoting was carried out using either “classic” one-by-one pivoting or using the slice pivoting scheme defined by: 1 × 5; 2 × 10; 3 × 15; 4 × 20; 5 × 25; 6 × 30; 7 × 35; 8 × 40; 9 × 45; 20 × 50.

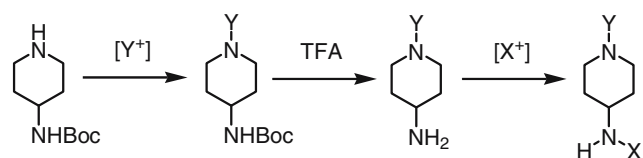
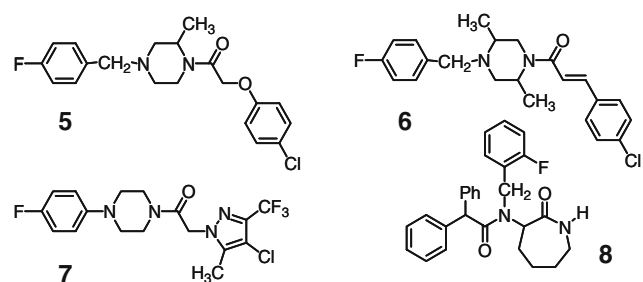
**Fig. 7** Synthetic scheme upon which the CCR1 library design was based**Fig. 8** CCR1 antagonists used as queries for ChemSpace searches and sublibrary design

Table 4 Topomeric product similarity to individual queries across the full CCR1 virtual library and effect of filtering steps

Query structure	Filtering step			
	Similarity to query ($r_p = 300$)	Drug-likeness	Reaction compatibility	Physical properties
5	3,355,771	2,561,582 (76%)	1,647,359 (49%)	795,908 (24%)
6	4,707,801	3,442,682 (73%)	1,934,608 (41%)	949,910 (20%)
7	163,217	133,444 (82%)	82,788 (51%)	24,849 (15%)
8	502,316	372,897 (74%)	137,886 (27%)	51,262 (10%)

One-by-one pivoting yielded a very skewed distribution of products (Table 5 and Fig. 9). A very similar distribution was seen in the design when a different random number seed was used, so it is not an accident of the “greedy” nature of the design algorithm. Changing to a slice pivoting scheme allowed the program to guide the design towards a solution where more products are topomerically similar to at least two of the queries. These results are presented graphically in Fig. 9 and are shown schematically in Fig. 10. Note that the increase in the number of products similar to 8 is most evident in the reduction in the number of hits falling beyond r_p . Again, similar results were obtained when a different random number seed was used.

Discussion

Combinatorial chemistry has become a major force in drug discovery and development, with attention in recent years shifting from generalized libraries [17] to ones focused on particular target proteins [3]. Docking and scoring against the target is a viable approach when enough structural information is available, but this is generally not the case for GPCRs such as serotonergic and chemokine receptors. Here we have described how coupling the incremental construction approach used in OptDesign to a rapid means for assessing shape similarity can provide an alternative,

Table 5 Relationships among individual CCR1 topomer hitlists

	FullLibrary ^a	One-by-one pivot design ^b	Slice pivot design ^b
Full design	1,701,332 (2631 × 6700)	601 (14 × 50)	470 (13 × 50)
5 ∩ 6	23,420	555	224
5 ∩ 7	5,028	0	239
5 ∩ 8	695	0	8
6 ∩ 7	0	0	0
6 ∩ 8	0	0	0
7 ∩ 8	3,226	46	9
5 ∩ 7 ∩ 8	51	0	5

^a Number of products that pass all filters

^b Number of products included in the design at a density of 0.50

ligand-based strategy for designing focused sublibraries that target specific GPCRs.

A library focused on any single ligand is likely to be overly specific, so it is generally desirable to incorporate multiple reference ligands (queries) into a design. A direct way to accomplish this is by using a weighted average of the similarities to each individual query structure. The appropriate weights to use can be very dependent on details of the distribution of the products of interest, however, and it is hard to know how the weights should be set a priori. An alternative, less direct approach is to extract a consensus query such as a

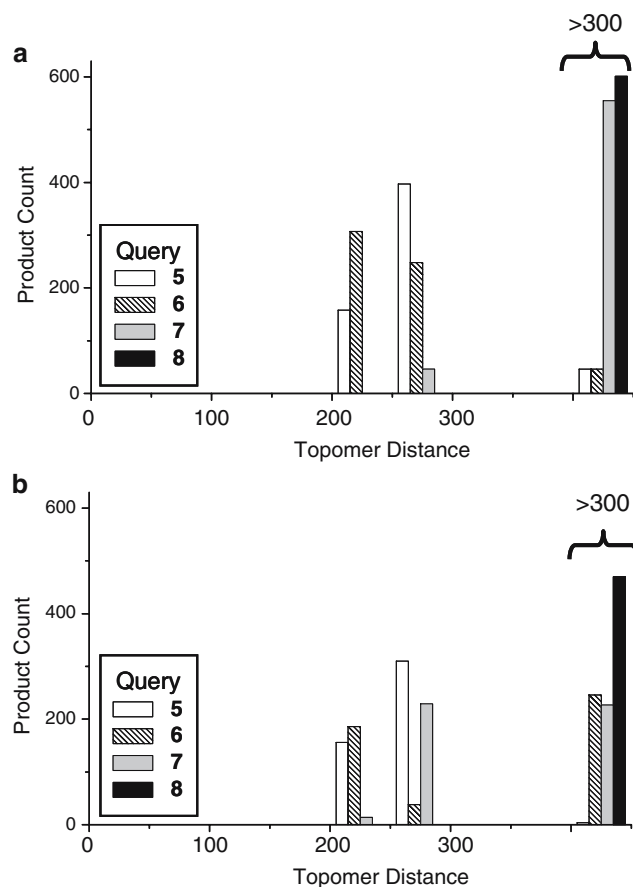


Fig. 9 Distribution of topomer distances to individual query structures for products included in focused 5HT_{1F} sublibraries. (a) One-by-one pivot design. (b) Slice pivot design

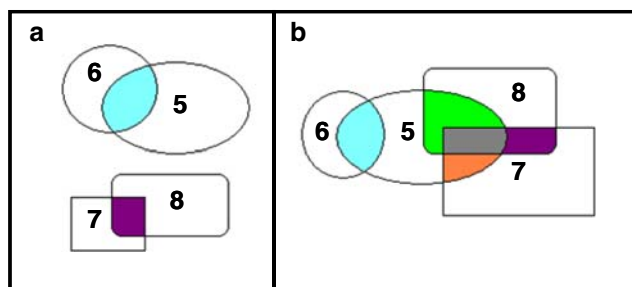


Fig. 10 Schematic representation of the distribution of products in the CCR1 sublibraries. Only products within the critical topomer distance $r_p = 300$ of *two* queries were allowed into the sublibrary. The white areas represent otherwise valid products that only hit one query. Areas do not relate directly to the number of products in each group or in the intersections between them. (a) One-by-one pivot design. (b) Slice pivot design

pharmacophore. Unfortunately, this strategy will focus primarily on products falling in the midst of *all* queries, and may miss important candidates that are similar to some—but not all—of them.

Introduction of a multi-objective scoring function makes it possible to optimize against multiple ligands simultaneously while specifying a minimal number of parameters. In the multiple objective genetic algorithm (MOGA) approach, each library is scored as a whole [10]. In contrast, the Pareto scoring scheme used here scores each candidate product separately; this makes the method much more suitable for generating sparse and multi-block designs.

OptDesign is a stochastic method. Indeed, that is key to the representativeness of the designs it produces. It follows that if the valid product space is very sparse—if, for example, there are too few products in the target library that are sufficiently similar to the queries provided—it will usually be difficult to build a good library. In particular, it is easy to pick starting points that lead to premature termination even under very loose density constraints. Worse, “classic” one-by-one pivoting will often produce very unbalanced designs wherein most products are similar to a single query.

The CCR1 library is a case in point. It is probably possible to obtain a useful library by looking at many runs using different random number seeds, but that is not a very efficient strategy. Instead, the balance in designs created from such sparse libraries was improved substantially by using slice pivoting in place of OptDesign’s standard one-by-one pivoting technique, evidently because doing so leads to a more equitable distribution of influence between the primary and secondary reagents. In particular, it enhanced the representation for products similar to query structure 7

well above the proportion seen in the full library (compare Table 4 with Fig. 9).

It bears noting that the approach described here is quite general, and could also be carried out using fast combinatorial docking scores [34] in lieu of topomer similarities from ChemSpace. Indeed, although these particular designs have yet to be synthesized or evaluated for biological activity, variations on the strategy employed have been successfully used to create focused GPCR and kinase screening libraries with confirmed activity against the respective target classes [35].

Acknowledgements The authors would like to thank Dick Cramer, Mike Lawless, Jon Swanson and Rob Jilek of Tripos, Inc., for their support and many helpful discussions about ChemSpace.

References

1. Filmore D (2004) *Mod Drug Discov* 11:24
2. Palczewski K, Kumasaka T, Hori T, Behnke CA, Motoshima H, Fox BA, Le Trong I, Teller DC, Okada T, Stenkamp RE, Yamamoto M, Miyano M (2000) *Science* 289:739
3. Krier M, de Araújo-Júnior JX, Schmitt M, Duranton J, Justiano-Basaran H, Lugnier C., Bouruignon J-J, Rognan D (2005) *J Med Chem* 48:3816
4. OptDesign[®] is distributed by Tripos, Inc., 1699 S. Hanley Road, St. Louis, MO 63144, USA (<http://www.tripos.com>)
5. Clark RD, Patterson DE, Soltanshahi F, Blake JF, Matthew JB (2000) *J Mol Graph Model* 18:404
6. Clark RD (1997) *J Chem Inf Comput Sci* 37:1181
7. Clark RD, Langton WJ (1998) *J Chem Inf Comput Sci* 38:1079
8. Clark RD, US Patent No. 6,535,819 (2003) OptiSim[™] is licensed by Tripos, Inc., 1699 S. Hanley Road, St. Louis, MO 63144, USA (<http://www.tripos.com>)
9. Clark RD, Kar J, Akella L, Soltanshahi F (2003) *J Chem Inf Comput Sci* 43:829
10. Gillet VJ, Willett P, Fleming PJ, Green DVS (2002) *J Mol Graph Model* 20:491
11. Patterson DE, Cramer RD, Ferguson AM, Clark RD, Weinberger LE (1996) *J Med Chem* 39:3049
12. Cramer RD, Poss MA, Hermsmeier MA, Caulfield TJ, Kowala MC, Valentine MT (1999) *J Med Chem* 42:3919
13. Pauwels PJ, 5-HT Receptors and Their Ligands, www.tocris.com
14. Horuk R, Ng HP, (2000) *Med Res Rev* 20:155
15. Lüttichau HR, Schwartz TW, (2000) *Drug Discov Dev* 3:610
16. Kaplan AP (2001) *Int Arch Allergy Immunol* 124:423
17. Lainton JAH, Allen MC, Burton M, Cameron S, Edwards TRG, Harden G, Hogg R, Leung W, Miller S, Morrish JJ, Rooke SM, Wendt B (2003) *J Comb Chem* 5:400
18. ChemSpace[®] is licensed by Tripos, Inc., 1699 S. Hanley Road, St. Louis, MO 63144, USA (<http://www.tripos.com>); Cramer, R.D. and Patterson, D.E., US Patent No. 6,240,374 (2001)
19. Cramer RD, Patterson DE, Clark RD, Soltanshahi F, Lawless MS (1998) *J Chem Inf Comp Sci* 6:1010
20. Andrews KM, Cramer RD (2000) *J Med Chem* 43:1723
21. Cramer RD, Jilek RJ, Guessregen S, Clark SJ, Wendt B, Clark RD (2004) *J Med Chem* 47:6777

22. ClogP is a product of BioByte Inc., Pomona CA
23. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ (1997) *Adv Drug Del Rev* 23:4
24. Ash S, Cline MA, Homer RW, Hurst T, Smith GB (1997) *J Chem Inf Comput Sci* 37:71
25. Integer arithmetic is used, so 50% of 3 is 2
26. Waly MA (2000) *Boll Chim Farm* 139:217
27. Xu Y-C, Johnson KW, Phebus LA, Cohen M, Nelson DL, Schenck K, Walker CD, Fritz JE, Kaldor SW, Le Tourneau ME, Murff RE, Zgombick JM, Calligaro DO, Audia JE, Schaus JM (2001) *J Med Chem* 44:4031
28. Fritz JE, Kaldor SW, Liang SX, Singh U, Xu Y-C (1998) *Internatl. Patent No. WO9815545*
29. Roland C, Gozalbes R, Nicolai E, Paugam M-F, Coussy L, Barbosa F, Horvath D, Revah F (2005) *J Med Chem* 48:6563
30. Pennell AMK, Aggen JB, Wright JJK, Sen S, McMaster BE, Dairaghi DJ (2003) *Internatl. Patent No. WO03105853*
31. Bauman JG, Buckman BO, Ghanam AF, Hesselgesser JE, Horuk R, Islam I, Liang M, May KB, Monahan SD, Morrissey MM, Ng HP, Wei GP, Xu W, Zheng W (1998) *Internatl. Patent No. WO9856771*
32. Wellner E, Sandin H (2005) *Internatl. Patent No. WO05080362*
33. McMaster B (2003) *Internatl. Patent No. WO03105857*
34. Sprous DG, Lowis DR, Leonard JM, Heritage T, Burkett SN, Baker DS, Clark RD (2004) *J Comb Chem* 6:530
35. LeadDiscovery™ screening libraries are distributed by Tripos, Inc., 1699 S. Hanley Rd., St. Louis, MO 63144, USA (<http://www.tripos.com>)