SHOP: Scaffold HOPping by GRID-Based Similarity Searches

Rikke Bergmann,*[,†] Anna Linusson,[‡,§] and Ismael Zamora[||,⊥]

Department of Medicinal Chemistry, University of Copenhagen, Faculty of Pharmaceutical Sciences, Universitetsparken 2, DK-2100 Copenhagen Ø, Denmark, AstraZeneca R&D Mölndal, SE-43183 Mölndal, Sweden, Lead Molecular Design, Valles 96-102, Loc 27, 08172 Sant Cugat de Valles, Barcelona, Spain, Grup de Recerca Biomédica, IMIM Universitat Pompeu Fabra, Aiguader, 80 08124 Barcelona, Spain

Received October 26, 2006

A new GRID-based method for scaffold hopping (SHOP) is presented. In a fully automatic manner, scaffolds were identified in a database based on three types of 3D-descriptors. SHOP's ability to recover scaffolds was assessed and validated by searching a database spiked with fragments of known ligands of three different protein targets relevant for drug discovery using a rational approach based on statistical experimental design. Five out of eight and seven out of eight thrombin scaffolds and all seven HIV protease scaffolds were recovered within the top 10 and 31 out of 31 neuraminidase scaffolds were in the 31 top-ranked scaffolds. SHOP also identified new scaffolds with substantially different chemotypes from the queries. Docking analysis indicated that the new scaffolds would have similar binding modes to those of the respective query scaffolds observed in X-ray structures. The databases contained scaffolds from published combinatorial libraries to ensure that identified scaffolds could be feasibly synthesized.

Introduction

Numerous computational approaches can now be taken to identify new lead compounds in the drug discovery process, but the ideal approach to take in specific cases depends on the available information. When no structural information about a target is available, drug discovery is usually focused on known ligands. The development of parallel synthesis techniques, combinatorial chemistry, and statistical designs for selecting building blocks has greatly facilitated the drug discovery process.[1−3] However, there have been relatively few publications regarding the description and design of the scaffolds to which the building blocks are attached, even though the scaffolds used to create libraries strongly affect biological activity.[4] Moreover, there are many potential reasons in drug discovery for substituting the central framework of a chemical entity, for example, to identify chemotypes that have improved pharmacokinetic properties or lower toxicity and/or are not protected by intellectual property rights. Scaffold hopping is an approach for obtaining new drug leads for targets with known ligands. Schneider et al. defined scaffold hopping as: "Identification of iso-functional molecular structures with significantly different molecular backbones",[5] that is, preserving the 3D interaction properties of a scaffold while changing the structural skeleton. However, there are few methods available for performing scaffold hopping in an automated way.[6,7] The difficulty is to find methods that can link 3D molecular properties of any scaffold with geometrical information and synthetic feasibility. One of the first publicly available programs for scaffold hopping was CAVEAT, which uses geometrical features such as vector pairs to search for new scaffolds in a database.[8] Information about chemical groups and interaction potential is not included in CAVEAT models and, thus, scaffold properties must be introduced in a postprocessing step.[9] Many of the more recent methods define scaffolds as relatively rigid ring systems[7,8,10] and some are directed toward de novo design, which limits the feasibility of synthesizing the identified scaffolds,[11,12] while other methods are limited to specific types of ligands, such as peptides.[13] Most methods define scaffolds as entire ligands and only few[7,8] address the central skeleton as a fragment that can be substituted while keeping the substituents. This would be desirable when properties such as favorable absorption, distribution, metabolism, and excretion/toxicity (ADME/Tox), affinity, or selectivity can be attributed to substituents. Recently published 3D scaffold hopping methods outperform 2D methods,[14,15] and shared features of these methods are that they consider flexibility issues as well as geometry and pharmacophore-like molecular properties.

The study presented here evaluated and validated a method (SHOP)[16] that considers geometrical features (distance and dihedral angle) and the scaffold shape in conjunction with alignment-independent GRID molecular interaction field (MIF) descriptors. In a prior study, SHOP was successfully used for scaffold hopping using a template scaffold derived from a thrombin-ligand X-ray structure.[17] Here, the methodology was thoroughly evaluated and shown to be capable of finding scaffolds for several targets, including novel, divergent scaffolds that bind in a similar fashion to the query scaffolds according to docking analyses. Three different targets relevant for drug discovery were used: human α-thrombin and HIV-1 protease (which are involved in blood coagulation and important in AIDS treatments, respectively) and influenza virus neuraminidase. First a detailed investigation of how each descriptor affected the outcome of scaffold searches was carried out using well-established statistical methods and the search procedure was optimized. The method was then applied in real case scenarios to search for scaffolds in a database of compounds for which synthetic routes are known to be available.

Methods

SHOP Methodology. The SHOP procedure identifies new scaffolds in a database by analyzing the similarity of their 3D structures to that of a query scaffold. The objective of such analyses is to find substitutes for the query scaffold, while retaining the geometry, shape, and interaction patterns of the central fragment

* To whom correspondence should be addressed. Tel.: + 45 35 30 61 62. Fax: + 45 35 30 60 40. E-mail: rb@farma.ku.dk.
† University of Copenhagen.
‡ AstraZeneca R&D Mölndal.
§ Current address: Department of Chemistry, Umeå University, SE-901 87 Umeå, Sweden.
|| Lead Molecular Design.
⊥ IMIM Universitat Pompeu Fabra.

10.1021/jm061259g CCC: $37.00
Published on Web 05/10/2007

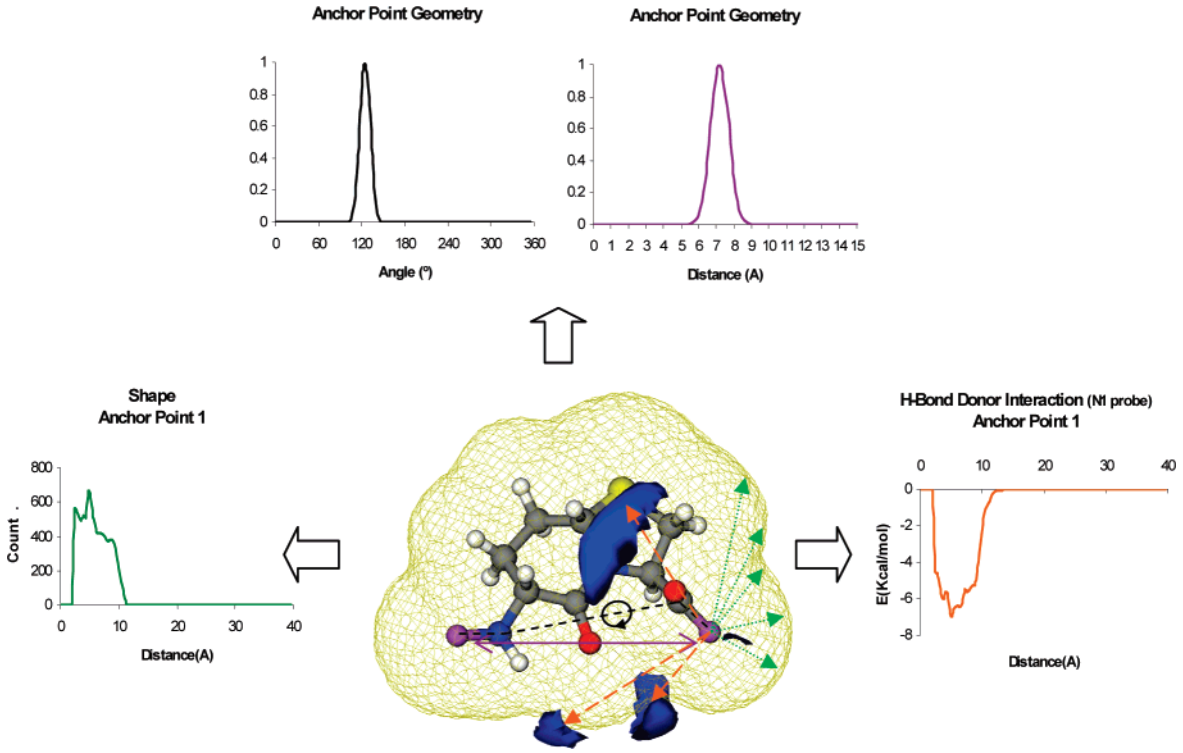

Figure 1. SHOP descriptors computed for a scaffold with two anchor points (shown in purple). The geometric features were represented by Gaussian functions, distance depicted in purple (full line), and the dihedral angle in dashed black. The scaffold shape was described by the interaction with the polar N1 probe at a level of 1.0 kcal/mol. The distance from each anchor point on the scaffold to each point on the calculated grid energy surface (beige net) versus the distance was recorded (green corellogram and dotted green arrows). The anchor-GRIND is exemplified by interactions with the donor probe N1. The most favorable energy of interaction was recorded for each distance bin (0.4 Å wide) and plotted against the distance (orange correlogram and dashed orange arrows).

of the query ligand. The descriptors used in the search are anchor point specific, where "anchor points" refer to positions on a scaffold to which building blocks could be attached (reactive sites). This means that distances and dihedral angles are calculated between the anchor points, and also, the shape and interaction pattern is encoded with reference to each reactive site. These features allow the chemist to exchange the central fragment *and* retain building blocks with desirable properties, for example, favorable binding, selectivity, or ADME/Tox parameters.

In the SHOP program, the user must provide the 2D or 3D structure of the query scaffold for which the anchor points are defined as dummy atoms. A 2D to 3D conversion will take place if a 2D query scaffold is used. The search database can either be user-generated or the CombChem-DB can be used (as in this study), which is provided with SHOP v. 2.0 or available free of charge from www.moldiscovery.com. Once these are supplied, the method is fully automatic and the program will return a list of the scaffolds in the database, sorted in order of similarity to the query. SHOP is suitable for substituting either relatively rigid central fragments or flexible scaffolds for which the bioactive conformation is known from X-ray crystallography analyses, even though the method as such is not restricted by ligand flexibility. If more than one potential scaffold conformation is identified, it is advisable to run several queries using different search conformations. Correspondingly, the conformational space of the central fragment in the search database should be covered by including several low-energy conformations for each scaffold. A database manager is provided with the program, which will supply multiple conformations of the scaffolds if desired. Overall, the method is fast, automated, and straightforward to use. It is intended for medicinal chemists as well as computational chemists. Details of the different parts of the SHOP program are described below.

Scaffold Characterization. The scaffolds were characterized using SHOP (v. 1.0 for Linux and Windows) and three sets (geometrical, shape, and GRIND) of molecular descriptors (Figure 1). Default settings for descriptor calculations were used.

(a) In SHOP, the scaffolds are geometrically described according to the distances and dihedral angles between their anchor points. The distances are recorded in bins of 0.4 Å, the dihedral angles are transformed into binned distances of 5 degrees, and both are described as Gaussian functions to allow overlap between different bins in the similarity analysis.

(b) The shape of the scaffolds is based on frequency analysis of distances between each anchor point and the scaffold surface computed with MIFs. To obtain the shape description, the GRID program[18,19] is automatically called, within SHOP, to calculate interaction energies in a box-shaped grid surrounding the scaffold at a 5 Å distance. The distances between each anchor point and each grid point with an energy value greater than 1 kcal/mol (determined using the noncharged polar N1 probe) are counted. A shape description is created by binning the number of distances as a function of the distance in 0.4 Å wide bins.

(c) The third kind of descriptors are based on a modified version of the GRID alignment independent descriptors (GRIND).[20] The potential interactions that the scaffolds could have with the target macromolecule were estimated as favorable interaction energies calculated between the scaffold and a GRID-probe and registered as a function of the distances between the fixed anchor points and the interaction points.[21] By examining sites of potential interaction with a protein rather than the position of a group or atom with a property of interest in the scaffolds, ligands could be discovered that are likely to have similar interactions as the query scaffold involving different parts of the scaffolds. The interaction profiles are determined using the GRID program, called by SHOP, and five probes corresponding to five interaction properties: DRY (hydrophobic), N1 (hydrogen bond donor), O (hydrogen bond acceptor), N1+ (positive charge), and O− (negative charge). A grid step of 0.5 Å is used by SHOP, while the other GRID parameters are kept at their default values. The distances between each of the anchor points and the MIF points are computed, and the corresponding energy values are registered. The interaction energies are binned according to distances, keeping the number of bins fixed to 100

and setting the bin space to 0.8 times the grid step, that is, 0.4 Å, so all energies corresponding to distances between 10.0 and 10.4 Å are considered in the same bin. The most favorable energy value was selected for each of the distance bins to represent the five interaction profiles based on the five probes.

The anchor points are represented as hydrogen atoms to facilitate the GRID MIF computations. Such substitution could lead to the detection of false interactions between the probes and the hydrogen atoms that represent the anchor points. To avoid this possibility, all the MIF points that are closer than 3 Å to the anchor point are removed from the descriptor calculation within the SHOP program.

Similarity Calculations. The similarity between the query scaffold and the scaffolds in the database described by the descriptors are compared using the Carbó index.[22] The total similarity Sim_{tot} is given as the sum of each similarity value Sim of the eight descriptor sets *d*, the five anchor-GRIND (hydrophobic, donor, acceptor, negative, and positive charges) and the shape, distance, and dihedral characterizations, using individual weights *w* for the similarity indices according to eq 1.

$$\mathrm{Sim}_{\mathrm{tot}} = \sum w_{\mathrm{d}}\,\mathrm{Sim}_{\mathrm{d}} \qquad (1)$$

The weights can be modified by the user to tailor the scaffold search and, hence, focus on features of the query scaffold that are known to be important. The similarities of all possible alignment combinations are compared, for example, the comparison of two compounds with two anchor points A1, A2 and B1, B2, respectively, would give the possible similarity sums (A1;B1 + A2;B2) and (A1;B2 + A2;B1). The anchor point alignment giving the highest similarity is selected and reported in the ranking list.

Database Preparation. Scaffolds from known combinatorial libraries, synthesized for biological testing, were extracted from published literature, and 2D structures were manually constructed.[23−29] The anchor points were introduced manually in the 2D depiction of the molecules phase as R groups. Confort (v. 6.0) was used for both 2D to 3D conversion and conformer generation.[30,31] A diverse set of conformers for each scaffold was stored in two databases (max. 40 conformers, and max. 10 kcal/mol above the global minimum), and the R-groups were converted into dummy atoms. The database (DB-R) used for recovering test set scaffold included 471 compounds and a total of 9719 conformers containing scaffolds from references 23, 24, and 29. For validation purposes, DB-R was spiked with known biologically active scaffolds, for which the bioactive conformation was known from X-ray crystallography (see datasets). The larger database, CombChem-DB, containing 2500 compounds and 28 281 conformers in total was searched for new scaffolds able to substitute the known X-ray query scaffolds.[32] CombChem-DB contained scaffolds from references 23−29.

Datasets. A number of complexes of ligands and proteins representing three protein families (thrombin, HIV protease, and influenza virus neuraminidase) were selected from the Research Collaboratory for Structural Bioinformatics (RCSB) Protein Data Bank.[33] The criteria for selecting the test sets were that they should (a) include as many complexes as possible with similar ligands binding in a similar fashion; (b) have similar binding site conformations within each test set to extract ligands with comparable conformations and interaction patterns with the target protein; (c) allow different numbers of anchor points to be assigned to the selected test sets; and (d) the proteins should belong to different families and be relevant for drug discovery.

Thrombin. The set of thrombin complexes selected for the scaffold recovery included inhibitors that occupy the S3, S2, S1, and S1′ binding pockets of the thrombin active site. Interactions in the S3 and S2 pocket are hydrophobic in nature. Interacting residues in S2 are His57, Tyr60A, Trp60D, and Leu99, while S3 is formed by Leu99, Ile174, and Trp215. A hydrophobic channel in S1 leads to the acidic Asp189 and the backbone carbonyl groups of Gly219 and Phe227 at the bottom of the channel. The S1 pocket is also called the specificity pocket due to its specificity for a basic amine or guanidine group. Finally, residues Leu41, Cys44, His57, Cys58, Tyr60A, Trp60D, Lys60F, and Gly193 form the half-open S1′ pocket,[34−36] which contributes to the hydrophobic environment in the active site of thrombin.

For the validation study, six crystal structures (pdb codes: 1B5G, 1A2C, 1AY6, 1BA8, 1BMM, and 1TMB) were identified in which the ligands comprised unique scaffolds and had similar molecular properties and orientations in the binding site. The ligands were extracted and side chains were deleted in two ways yielding, in each case, a scaffold with two anchor points and another with three anchor points. The scaffolds with two anchor points represented ligands occupying the S1 and S2 pockets, while those with three anchor points also populated the S1′ pocket. The scaffolds of two ligands, extracted from 1A2C and 1AY6, were further modified by deleting a side chain resulting, in total, in eight scaffolds with two anchor points and a further eight with three anchor points (Figure 2a,b). The scaffold resulting from ligand 1B5G was selected as the query molecule for scaffolds with both two and three anchor points. This scaffold forms three direct hydrogen bonds and one water-mediated bond with thrombin. In addition, the hydrophobic bicyclic ring system fits well into the S2 pocket surrounded by Tyr60A, Trp60D, and Leu99. The crystal structure of thrombin with the selective inhibitor SDZ 229-357, K_i 145 nM (pdb code: 1BHX),[37] was chosen to test SHOP's ability to find new thrombin scaffolds. As shown in Figure 3, the interactions between thrombin and SDZ 229-357 involve four direct hydrogen bonds, three water-mediated hydrogen bonds, and a salt bridge between SDZ 229-357's guanidine moiety and Asp189. Finally, the hydrophobic pockets S2 and S3, defined above, are filled by the bicyclic ring structure and a phenyl group. A two-anchor point scaffold was obtained by cutting the two amide bonds, resulting in the anchor points positioning in place of the amide nitrogen and the sulfamide sulfur.

HIV-1 Protease. HIV-1 protease is a symmetrical dimer with an active site at the interface between the two identical subunits. The selected test set of HIV-1 protease inhibitors bind to the S1 and S1′ pockets (which are identical) and the S2 and S2′ pockets (which are also identical). These subsites are mainly hydrophobic. S1/S1′ are formed by Ile23/23′, Gly27/27′, Ile50/50′, Thr80/80′, Pro81/81′, Ile84/84′, and the catalytic Asp25/25′, while S2/2′ are formed by Ala28/28′, Asp30/30′, Val32/32′, Ile47/47′, Ile50/50′, and Leu76/76′.[38,39] Seven crystal complexes (pdb codes: 1AJV, 1G2K, 1G35, 1AJX, 1BVG, 1HVR, and 1PRO) with cyclic ureas and cyclic sulfamide inhibitors were chosen, all of which have four aromatic side chains occupying the S1/S1′ and S2/S2′ pockets. These side chains were removed, resulting in three unique scaffolds, seven in total, comprising four anchor points (Figure 2c).

The scaffold from HIV-1 protease inhibitor AHA-006, K_i 19.1 nM,[38] from crystal structure 1AJV was selected as the query scaffold for both the scaffold recovery study and for investigating SHOP's potential for bioisosteric replacements. The cyclic ureas and cyclic sulfamides form direct hydrogen bonds with the amide N−H from Ile50/50′ as opposed to peptide inhibitors, for which binding to Ile50/50′ is mediated by a water molecule.[40] The query scaffold forms five hydrogen bonds with HIV-protease. Two of these are formed between the sulfamide oxygens and the backbone N−H between Ile50 and Ile50′. The two hydroxyl groups from the scaffold form hydrogen bonds with the catalytic aspartates Asp25 and Asp25′. Finally, the hydrophobic pockets S1/S1′ and S2/S2′ are occupied by the aromatic side chains of the inhibitor AHA-006 (Figure 4).

Neuraminidase. The active site of influenza virus neuraminidase is highly polar. Three arginine residues (Arg118, Arg292, and Arg371) interact electrostatically in the site with an acidic moiety of the natural substrate sialic acid as well as the inhibitor 2,3-didehydro-2-deoxy-*N*-acetylneuraminic acid (DANA). Furthermore, Asp151, Arg152, Glu227 (water-mediated), and Glu276 interact with the sialic acid. A hydrophobic region is formed near Trp178 and Ile222.[41−43]

Thirty-one neuraminidase complexes (pdb codes: 1F8B, 1F8C, 1F8D, 1F8E, 1IVF, 2QWC, 1NNB, 2QWD, 2QWE, 1NNC, 2QWF, 2QWI, 2QWJ, 2QWG, 1BJI, 2QWK, 2QWH, 1IND, 1IVC, 1IVE, 1ING, 1INH, 2QWB, 1MWE, 2BAT, 1INW, 1INY, 1INX, 1L7F,

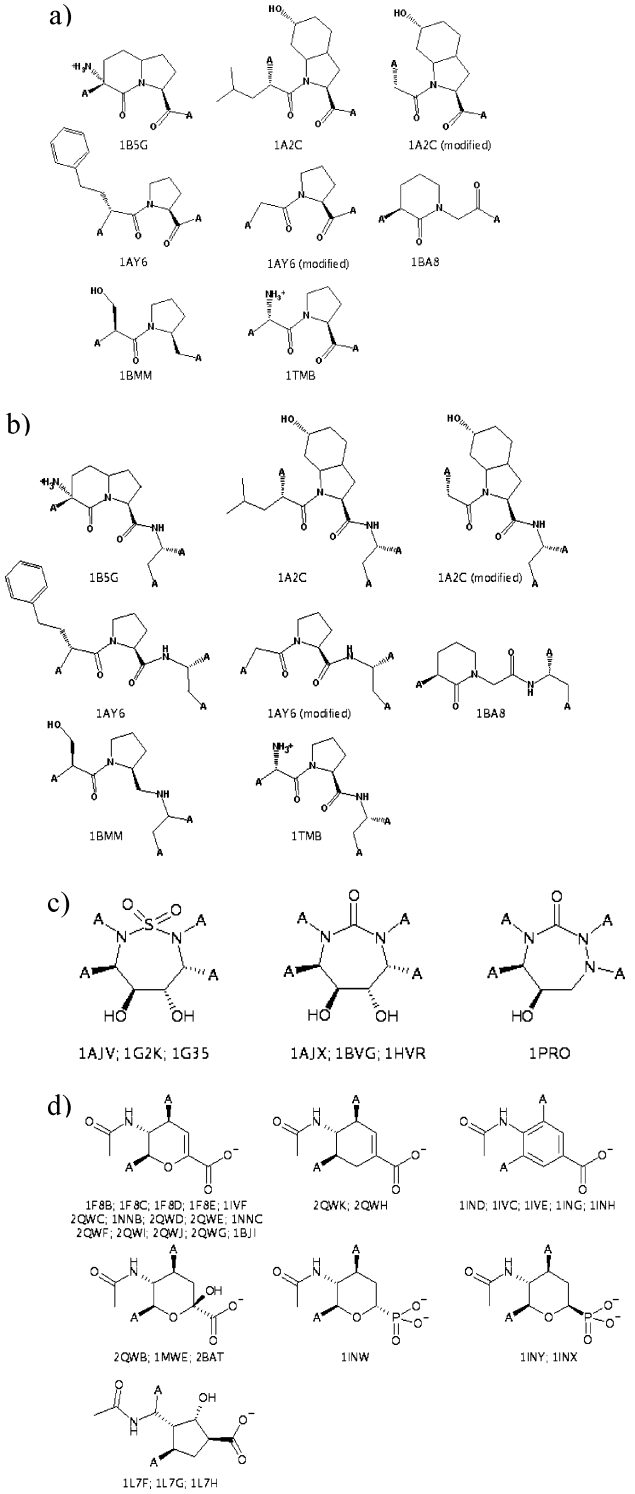

Figure 2. Scaffolds used for validation: (a) thrombin case study with two anchor points, (b) thrombin case study with three anchor points, (c) HIV case study with four anchor points, and (d) neuraminidase case study with two anchor points.

1L7G, and 1L7H) were selected from the PDB database from which scaffolds with two anchor points were created. The 31 scaffolds comprised seven unique central fragments (Figure 2d).

The selected query scaffold for scaffold recovery originates from the inhibitor DANA (1F8B), which displays many of the interactions with neuraminidase described above for sialic acid.[42] The interaction of Asp151 is substituted by another acidic amino acid, Glu119. The methyl group of the 5-$NHCOCH_3$ substituent is positioned in the hydrophobic region near Trp178 and Ile222.

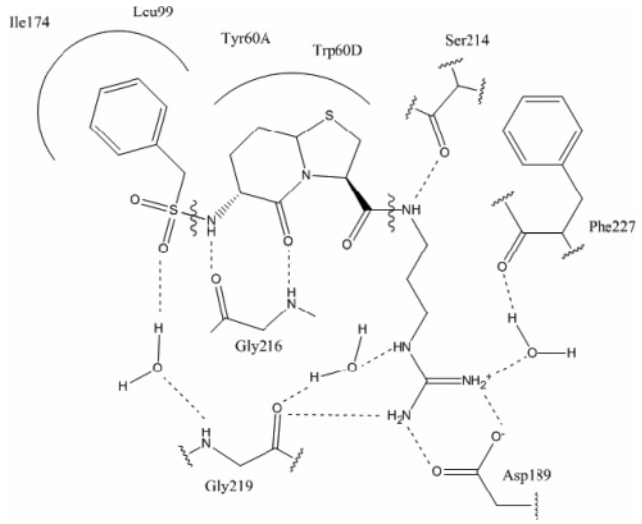

Figure 3. Interactions of the thrombin inhibitor SDZ 229-357 with thrombin and structural water molecules. Broken amide bonds in SDZ 229-357 indicate points where the compound was divided to obtain the scaffold for searching the CombChem-DB. The anchor points were attached in the place of the amide nitrogen and the sulfamide sulfur.

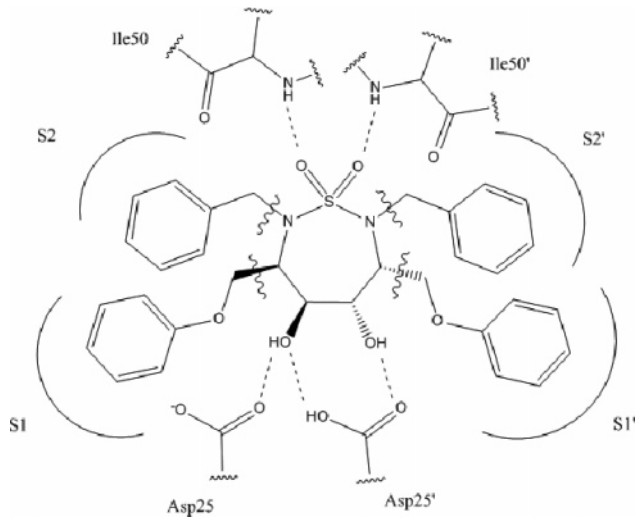

Figure 4. HIV-1 protease interactions with the inhibitor AHA-006 from the X-ray structure 1AJV. Sissile bonds indicate the bonds that were broken to obtain the query scaffold with four anchor points. Anchor points substitute the distal carbon atoms. One of several possible protonation states and the hydrogen-bonding pattern of the Asp25/25′ residues with AHA-006 are presented in the sketch.

The inhibitor 4-amino-DANA, K_i 40 nM,[42] from the X-ray structure 1F8C was chosen for the docking test. Figure 5 summarizes its interactions with neuraminidase. These are the same hydrogen bonds as described for DANA, with the addition of two extra hydrogen bonds from the 4-amino moiety to Asp151, one of which is mediated by a water molecule. A scaffold with three anchor points was obtained by deleting the side chains, as indicated in Figure 5.

Protein and Ligand Preparation. Structural modifications were performed using Sybyl (v. 6.9 for Irix).[44] The proteins within each test set were aligned based on their Cα atoms (Sybyl: Align Structures Using Homology) and ligands were extracted. Side chains of the ligands were removed and anchor points on the scaffolds were indicated by dummy atoms. The anchor points were selected based on maximum 3D overlap within a test set of scaffolds. All atom types were checked, manually changed when required, and hydrogens were added (Sybyl: Biopolymer Add Hydrogens).

The protein PDB files were prepared for docking, leaving only the protein chains and structural water molecules interacting with

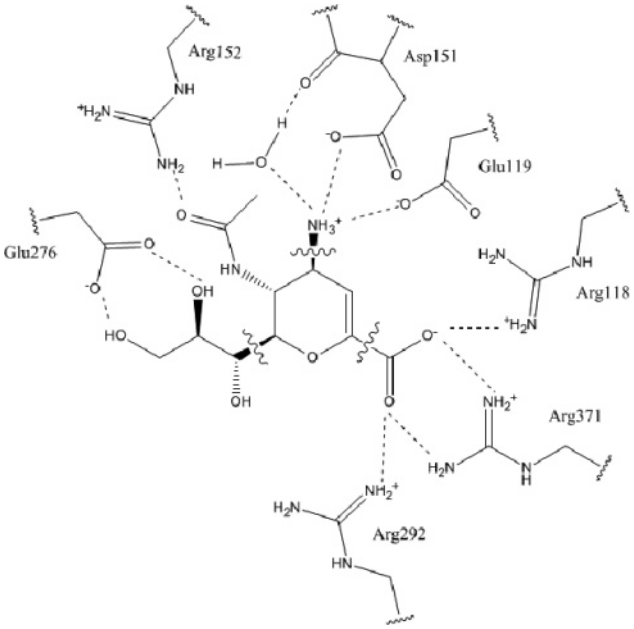

Figure 5. Neuraminidase inhibitor 4-amino-DANA interactions with influenza virus neuraminidase (1F8C). Broken bonds indicate the position of the anchor points in the selected three anchor point scaffold, substituting the amino-N, the carboxyl-C, and the first carbon in the polyhydroxyl side chain.

the ligands (four water molecules in 1BHX, none in 1AJV, and one in 1F8C). Hydrogens were added in standardized geometry in Maestro v. 6.0 for Linux.[45] The orientations of Asn and Gln side chain amides and His rings were checked using REDUCE.[46] No changes were suggested to residues closer than 12 Å to the binding site of the proteins 1BHX and 1AJV, so they were left unchanged. Three asparagines (Asn221, Asn294, and Asn346) within 12 Å of the neuraminidase 1F8C ligand were flipped. One of these, Asn294, was in direct Van der Waals contact with 4-amino-DANA. The orientations of water molecule hydrogens in the thrombin and neuraminidase cavities were optimized by minimization of the hydrogen atoms in the presence of the inhibitors using MacroModel v. 8.1 for Linux[47,48] and the MMFFs force field.

Scaffold Recovery. The ability of the SHOP method to recover known actives from DB-R was validated for each of the three targets: thrombin, HIV-1 protease, and neuraminidase. A systematic investigation of how each descriptor influenced the outcome of scaffold searches was carried out by applying statistical experimental design to the individual weights within the total similarity index used to rank the scaffolds in the database. The weights of the eight different molecular properties were tested at three levels (0.10, 0.45, and 0.90), and 162 combinations were selected using a combined D-optimal and fractional factorial design (81+81).[49,50] The results were compared with those obtained with a reference set of equal coefficients of 0.9 for all similarity indices. The scaffold recovery using the different settings of the individual weights for the similarity indices was evaluated in two ways: by identifying (A) how many of the known scaffolds were present among the 10, 20, 30, 40, 50, and 100 top-ranked suggestions and (B) how many scaffolds needed to be selected to recover 80, 90, and 100% of the known scaffolds. These two sets of responses were further used to build regression models using the different parameters as the X-matrix and the sets of responses, (A) and (B), as the Y-matrix. Partial least-square (PLS) projection to latent structures was used to determine how the different parameter settings affected the outcome of the similarity searches.[51−53] The number of significant components was decided by cross-validation.[54,55] A tailor-made set of parameters for each of the test cases was calculated based on the fitted PLS models using a Nelder Mead simplex method implemented in MODDE software.[49,56] Finally, the PLS models were used to identify a default set of parameters, yielding good results for all test cases.

Table 1. Results Obtained Using the Different Sets of Coefficients

	response set (A)						response set (B)		
coefficient set[a]	10	20	30	40	50	100	80%	90%	100%
Thrombin 2 Anchor Points									
reference set[b]	3	4	6	6	6	7	27	96	504
optimized 1	4	6	6	6	6	6	12	222	825
optimized 2	3	6	6	6	6	7	16	61	553
optimized (1−2)[c]	4	6	6	6	6	6	13	102	724
optimized (1−4)[d]	5	6	6	6	6	6	12	89	762
default[e]	3	4	5	5	5	7	58	66	508
Thrombin 3 Anchor Points									
reference set[b]	3	5	5	5	5	6	85	216	479
optimized 3	6	6	6	6	7	7	5	47	185
optimized 4	5	6	6	7	7	7	7	12	220
optimized (3−4)[f]	7	7	7	7	7	7	7	9	165
optimized (1−4)[d]	5	6	6	7	7	7	11	39	209
default	3	5	5	6	7	7	39	48	480
HIV Protease									
reference set[b]	3	4	4	6	6	6	37	38	117
optimized 5	7	7	7	7	7	7	4	5	8
optimized 6	6	7	7	7	7	7	4	5	11
optimized (5−6)[g]	7	7	7	7	7	7	4	5	6
default	5	7	7	7	7	7	10	11	21
Neuraminidase									
reference set[b]	10	20	27	28	30	31	48	49	50
optimized 7	10	20	29	31	31	31	28	30	31
optimized 8	10	20	30	31	31	31	28	29	30
optimized (7−8)[h]	10	20	30	31	31	31	28	29	30
default	10	20	29	31	31	31	28	30	31

[a] The optimized coefficients are related to the row numbers in Table 2. [b] All weights are set equal to 0.9. [c] The coefficient set is an average between optimized 1 and 2. [d] The coefficient set is an average between optimized 1, 2, 3, and 4. [e] The default set includes the coefficients 0.9 for distance, 0.45 for all interactions and the shape, and 0.1 for the dihedral angle. [f] The coefficient set is an average between optimized 3 and 4. [g] The coefficient set is an average between optimized 5 and 6. [h] The coefficient set is an average between optimized 7 and 8.

The experimental designs and subsequent PLS modeling were performed using MODDE software v. 6.0 for Windows.[49]

Methodology Test. The optimized coefficients were applied in searches for new scaffolds in the CombChem-DB for the three targets: thrombin, HIV-1 protease, and neuraminidase. In each case, the 10 top-ranked suggestions from SHOP were subjected to further investigation. The original side chains from the query scaffolds, 1BHX, 1AJV, and 1F8C (Figures 2, 3, and 4), were attached to the new scaffolds. The resulting compounds were minimized to convergence with MacroModel v. 8.1[47,48] using the MMFFs force field and the GB/SA solubility model, as implemented in MacroModel. Subsequently, they were docked into the proteins from which the query scaffolds were extracted. If the query scaffolds were substituted appropriately, the new docked structures were expected to have similar geometries and binding modes to the template scaffold obtained from X-ray crystallography. All dockings were made using GLIDE XP v. 4.0 for Linux.[57] Twenty poses for each compound were allowed, and otherwise, default settings were used. Docking poses were evaluated visually. The criteria used to select new compounds to illustrate the potential for bioisosteric replacements were that they should have side chains in similar positions to the ligand in the X-ray structure with the corresponding interactions; a high G-Score similar to or better than that obtained for docking the X-ray ligand itself; and a pose mimicking some or all of the query scaffold interactions.

To investigate the docking performance of GLIDE XP, the ligands from 1BHX, 1AJV, and 1F8C were removed from the binding sites and docked back into their respective binding sites after they were prepared as described above. The root-mean-square deviations (RMSDs) between the X-ray coordinates of the heavy atoms of the ligands and their docked poses were calculated. A RMSD of <1.5 Å was considered to indicate good reproduction of the binding mode.

Table 2. Tailor-Made Sets of Parameters Optimized for Each of the Different Test Cases

No	target	response set[a]	No of anchors	dist.[b]	lipo.[c]	don.[d]	acc.[e]	pos.[f]	neg.[g]	shape[h]	dih.[i]
1	thrombin	A	2	0.90	0.84	0.50	0.47	0.90	0.26	0.34	0.90
2	thrombin	B	2	0.82	0.86	0.49	0.44	0.47	0.51	0.57	0.49
3	thrombin	A	3	0.89	0.88	0.31	0.01	0.90	0.00	0.03	0.74
4	thrombin	B	3	0.90	0.90	0.45	0.00	0.54	0.18	0.00	0.00
5	HIV	A	4	0.90	0.29	0.90	0.02	0.90	0.90	0.85	0.00
6	HIV	B	4	0.90	0.31	0.89	0.35	0.38	0.38	0.57	0.00
7	NA	A	2	0.81	0.00	0.90	0.18	0.90	0.90	0.90	0.00
8	NA	B	2	0.45	0.18	0.90	0.90	0.45	0.00	0.90	0.00

[a] Response set (A): the number of known scaffolds that were present among the 10, 20, 30, 40, 50, and 100 A-ranked suggestions. Response set (B): The number of scaffolds needed to be selected to find 80, 90, and 100% of the known scaffolds. [b] Distance. [c] Lipophilic. [d] Donor. [e] Acceptor. [f] Positive. [g] Negative. [h] Shape. [i] Dihedral angle.

Results and Discussion

Scaffold Recovery. The known scaffolds for all three of the targets (thrombin, HIV-1 protease, and neuraminidase) were all recovered among the top-ranked scaffolds in the DB-R database using the SHOP methodology (Table 1). Furthermore, the weights of the eight similarity indices affected the results, and the tailor-made sets of weights for the different targets reflect the important molecular interaction properties between the scaffolds and the corresponding proteins (Table 2). The results for each of the three test sets are discussed in detail below.

Thrombin. Five and seven out of eight of the scaffolds with two anchor points and three anchor points, respectively, were recovered in the top 10-ranked scaffolds in the database using the tailor-made weights, which is highly satisfactory for drug discovery purposes. None of the queries managed to recover the 1BMM scaffold within the first 100 solutions (the top 1% rankings of the database). The most likely reasons for the failure to recover this scaffold are that it is the scaffold with the most strongly deviating dihedral angles (up to 147 degrees) and it has a substantially different shape compared to the majority of the scaffolds in the thrombin test sets.

The distances between the anchor points, the lipophilicity of the scaffolds, and the acceptor capabilities were the most important features for thrombin scaffolds (Table 2). The donor capability, the positive and negative charge, and the shape varied in importance, in accordance with the variation of these features in the test set (see Figure 2a,b). The fact that hydrophobic interactions were weighted highly for all thrombin models was not surprising because thrombin is known to have a highly hydrophobic active site. The selected part of the thrombin inhibitors defined as the scaffold in this investigation bind in the hydrophobic S2 pocket.[35,36] The interactions with a donor probe and/or a positive charge probe with hydrogen bond donor capabilities had medium weights in all the thrombin coefficient sets analyzed. This is not surprising because all of the thrombin scaffolds, shown in Figure 2a,b, have one or more carbonyl groups.

HIV Protease. All seven HIV scaffolds were recovered within the top ten ranked scaffolds (top 0.1% of the database) using the tailor-made set of weights. According to these weight values, the charged interactions and the hydrogen bond donors were important for the scaffold characterization of HIV protease inhibitors, as can be verified from the scaffolds obtained from the crystal structures. The negatively charged probe simulated the Asp25/25′, and the donor/positive charge covered the interaction between the urea/sulfone amide oxygens and the Ile50/50′ backbone. The shape descriptors had a strong influence in the total similarity calculations used for ranking, which is consistent with expectations because all were seven-membered cyclic structures. Furthermore, the weight values for the HIV protease-case revealed that hydrophobic characterizations of the scaffolds were of low importance for ranking the scaffolds. This is also consistent with expectations because the selected scaffolds contained several donors and acceptors and the aromatic side-chains of the cyclic urea and sulfamide inhibitors had been removed.

Table 3. Reproduction of Crystal Structure Inhibitor Poses by GLIDE XP

target	docking solution	RMSD[a]	G-score	E_{model}	energy
thrombin, 1BHX	2	0.43	−13.03	−126.1	−52.9
HIV-1 protease, 1AJV	1	0.40	−9.37	−124.0	−69.7
neuraminidase, 1F8C	2	0.42	−4.46	−75.7	−53.3

[a] All atoms were included in the RMS calculation.

Table 4. Docking of the New Compounds

target	new cmpd	docking solution[a]	G-score	E_{model}	energy
thrombin, 1BHX	T2	1	−10.42	−106.2	−65.7
HIV-1 protease, 1AJV	H2	3	−9.32	−128.8	−78.5
HIV-1 protease, 1AJV	H2b	1	−11.63	−126.4	−73.1
neuraminidase, 1F8C	N5	5	−3.26	−73.5	−54.4

[a] The docking solution refers to the rank of the pose within all poses proposed for the compound itself.

Neuraminidase. All 31 neuraminidase test set scaffolds were recovered in the top 31 ranked scaffolds using optimized coefficients. As observed in the search for HIV protease scaffolds, hydrophobic interactions were of low importance for the recovery of neuraminidase scaffolds, which is not surprising because the neuraminidase binding site has a polar character. Consequently, the hydrogen bond donor/acceptor descriptors, with and without charges, made strong contributions, as did the shape descriptors, which was also unsurprising given the similarities of the selected scaffolds in this respect (Figure 2d).

Default Set of Coefficients. The results across targets suggest that distance was generally highly important and the dihedral angle was less important. The interaction properties and shape varied greatly depending on important features of the scaffolds. Therefore, appropriate default settings could be 0.9 for distance, 0.45 for all of the interaction properties and shape, and 0.1 for the dihedral angle (see Table 1 for results). The outcome using the default settings was good; four and five out of seven of the scaffolds with two and three anchor points binding to thrombin, respectively, and seven out of seven of those binding to HIV protease were recovered within the top 20 rankings of the database, and all of the 31 neuraminidase scaffolds were within the top 31 rankings. If no knowledge is available about important features of a target, these seem to be good starting options.

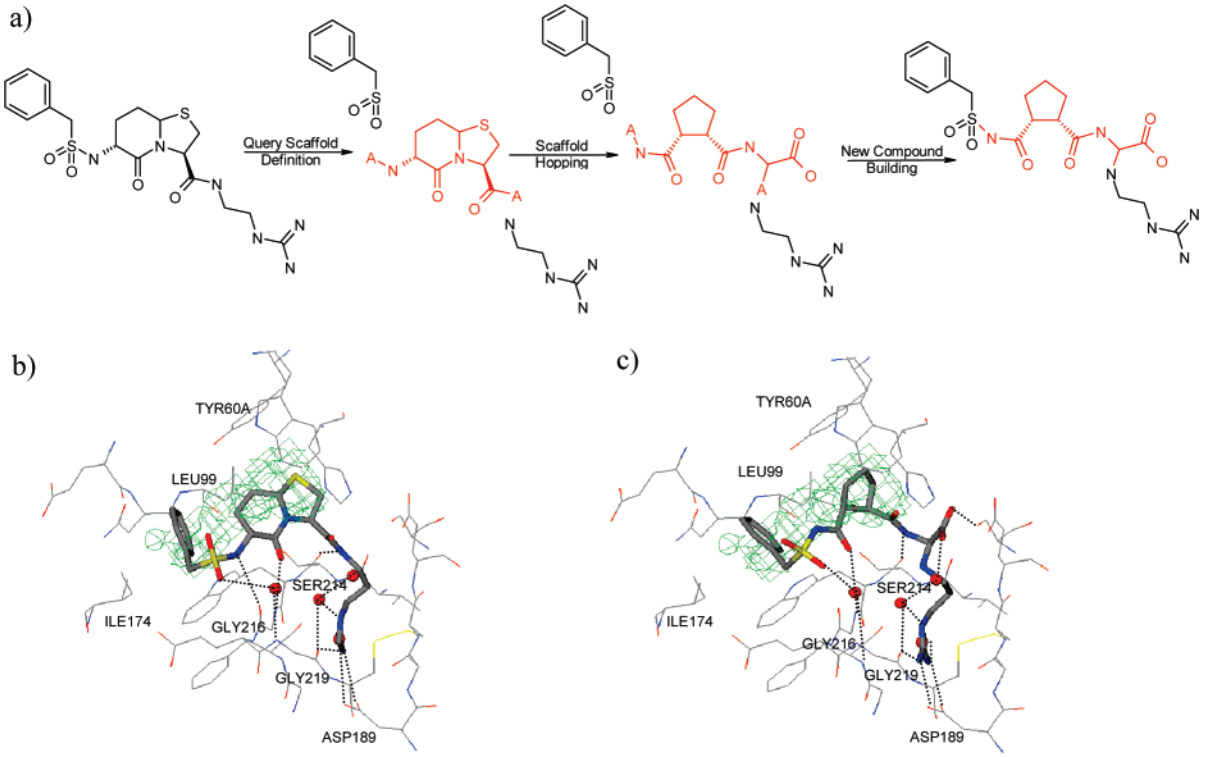

Figure 6. The test procedure for the thrombin case: (a) First the query scaffold was defined and the CombChem-DB was searched for scaffolds fulfilling the descriptor criteria calculated by SHOP. Ten new compounds were assembled using the most highly ranked new scaffolds from the CombChem-DB and the building blocks from the original compound. The exposed example is compound T2 (ranked second). The new compound T2 displayed many of the parent interactions (b) when docked into the thrombin binding site of the crystal structure 1BHX (c).

Docking Test. The ability of SHOP to identify isofunctional structures in large datasets was assessed by searching CombChem-DB for scaffolds for each of the three target proteins. One example of the outcome of such substitution and search is presented for each of the three test sets. Table 3 summarizes the results from the docking reproduction of the crystal structure ligand poses, all of which were docked with an RMSD of 0.4 Å to the corresponding crystal structure. Table 4 presents docking results for the predicted new compounds. In each case, the G-scores were similar to those for the corresponding crystal structures.

Thrombin. The T2 scaffold from the CombChem-DB was ranked second in the SHOP search using the optimized coefficients for thrombin scaffolds with two anchor points. Figure 6a illustrates how the SDZ 229-357 building blocks from the crystal structure were reattached to T2 to form the new compound. In Figure 6b, the interactions between SDZ 229-357 and thrombin are illustrated, and the docked pose of the new compound is presented in Figure 6c. The original arginine and phenyl side chains both docked into their parent positions from the 1BHX crystal structure, and the side-chain interactions made by the thrombin inhibitor SDZ 229-357 were thus retained. The new scaffold included a pentacyclic proline-like ring (without the nitrogen) that fitted well into the hydrophobic S2 pocket, in a similar manner to the proline in the motif D-Phe-Pro-Arg of the natural substrate fibrinogen.[36] Most of the hydrogen bonds were retained and several new ones were gained. The interaction with Gly216 was not seen in the docking pose of T2, but was observed in other scaffolds among the top 10. A new interaction between Ser195 and a carboxylic acid in the T2 scaffold occurred. The original reference for this scaffold shows that it was developed during the course of a general method for making unsymmetrically functionalized diamides from diacids.[58] However, this does not exclude the possibility that the T2 scaffold could bind to α-thrombin.

HIV Protease. Using the four anchor point scaffold 1AJV as query input, the second-ranked CombChem-DB scaffold was H2 (Figure 7a). After attaching building blocks to the H2 scaffold and subjecting it to the docking procedure, most of the AHA-006 interactions were reproduced. The hydrogen bonds to the Asp25/25′ were donated by a hydroxyl group, as in the parent scaffold, and an amide NH. A phoshonate group was the bioisosteric replacement for the sulfone amide, its sp^2 oxygen interacting with the Ile50′. Moreover, the aromatic side chains were oriented such that they occupied the same hydrophobic pockets as AHA-006 (Figure 7b). Its G-score and energy were similar to those of AHA-006 (Tables 3 and 4). Due to poor chemical stability of the acetal moiety, which was formed when the building blocks were attached, a modified version of the new compound leaving out the acetal was constructed and docked (Figure 7c). As well as retaining all the interactions formed by H2, this compound (H2b) also formed a hydrogen bond from the Ile50 backbone NH to a sp^3 oxygen of the phosphonate group, then fulfilling all the interactions made by AHA-006 in the crystal structure (Figure 7d). Furthermore, the Glide score was improved by two units compared to both AHA-006 and H2 (Table 4). The synthesis reference for the scaffold reveals that it was developed for a combinatory library targeting aspartic acid proteases like HIV-1 protease.[59] The compound was patented in 1999 by Carroll et al. for aspartic proteases exemplified by plasmepsin and cathepsin D.[60] Here a retro-synthetic pathway is presented, illustrating a possible way to make the final compound. The first step in the suggested retro-synthesis pathway was described by Cavallaro et al.[61] and the second by Dolle et al.[59] (Figure 7d).

Neuraminidase. When using the three anchor point 1F8C scaffold as a query, the fifth-ranked scaffold N5 from the

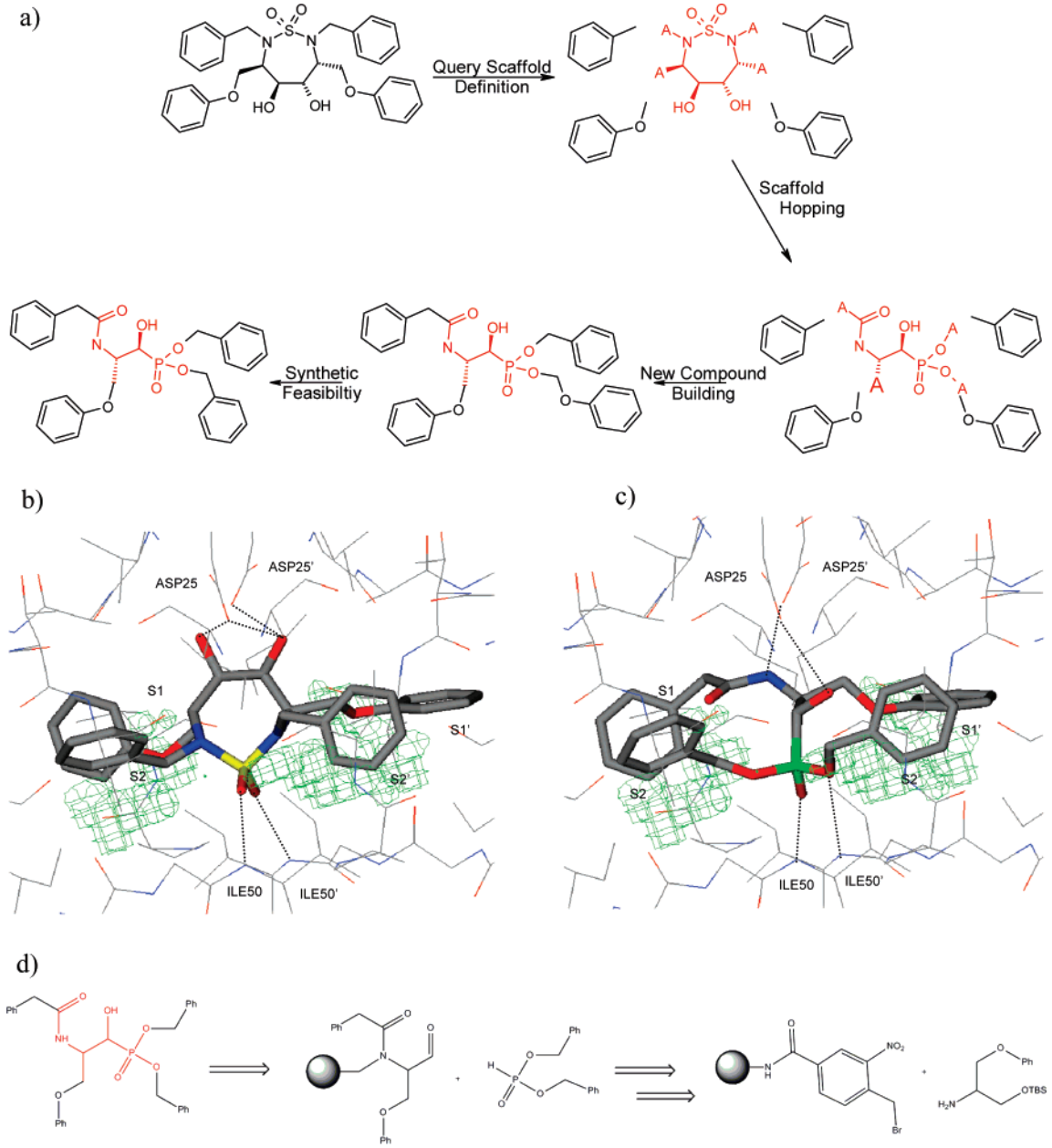

Figure 7. Test procedure for the HIV test case: (a) The 1AJV query scaffold was defined and used for searching the CombChem-DB. The original building blocks were added to the top 10 scaffolds, considered by SHOP to be most similar to the query, here exemplified by scaffold 2 (H2). Due to poor chemical stability and synthetic feasibility of the acetal group in H2 formed when attaching the substituents, an oxygen atom was removed from the building block. The novel compound (H2b) interacted in the same way as the original compound (b) when it was docked into the HIV-1 protease active site (c). (d) A retro synthetic pathway was possible with a look into the original reference of the new scaffold (see text for details). The reference revealed that the scaffold was originally produced as part of a library of aspartic protease inhibitors.

CombChem-DB seemed to be the most interesting among the top ten according to the criteria listed in the methods section (Figure 8a). When docked, the new compound N5 covered most of the interactions with neuraminidase displayed by 4-amino-DANA (Figure 8b,c). The six-membered didehydro ring system of 4-amino-DANA was substituted by a triazole ring that was able to position the side chains in the same way as the 1F8C query scaffold. The carboxylic acid was predicted to bind exactly as in the crystal structure making salt bridges to the arginines Arg118, Arg292, and Arg371. The amine was positioned at the same location as in the parent compound, reproducing the three interactions made by the 4-amino-DANA in the crystal structure. These were salt bridges to Glu119 and Asp151 and water-mediated hydrogen bonds to the backbone carbonyls of Asp151 and Trp178. Moreover, the polyhydroxyl side chain was positioned to make hydrogen bonds between the two distal hydroxyl groups and the acidic side chain of Glu276, similar to 4-amino-DANA. Two new hydrogen bonds were formed: one donated by Arg224 to the oxygen of the central hydroxyl group and the other between the third proximal hydroxyl group and Glu277. The G-score of the new compound was slightly lower than that of 4-amino-DANA (Table 4). The reference for the N5 scaffold indicates that it was originally made as part of a general combinatorial library of pyrazole- and triazole-containing compounds that were not directed toward any specific target.[62]

Conclusions

Excellent results were achieved for recovery of all chosen test sets when sets of coefficients optimized by statistical

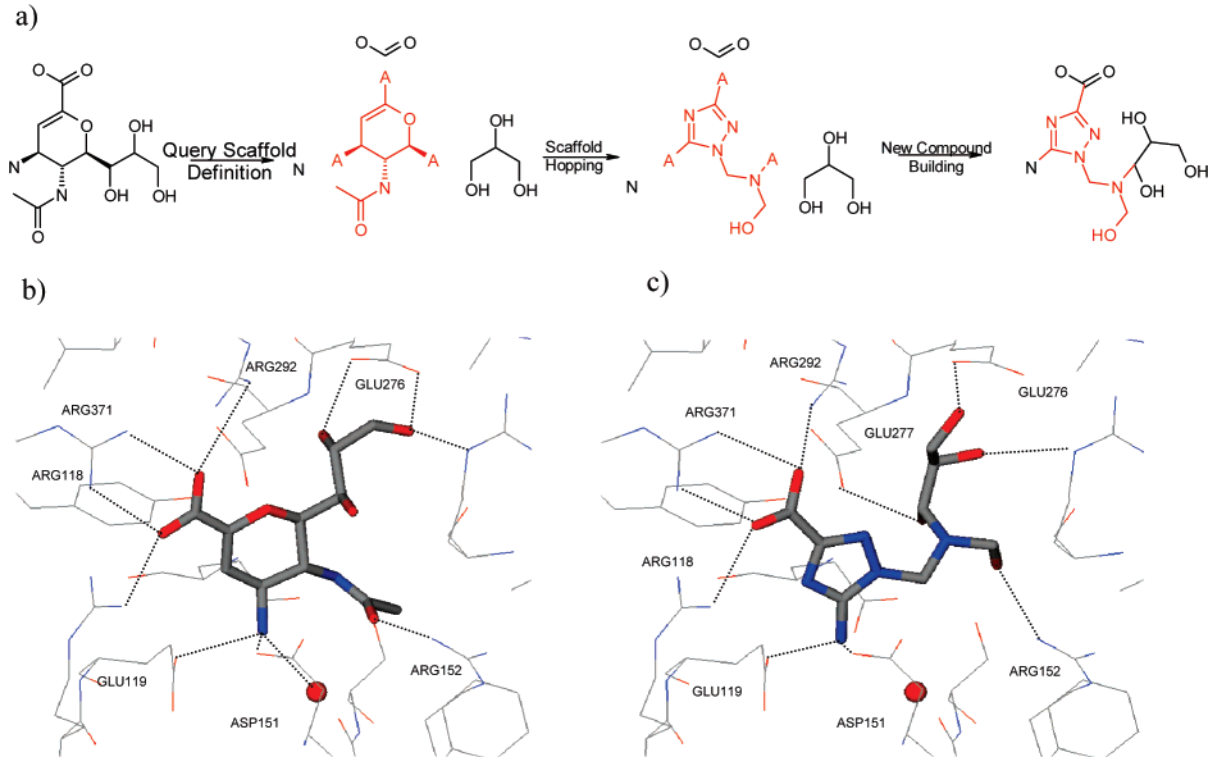

Figure 8. The test procedure for the neuraminidase case: (a) A scaffold with three anchor points was selected for searching the CombChem-DB. The top 10 scaffolds from the search were reappended to the removed building blocks, here exemplified by the N5 compound. Nine out of 10 hydrogen bonds involved in the interactions between the inhibitor 4-amino-DANA with neuraminidase (b) were reproduced by N5 when it was docked into the 1F8C binding cavity (c). Two additional hydrogen bonds were formed between N5 and Arg224 and Glu277 (not shown).

experimental design were employed in the similarity measurements in SHOP. Even a default set of coefficients gave good results for all three test sets. When optimized coefficients were used, five out of eight and seven out of eight of the thrombin test sets with two and three anchor points, respectively, and all seven of the HIV test set protease scaffolds were recovered within the top 10 ranked scaffolds and all 31 neuraminidase scaffolds were within the top 31 rankings. Even when the standardized set of coefficients were used, only slightly lower numbers of X-ray test set scaffolds were recovered with high similarity scores to the query scaffolds: four out of eight and five out of eight of the thrombin test sets with two and three anchor points, respectively, and all seven of the HIV test set protease scaffolds were recovered within the top 20 ranked scaffolds and again all 31 neuraminidase scaffolds were within the top 31 rankings. These are highly promising results, illustrating the power of the SHOP method even with the suggested default set of coefficients. When optimized coefficients were used to search the CombChem-DB, new scaffolds were identified and successfully docked, showing quite similar interaction patterns to the query scaffolds. The docking scores were comparable to those obtained for the crystal structure ligands, and the poses were highly ranked. Because the chemical structures of the new scaffolds were quite unrelated to the query scaffolds, it follows that a selection of bioisosteres was obtained and valuable new insights and information were acquired by using SHOP. A search for new leads was beyond the scope of the present study, the purpose of which was to examine the power of the method using a few test cases. Therefore, only one example of scaffold substitution tested by docking was presented for each target, although several interesting poses were observed among the new scaffolds from CombChem-DB. A database based on combinatorial libraries was used for this purpose, ensuring accessibility to the scaffolds, although synthetic pathways for the entire compounds with attached substituents must be designed to ensure that they are stable compounds that can be viably synthesized. This we attempted in the HIV-1 protease test example by designing a retro synthetic pathway with a promising outcome.

Acknowledgment. AstraZeneca R&D Mölndal, Drug Research Academy (DRA) and the Swedish Research Council are gratefully acknowledged for financially supporting the validation of SHOP. We are indebted to Monica Clementi for manually constructing the databases DB-R and CombChem-DB.

Supporting Information Available: Summary of two multi-Y PLS models for the 162 different weight combinations and the two sets of responses. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(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) Linusson, A.; Gottfries, J.; Lindgren, F.; Wold, S. Statistical molecular design of building blocks for combinatorial chemistry. *J. Med. Chem.* **2000**, *43*, 1320−1328.

(3) Linusson, A.; Gottfries, J.; Olsson, T.; Ornskov, E.; Folestad, S.; Norden, B.; Wold, S. Statistical molecular design, parallel synthesis, and biological evaluation of a library of thrombin inhibitors. *J. Med. Chem.* **2001**, *44*, 3424−3439.

(4) Sauer, W. H.; Schwarz, M. K. Molecular shape diversity of combinatorial libraries: A prerequisite for broad bioactivity. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 987−1003.

(5) Schneider, G.; Neidhart, W.; Giller, T.; Schmid, G. "Scaffold-hopping" by topological pharmacophore search: A contribution to virtual screening. *Angew. Chem., Int. Ed.* **1999**, *38*, 2894−2896.

(6) Abolmaali, S. F.; Ostermann, C.; Zell, A. The compressed feature matrix−A novel descriptor for adaptive similarity search. *J. Mol. Model.* **2003**, *9*, 66−75.

(7) Bohl, M.; Dunbar, J.; Gifford, E. M.; Heritage, T.; Wild, D. J.; et al. Scaffold searching: Automated identification of similar ring systems for the design of combinatorial libraries. *Quant. Struct.-Act. Relat.* **2002**, *21*, 590−597.
(8) Lauri, G.; Bartlett, P. A. CAVEAT: A program to facilitate the design of organic molecules. *J. Comput.-Aided Mol. Des.* **1994**, *8*, 51−66.
(9) Yang, Y.; Nesterenko, D. V.; Trump, R. P.; Yamaguchi, K.; Bartlett, P. A.; Drueckhammer, D. G. Virtual hydrocarbon and combinatorial databases for use with CAVEAT. *J. Chem. Inf. Model.* **2005**, *45*, 1820−1823.
(10) Lee, M. L.; Schneider, G. Scaffold architecture and pharmacophoric properties of natural products and trade drugs: Application in the design of natural product-based combinatorial libraries. *J. Comb. Chem.* **2001**, *3*, 284−289.
(11) Lloyd, D. G.; Buenemann, C. L.; Todorov, N. P.; Manallack, D. T.; Dean, P. M. Scaffold hopping in de novo design. Ligand generation in the absence of receptor information. *J. Med. Chem.* **2004**, *47*, 493−496.
(12) Todorov, N. P.; Dean, P. M. Evaluation of a method for controlling molecular scaffold diversity in de novo ligand design. *J. Comput.-Aided Mol. Des.* **1997**, *11*, 175−192.
(13) Chianelli, D.; Kim, Y. C.; Lvovskiy, D.; Webb, T. R. Application of a novel design paradigm to generate general nonpeptide combinatorial scaffolds mimicking beta turns: synthesis of ligands for somatostatin receptors. *Bioorg. Med. Chem.* **2003**, *11*, 5059−5068.
(14) Jenkins, J. L.; Glick, M.; Davies, J. W. A 3D similarity method for scaffold hopping from known drugs or natural ligands to new chemotypes. *J. Med. Chem.* **2004**, *47*, 6144−6159.
(15) Zhang, Q.; Muegge, I. Scaffold hopping through virtual screening using 2D and 3D similarity descriptors: Ranking, voting, and consensus scoring. *J. Med. Chem.* **2006**, *49*, 1536−1548.
(16) *SHOP*, v. 1.0 for Linux; Molecular Discovery Ltd.: 215 Marsh Road, HA5 5NE Pinner, Middlesex, U.K.
(17) Ahlstrom, M. M.; Ridderstrom, M.; Luthman, K.; Zamora, I. Virtual screening and scaffold hopping based on GRID molecular interaction fields. *J. Chem. Inf. Model.* **2005**, *45*, 1313−1323.
(18) *GRID*, v. 22 for Linux; Molecular Discovery Ltd.: 215 Marsh Road, HA5 5NE Pinner, Middlesex, U.K.
(19) Goodford, P. J. A computational procedure for determining energetically favourable binding sites on biologically important macromolecules. *J. Med. Chem.* **1985**, *28*, 849−857.
(20) Pastor, M.; Cruciani, G.; McLay, I.; Pickett, S.; Clementi, S. GRid-INdependent descriptors (GRIND): A novel class of alignment-independent three-dimensional molecular descriptors. *J. Med. Chem.* **2000**, *43*, 3233−3243.
(21) Fontaine, F.; Pastor, M.; Zamora, I.; Sanz, F. Anchor-GRIND: Filling the gap between standard 3D QSAR and the GRid-INdependent descriptors. *J. Med. Chem.* **2005**, *48*, 2687−2694.
(22) Carbo, R.; Leyda, L.; Arnau, M. How similar is a molecule to another? An electron density measure of similarity between two molecular structures. *Int. J. Quantum Chem.* **1980**, *17*, 1185−1189.
(23) Dolle, R. E. Comprehensive survey of combinatorial library synthesis: 1999. *J. Comb. Chem.* **2000**, *2*, 383−433.
(24) Dolle, R. E. Comprehensive survey of combinatorial library synthesis: 2000. *J. Comb. Chem.* **2001**, *3*, 477−517.
(25) Dolle, R. E. Comprehensive survey of combinatorial library synthesis: 2001. *J. Comb. Chem.* **2002**, *4*, 369−418.
(26) Dolle, R. E. Comprehensive survey of combinatorial library synthesis: 2002. *J. Comb. Chem.* **2003**, *5*, 693−753.
(27) Dolle, R. E. Comprehensive survey of combinatorial library synthesis: 2003. *J. Comb. Chem.* **2004**, *6*, 623−679.
(28) Dolle, R. E. Comprehensive survey of combinatorial library synthesis: 2004. *J. Comb. Chem.* **2005**, *7*, 739−798.
(29) Dolle, R. E.; Nelson, K. H., Jr. Comprehensive survey of combinatorial library synthesis: 1998. *J. Comb. Chem.* **1999**, *1*, 235−282.
(30) *Confort*, v. 6.0 for Linux; Tripos Inc.: 1699 South Hanley Rd, St. Louis, MO 63144-2919, U.S.A.
(31) Perlman, R. S.; Balducci, R. *Confort: A novel algorithm for conformational analysis*; National Meeting of the American Chemical Society, New Orleans, 1998; American Chemical Society: Washington, DC, 1998.
(32) CombChem-DB can be downloaded from http://www.moldiscovery-.com.
(33) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235−242.
(34) St Charles, R.; Matthews, J. H.; Zhang, E.; Tulinsky, A. Bound structures of novel P3−P1′ beta-strand mimetic inhibitors of thrombin. *J. Med. Chem.* **1999**, *42*, 1376−1383.
(35) Nilsson, J. W.; Kvarnstrom, I.; Musil, D.; Nilsson, I.; Samulesson, B. Synthesis and SAR of thrombin inhibitors incorporating a novel 4-amino-morpholinone sscaffold: Analysis of X-ray crystal structure of enzyme inhibitor complex. *J. Med. Chem.* **2003**, *46*, 3985−4001.
(36) Lange, U. E.; Baucke, D.; Hornberger, W.; Mack, H.; Seitz, W.; Hoffken, H. W. D-Phe-Pro-Arg type thrombin inhibitors: Unexpected selectivity by modification of the P1 moiety. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 2029−2033.
(37) Wagner, J.; Kallen, J.; Ehrhardt, C.; Evenou, J. P.; Wagner, D. Rational design, synthesis, and X-ray structure of selective noncovalent thrombin inhibitors. *J. Med. Chem.* **1998**, *41*, 3664−3674.
(38) Backbro, K.; Lowgren, S.; Osterlund, K.; Atepo, J.; Unge, T.; Hulten, J.; Bonham, N. M.; Schaal, W.; Karlen, A.; Hallberg, A. Unexpected binding mode of a cyclic sulfamide HIV-1 protease inhibitor. *J. Med. Chem.* **1997**, *40*, 898−902.
(39) Abdel-Rahman, H. M.; Al-karamany, G. S.; El-Koussi, N. A.; Youssef, A. F.; Kiso, Y. HIV protease inhibitors: Peptidomimetic drugs and future perspectives. *Curr. Med. Chem.* **2002**, *9*, 1905−1922.
(40) Hilgeroth, A. Dimeric 4-aryl-1,4-dihydropyridines: Development of a third class of nonpeptidic HIV-1 protease inhibitors. *Mini-Rev. Med. Chem.* **2002**, *2*, 235−245.
(41) Wilson, J. C.; von Itzstein, M. Recent strategies in the search for new anti-influenza therapies. *Curr. Drug Targets* **2003**, *4*, 389−408.
(42) Smith, B. J.; Colman, P. M.; Von, Itzstein, M.; Danylec, B.; Varghese, J. N. Analysis of inhibitor binding in influenza virus neuraminidase. *Protein Sci.* **2001**, *10*, 689−696.
(43) Varghese, J. N.; McKimm-Breschkin, J. L.; Caldwell, J. B.; Kortt, A. A.; Colman, P. M. The structure of the complex between influenza virus neuraminidase and sialic acid, the viral receptor. *Proteins* **1992**, *14*, 327−332.
(44) *Sybyl*, v. 6.9 for Irix; Tripos Inc.: 1699 South Hanley Rd, St. Louis, MO 63144-2919 U.S.A.
(45) *MAESTRO*, v. 6.0 for Linux; Schrödinger LLC: New York, U.S.A.
(46) Lovell, S. C.; Davis, I. W.; Arendall, W. B., III; de Bakker, P. I.; Word, J. M.; Prisant, M. G.; Richardson, J. S.; Richardson, D. C. Structure validation by Calpha geometry: phi, psi, and Cbeta deviation. *Proteins* **2003**, *50*, 437−450.
(47) *MacroModel*, v. 8.1 for Linux; Schrödinger LLC: New York, U.S.A.
(48) Mohamadi, F.; Richards, N. G. J.; Guida, W. C.; Liskamp, R.; Lipton, M.; Caufield, C.; Chang, G.; Hendrickson, T.; Still, W. C. Macromodel−An integrated software system for modeling organic and bioorganic molecules using molecular mechanics. *J. Comput. Chem.* **1990**, *11*, 440−467.
(49) *MODDE*, v. 6.0 for Windows; Umetrics: Box 7960, SE-907 19 Umeå, Sweden.
(50) Box, G. E. P.; Hunter, W. G.; Hunter, J. S. *Statistics for experimenters. An introduction to design, data analysis and model building*; John Wiley & Sons, Inc.: New York, 1978.
(51) Burnham, A. J.; MacGregor, J. F.; Viveros, R. Latent variable multivariate regression modeling. *Chemom. Intell. Lab. Syst.* **1999**, *48*, 167−180.
(52) Wold, S. PLS for multivariate linear modeling. *Chemometric methods in molecular design*; VCH: Weinheim, 1995; pp 195−218.
(53) Wold, H. The basic design and some extensions. *Systems under indirect observation*; North-Holland Publishing company: Amsterdam, 1982; pp 1−53.
(54) Stone, M. Cross-validatory choice and assessment of statistical predictions. *J. R. Stat. Soc.* **1974**, *36B*, 111−133.
(55) Wold, S. Cross-validatory estimation of the numbers of components in factor and principal component models. *Technometrics* **1978**, *20*, 397−405.
(56) Nelder, J. A.; Mead, R. A simplex method for function minimization. *Comput. J.* **1965**, *7*, 308−313.
(57) *GLIDE*, v. 4.0 for Linux; Schrödinger, LLC: New York, U.S.A.
(58) Wahhab, A.; Leban, J. A simple procedure for the solid phase synthesis of unsymmetrically functionalised diamides from symmetric diacids. *Tetrahedron Lett.* **1999**, *40*, 235−238.
(59) Dolle, R. E.; Herpin, T. F.; Shimshock, Y. C. Solid-phase synthesis of α-hydroxy phosphonates and hydroxystatine amides. Transition-state isosteres derived from resin-bound amino acid aldehydes *Tetrahedron Lett.* **2001**, *42*, 1855−1858.
(60) Carroll, C. D.; Dolle, R. E.; Shimshock, Y. C.; Herpin, T. F.; Hansen, P. E. Glycol and hydroxyphosphonate peptidomimetics as inhibitors of aspartyl proteases. U.S. Patent US005,962,506A, Oct. 5, 1999.
(61) Cavallaro, C. L.; Herpin, T.; McGuinness, B. F.; Shimshock, Y. C.; Dolle, R. E. Allylindium and allylboronic acid pinacolate: Mild reagents for the allylation of resin-bound aldehydes. Application to the solid-phase synthesis of hydroxypropylamines *Tetrahedron Lett.* **1999**, *40*, 2711−2714.
(62) Touzani, R.; Garbacia, S.; Lavastre, O.; Yadav, V. K.; Carboni, B. Efficient solution phase combinatorial access to a library of pyrazole- and triazole-containing compounds. *J. Comb. Chem.* **2003**, *5*, 375−378.

JM061259G