Reagent Selector: Using Synthon Analysis to Visualize Reagent Properties and Assist in Combinatorial Library Design

Ralph T. Mosley,*,† J. Christopher Culberson,*,‡ Bryan Kraker,† Bradley P. Feuston,‡ Robert P. Sheridan,† John F. Conway,‡ Joseph K. Forbes,‡ Subhas J. Chakravorty,‡ and Simon K. Kearsley†

Department of Molecular Systems, RY50SW-100, Merck Research Laboratories, P. O. Box 2000, Rahway, New Jersey 07065, and Department of Molecular Systems, WP53F-301, Merck Research Laboratories, Sumneytown Pike, P.O. Box 4, West Point, Pennsylvania 19486

Received March 21, 2005

Reagent Selector is an intranet-based tool that aids in the selection of reagents for use in combinatorial library construction. The user selects an appropriate reagent group as a query, for example, primary amines, and further refines it on the basis of various physicochemical properties, resulting in a list of potential reagents. The results of this selection process are, in turn, converted into synthons: the fragments or R-groups that are to be incorporated into the combinatorial library. The Synthon Analysis interface graphically depicts the chemical properties for each synthon as a function of the topological bond distance from the scaffold attachment point. Displayed in this fashion, the user is able to visualize the property space for the universe of synthons as well as that of the synthons selected. Ultimately, the reagent list that embodies the selected synthons is made available to the user for reagent procurement. Application of the approach to a sample reagent list for a G-protein coupled receptor targeted library is described.

INTRODUCTION

Large or small, directed toward a particular target or to enhance diversity of a sample repository, solid or solution phase synthesis, it all comes down to one thing in designing a combinatorial library: selection of the reagents. Because of the overwhelming number of reagents available for incorporation into a library, combinatorial chemistry often becomes computationally assisted chemistry.[1,2] Software has been developed,[3] and some has been made commercially available,[4] that purports to make the selection of reagents a manageable process whether by focusing on characteristics of the product library,[5] for example, meeting diversity standards,[6] or by targeting properties of the reagents themselves, for example, clustering them into families.[7]

At Merck Research Laboratories (MRL), a multiyear initiative is underway in which the sample repository is being enhanced by the addition of combinatorial libraries designed to range in size from ∼3 000 to ∼20 000 chemical entities. Rather than use software to automate the design of these libraries or have the molecular modeling group design them, it was decided that MRL medicinal chemists should be employed in the library design process. The chemists are responsible for both the identification of the library core or scaffold and the selection of the reagents amenable to combinatorial synthesis. Their overarching goal is to design a library that could provide a lead structure they themselves would be happy to use as a starting point en route to a new medicine. The objective of our group has been to design and implement an ensemble of software tools to facilitate this effort without encroaching upon the chemists' mandate.

The Virtual Library Tool Kit (VLTK), of which the Reagent Selector (RS) and Synthon Analysis (SA) are components, has been developed as an intranet-based software tool to meet this need. As will be discussed more fully elsewhere,[8] VLTK encapsulates the workflow needed to select reagent groups, transform these into synthons from which the ultimate reagents are identified, and assemble the synthons into a virtual library. The virtual library can then be assessed on the basis of its profile of calculated physicochemical properties relative to that of the sample repository or other libraries before any reagents are purchased and synthesis of the library initiated. The focus of this report is on the Reagent Selector, with particular emphasis on the conceptualization, design, and implementation of the Synthon Analysis component.

RS: QUERY BUILDING

Selecting a list of suitable reagents of a particular type is a surprisingly complex task. A substructure query must be constructed that is able to reliably retrieve the desired functionality and exclude moieties that could result in undesirable side products or be unsuitable for a lead compound. Then, one must ensure that the reagents are available in sufficient quantity, at reasonable cost, and in good purity and are deliverable in a timely fashion. Given the number of libraries planned for synthesis, this could become a burdensome, redundant exercise beyond the tolerance of medicinal chemists who are to be enlisted in this process.

Our solution has been to provide curated lists of reagent groups. In a fashion that will be more fully described

* Corresponding author tel.: (R.T.M.) 732-594-3126, (J.C.C.) 215-652-7537; e-mail: ralph_mosley@merck.com (R.T.M.) or chris_culberson@merck.com (J.C.C.).

† Department of Molecular Systems, NJ.

‡ Department of Molecular Systems, PA.

10.1021/ci050090x CCC: $30.25
Published on Web 07/14/2005

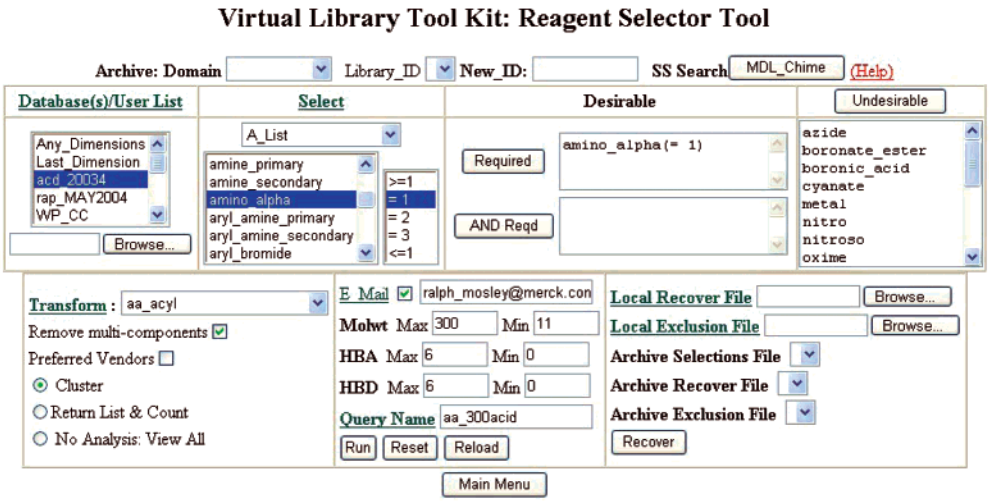


Figure 1. Reagent Selector query builder web interface, which has been completed to identify a group of amino acids for incorporation into the somatostatin agonist receptor library.

elsewhere,[9] upon each major release of the Available Chemicals Directory (ACD) or internal specialized reagent databases, we construct a derivative database from which clearly unsuitable compounds are eliminated (for instance, because they contain covalently bound metals or radioactive isotopes). Then, series of substructure searches using both ISIS and Daylight tools are undertaken to identify ~140 reagent types. These types are used to populate a pull-down menu item on the Reagent Selector query builder page (see Figure 1), which enables the user to create Boolean operable list sets of reagents. The Boolean "NOT" category is automatically populated with ~20 groups considered undesirable in a lead compound. Selection of certain reagent groups will automatically amend the "Undesirable" list. For example, selecting a less reactive species as a reagent, for example, a "secondary amine," will add the similar, more reactive species, for example, a "primary amine," to the "Undesirable" set automatically. This should curtail the selection of potentially troublesome reagents. Finally, the user can set the number of times in which the reactive moiety can occur in each reagent on the basis of the intended chemistry. In this fashion, a user can specify reagents that contain a specific count (1, 2, or 3) of an acid (the default is 1) or, alternatively, require that at least one acid be present and permit the inclusion of di- and triacids in the set. The list of reagents resulting from the Boolean operations on the list sets can be further winnowed down on the basis of the number of hydrogen bond donors/acceptors and the molecular weight prior to further analysis of the reagents.

However, it would be nonsensical to perform this filtering using the entire reagent! Some reagents have large leaving groups that could interfere with proper clustering or result in the exclusion of reagents on the basis of molecular weight. Cutoffs should only be applied to the portion of the reagent that is to be incorporated into the library molecules, that is, the "synthon". To enable synthon-centric filtering and to prepare for synthon analysis, a series of reagent transformations has been implemented. Rather than replicate synthetic steps in which each step has its own transform, that is, A → B, B → C, and C → D, virtual transformations have been designed to accomplish this in one step: A → D.

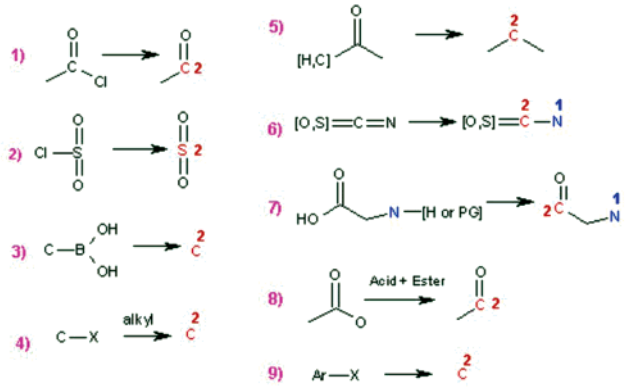


Figure 2. The nine transformations that comprise the "Capping" transformation group. Each depicts the reagent family and conversion into its synthon along with the markup employed in the creation of the virtual library.

Transformations are encoded with the programmable chemistry perceiver (PCP), an enhanced version of PATTY[10] (equivalent in function to SMARTS, using Daylight software). Each transformation performs several functions (see Figure 2):

(1) Delete atoms and adjust bond orders to reflect incorporation into the final product.

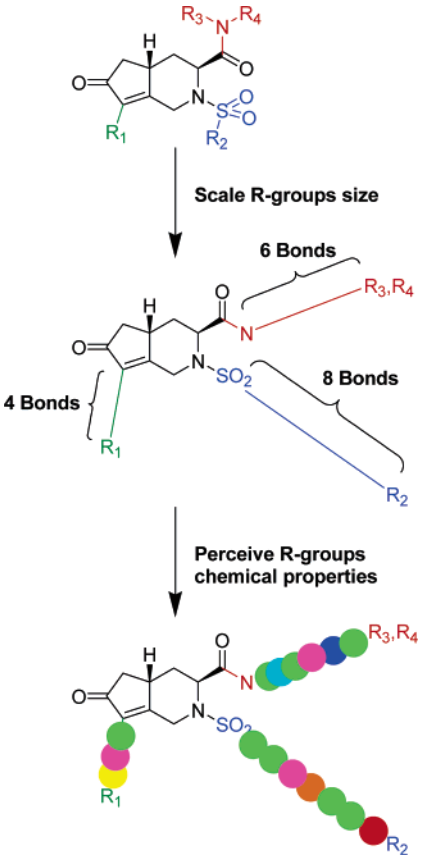
(2) Mark the atoms that will serve as attachment point(s) into the library scaffold. These are further prioritized on the basis of predefined reactivity such that the most reactive species, for example, a primary amine, is marked in preference to the less reactive species, for example, a secondary amine, if both should occur in a reagent.

(3) Perceive the moieties present in the synthon, and categorize them into chemical "types" as a function of the topological bond distance from the attachment point. The atoms that comprise the reactive species, for example, the carbonyl of an acylhalide, are excluded from this perception because they are invariant.

(4) Calculate the molecular weight of the synthon.

The chemical types[11] include neutral H-bond donors and acceptors, anions and cations, polar/unspecified H-bonding groups (species that can function as both an H-bond donor

Scheme 1. Schematic of the Original Concept for R-Group Analysis. The Color Code for the Properties Could Be Those Depicted in Figure 3, For Example



and acceptor, e.g., hydroxy), and hydrophobic groups. Because of the anticipated application of this tool, the hydrophobic grouping was replaced with three new subcategories: saturated hydrophobic groups, unsaturated hydrophobic groups, and aromatic centroids.

Approximately 120 transformations have been developed to date that reflect the growing wealth of chemistry amenable to combinatorial approaches. Several of these are grouped into larger families of transformations to facilitate the creation of synthons from chemical groups of similar utility. For example, acylhalides may be converted into acyl synthons by using the "Acylhalide" transformation, but this can also be accomplished using the "Capping" transformation. As depicted in Figure 2, the "Capping" transformation encompasses a total of nine transformations commonly employed to "cap" a nucleophilic center.

With synthons in hand, filters can be applied to remove synthons that fall outside the H-bond donor/acceptor count limits as well as the molecular weight cutoffs. Synthons are then clustered using Daylight fingerprints, Tanimoto similarity, and the nonhierarchical Jarvis−Patrick clustering algorithm.[12] Clustering reduces the number of reagents that the user needs to inspect and is the penultimate step to synthon analysis.

RS: SYNTHON ANALYSIS

Synthon analysis was initially envisioned as a form of R-group or side-chain analysis: a way to visualize the chemical properties of R-groups as they extend out from a central core.[13,14] As depicted in Scheme 1, the R-groups would be redrawn to reflect the relative distance each extends out from the scaffold and the vector populated with the perceived chemical properties. This approach could be useful in assessing the chemical property coverage for a series of substituents off a common scaffold, developing a structure−activity relationship (SAR) for those substituents, or assisting in generating a pharmacophore. It also has obvious potential as a technique to graphically assess the chemical diversity of the reagents being evaluated for incorporation into a combinatorial library. However, given the number of reagents involved, the number of chemical descriptors being used, and other practical considerations, this tool would have to be recast in order to be useful in selecting reagents.

There are several key features that need to be incorporated into this tool. These include the display of reagents, quick transitions between cluster centroids and the set of compounds that they represent, access to purchasing information, updated molecular weight profiles and counts for reagents selected, and, perhaps most importantly, an easy way to select or deselect reagents. In addition, synthon analysis, as a web-based graphical aid to selecting reagents, should have the following abilities:

(1) Focus on only one R-group or synthon at a time.

(2) Be able to display the chemical types and their frequencies for both the "universe" of synthons that passed previous filters and the "shopping cart" of synthons the user has selected for the library.

(3) Employ a metric to indicate the distance of a chemical property from the attachment point.

(4) Provide "live" updates to the graphics as users make selections.

(5) Keep the bulk of the operations and information in one window that fits onto a typical monitor screen without scrolling.

As currently implemented, the SA tool meets all of these requirements and addresses others discovered during its development. A screen capture from a typical SA session is depicted in Figure 3. In the plot, the chemical types for each synthon are mapped on the *x* axis and the distance (in bonds) from the attachment atom is plotted on the *y* axis. (We investigated using the through-space distances for 3D structures of synthons generated by Corina,[15] but this proved less suitable.) Two columns are associated with each chemical property. The first represents the chemical property space for the "universe" of synthons, those which have passed through the filtering steps previously described. The second is created and added to as a compound is selected and put into the "shopping cart" by the user. The numbers within the shapes found in each column, circles to represent the universe and diamonds to represent the shopping cart (or the primary selections), indicate the frequency of occurrence for that chemical type at that particular distance. Several other approaches to displaying frequency were explored including changing the relative size of the circles and diamonds; all proved unsatisfactory. In addition to the eight chemical types, two columns representing the molecular weights of the synthons are available to help the user maintain a healthy molecular weight profile for the library. (The "distance metric", in this case, is in increments of 50 Daltons rather than bond lengths.) When a user clicks on any of the circle or diamond icons, the structures corresponding to that subset of synthons are displayed.

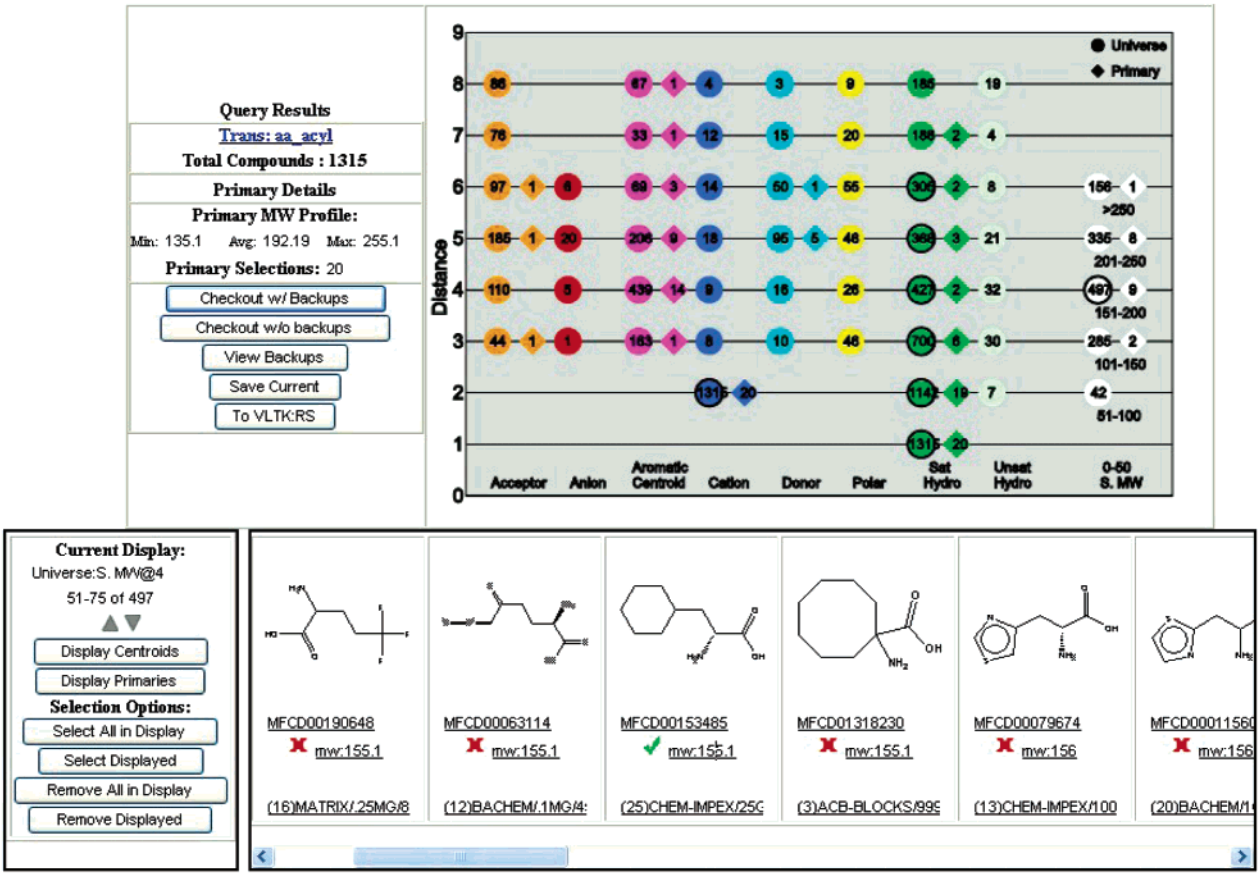


Figure 3. Reagent Selector and Synthon Analysis interface. The chemical properties of the universe of ~1300 amino acids are compared to those of the 20 selected for the SSTR2 library.

RS: SELECTING REAGENTS

Aside from the SA component, other features have been included to facilitate reagent selection. These are the structure display box, the filmstrip and its control panel, and the data center. Each structure display box shows the reagent from which a synthon is derived, its ID tag, synthon molecular weight, and the best source for the compound. Clicking on the molecular weight of a structure causes the chemical types from that synthon to be highlighted in the SA graphic, permitting the user to preview the impact of including it in the primary reagent list. If the user is in the "centroid view" mode, in which only the centroids of the clusters are displayed, a count of the compounds clustered with each centroid structure will also be indicated (and the count acts as a link to display the other members of the cluster). Importantly, a compound can be added to or deleted from the shopping cart simply by toggling the red "X" to the green check.

To maintain a one page display, the structure display boxes are arranged in a gliding "filmstrip" view, which enables the user to see up to 25 reagents beneath the SA graphic before stepping to the next 25 using the "elevator" buttons in the control panel to its left. Initially loaded with the reagent cluster centroids, selecting a chemical type in the SA graphic will load the filmstrip view with the reagents that constitute that group whether from the universe or the shopping cart. For example, the reagents displayed in Figure 3 were loaded into the filmstrip by selecting the fourth molecular weight bucket (151−200) for the universe. The control panel indicates which group is being displayed in the filmstrip and also enables the user to quickly step back to the centroid view mode or a display of the reagents in the shopping cart. Additionally, this panel provides a second way to include or exclude large sets of reagents at one time.

The data center provides live feedback as to the number of reagents selected; their average, maximum, and minimum molecular weight; the number of reagents available; and the synthon transformation. This panel also includes the check-out function as well. The user may elect to proceed with just receiving a list of the primary reagents in the shopping cart mailed back. However, it is important to have a list of secondary reagents that can act as backups should the primary reagents be unavailable or prove to be synthetically intractable. Typically, then, users will have both the list of primary and secondary reagents sent to them, or they will first view the primary and secondary selections in the Backup Interface page and make final adjustments to the reagents. They can delete a primary entirely, reprioritize the order in which backups would be employed should the primary not be available, or even promote a secondary reagent into the primary role before ultimately checking out and receiving the appropriate reagent list via e-mail. A screen capture of the Backup Interface page is shown in Figure 4.

RS: IMPLEMENTATION

One of the requirements in the development of VLTK has been to provide an easily accessible interface for chemists to use during the library design process. Ideally, this tool

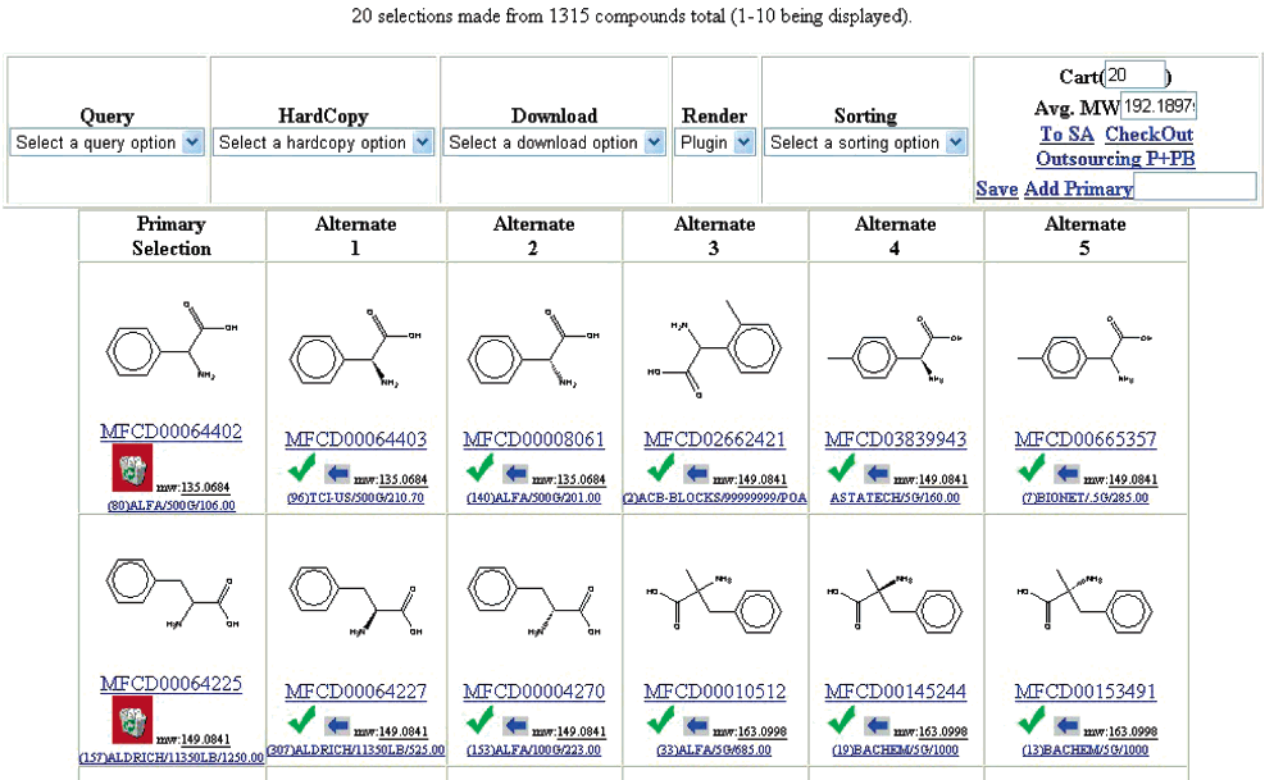


Figure 4. View of the Backup Interface HTML page, which permits final assessment of amino acids selected for the somatostatin agonist library in the first column and the five secondary reagents automatically selected for each primary. Icons beneath each structure are used to make final revisions to these lists.

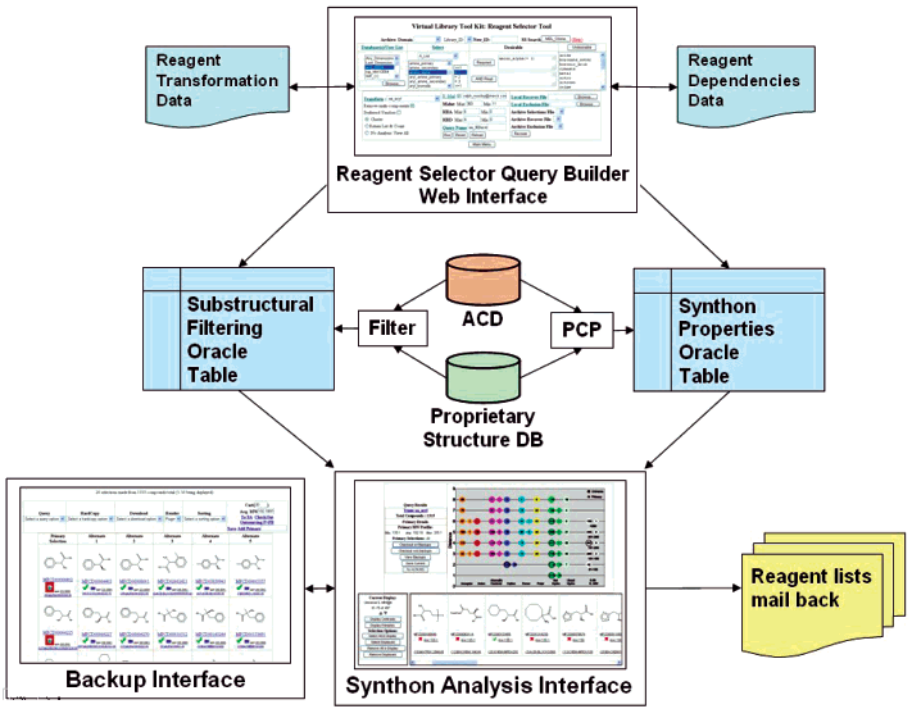


Figure 5. Implementation of Reagent Selector and general workflow.

would not have to be redistributed to the desktop for each update of databases or release of software. Additionally, it would facilitate user acceptance if they already were accustomed to the "look and feel" of the reagent selector. The intranet seemed to provide the best answer for all of these needs. The web is used internally to deliver a variety of content to MRL scientists. The available resources and user acceptance of the web made it a viable option for developing a reagent selection tool. In addition, it enabled immediate access to updated code from the web server, a tremendous advantage during the prototyping phase of the project. The current implementation of the Reagent Selector and Synthon Analysis is depicted in Figure 5.

Initially, the reagent query page was based on a previous, internally developed application for building combinatorial libraries. The current version is driven by two files. One

contains reagent dependencies that "hardwire" reactivity profiles and assist the user in making effective queries. For example, if aldehydes are selected as a desired reagent group, this file is accessed and the more reactive species that undergo similar reactions, such as acylhalides and isocyanates, are added to the "Undesirable" group filter list. Conversely, if a Grignard reagent is desired, this file is accessed and the reagent group that contains metal is removed from the "Undesirable" filter list. The second file contains information about the currently available transformations along with a mapping to a substructural depiction of the reagent transformation. This enables the user to visually inspect the result of the transformation of a template reagent into a synthon prior to performing the synthon analysis step.

During the prototyping phase, it was quickly determined that performing all the operations (filtering substructures, transforming reagents, clustering synthons, and calculating the chemical types) "on the fly" was much too slow for a web-based tool. Our initial solution was to store the results of substructural filters in an Oracle table while maintaining on the fly analysis of the synthons. Although this was effective with small sets of reagents, it failed for larger sets. Ultimately, it became necessary to perform the PCP transformation step and calculate the synthon properties for all viable reagent/transformation combinations beforehand and store these in a second Oracle table. As a result, it was possible to achieve the desired interactive response regardless of the size of the reagent set.

Concurrent with getting data generation, and access to it, under control, we were also optimizing the user friendliness and feature set of the reagent selector. The synthon analysis visualization was initially prototyped using Excel, with which we first developed the circle and diamond plots in Figure 3 using small sets of reagents. A combination of Excel, Visual Basic, and MDL's Excel add-ins were used to further evolve the Synthon Analysis prototype and assess the interactivity and speed of the tool. Although helpful in determining the requirements for using SA, the goal to provide a single user interface eventually precluded continued development in Excel. Ultimately, the Synthon Analysis tool was incorporated into the reagent display and selection page directly. In the current implementation of RS/SA, an HTML style sheet defines placement of the page elements; scalable vector graphics (SVG) supply the chemical types graph; the ChemDraw plug-in provides the reagent structures; and JavaScript controls communication between the SVG graph elements, toggles, buttons, and other page elements, unifying this diverse set of tools into a seamless user interface.

Another concern arose as reagents for designed libraries were being assembled and syntheses started. The availability of reagents and their costs are constantly changing, and some reagents prove to be unexpectedly unreactive. Thus, it became apparent that alternative reagents would have to be identified as backups to the primary reagents. Rather than perform this task on an ad hoc basis months after the library was designed, secondary reagents are identified at the same time that the primary reagents are selected. This is a relatively straightforward process using the nearest neighbor (NN) calculation performed during the Jarvis−Patrick clustering step. Rather than discard the NN data upon completion of synthon clustering, it is kept and used to generate the list of backup reagents for each of the selected reagents upon "check out" from the RS/SA. The check-out process results in an e-mail being sent to the user with a list of the reagents selected and, at their discretion, the list of alternative reagents automatically selected as well. Alternatively, the user may opt to step to the Backup Interface HTML form that uses the ChemDraw plug-in and JavaScript to enable final assessment and refinement of the primary reagents and their alternatives before committing to the check out.

It should be noted here that the development of the RS/SA collection of tools was facilitated by using PowerPoint slides to "storyboard" the anticipated workflow. In this way, it was possible to identify the best practices for a workflow segment, describe the impact of "enhancements" to the evolving tool, and resolve many of the issues inherent in software design prior to committing programming resources. With the assistance of storyboards and prototypes of novel functions, it has been possible to implement the Reagent Selector with the Synthon Analysis interface to provide the speed, interactivity with the user, and "look and feel" desired for an application of this kind. It is used regularly in a production setting to facilitate the selection of reagents targeted toward covering the preferred chemical property space for a particular R-group by the nonspecialist user.

RS: SYSTEM DESIGN

A complete description of VLTK, in which RS/SA is a component, is available elsewhere.[8] Here, we highlight key programs used to produce the RS/SA system described above. The ACD database supplied by MDL is used as a starting point for reagents. Problematic structures that contain isotopes, superatoms, and radicals are removed from the list of potential reagents. Additionally, molecules that contain metals not common in organic synthesis as well as reagents with generic atom types such as "X" or "Y" are discarded. The vendor information available in ACD is also retained, as this will be displayed along with the compounds to assist in reagent selection.

A series of ISIS and Daylight queries have been developed for 136 commonly used reagents, for example, carboxylates, and are the subject of a separate communication.[9] The ISIS queries are run over the modified ACD using the ISISdirect Oracle cartridge to produce lists of reagents for each of the substructure queries. These lists are then augmented to include the number of occurrences for a particular substructure in each reagent using the results of the Daylight queries. The MFCD identification number, substructure counts, vendor information, and molecular weight are loaded into an Oracle table (VLTK) to facilitate performance.

The modified ACD database is processed through a series of ~120 transformations. These are similar to "clipping" operations used in other library enumeration software. Once transformed, the topological distances to pharmacophoric points are calculated. Transformations, property perception, and distance calculations are accomplished using a series of Perl scripts, C code, and chemical patterns encoded in the PCP program. The MFCD number, transformation name, and pharmacophoric property distances are loaded into a second Oracle table (SA) used exclusively by the Reagent Selection process.

On the basis of a series of user specified criteria, as typified in Figure 1, an SQL statement is generated joining the VLTK

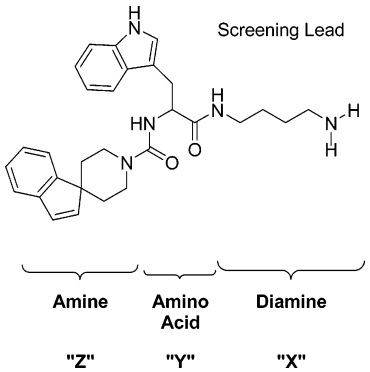


Figure 6. Disconnection of the screening lead into subunits used as the basis for the somatostatin receptor targeted combinatorial library. Twenty diamines and amino acids and 79 amines were used in the basis set of that library.

and SA tables. The reagent_ids identified from the query are used to extract a TDT file. The Daylight clustering toolkit is used to generate a nearest neighbor list and Jarvis−Patrick clustering of the reagents that meet the user specified criteria. The reagent_ids, clustering information, and property distances are processed using a Perl script to produce an XML-based graph suitable for display in the Adobe SVG plug-in as well as the HTML and JavaScript necessary to drive the user interactions. The nearest neighbor information is carried along in the data file and ultimately used to produce alternate reagents to the primary selections if that is desired by the user.

In summary, we rely on the ACD database from MDL, the Daylight toolkit, and the internally developed PCP. The rest of the system is mainly Perl script since the majority of the tasks, after the initial perception and processing of reagents, are text manipulations.

APPLICATION

To better illustrate the application of the RS/SA tool, we have applied it to a previously described combinatorial approach to designing agonists for the somatostatin receptor.[16,17] Discovered through molecular-modeling directed screening of the MRL sample repository, the lead compound was well-suited to the application of combinatorial chemistry with the goal to identify active and subtype selective somatostatin receptor agonists. At the time of the library synthesis, reagents were selected on the basis of known SARs for somatostatin agonists and the availability of reagents via visual inspection of many clusters of potential reagents. Theoretically, 131 670 compounds were prepared in this library of 79 mixtures when stereo- and regioisomers for the 20 diamines, 20 amino acids, and 79 amines are taken into account (see Figure 6). When the RS query builder was used to replicate the requirements for each of the components, the following reagent counts were returned using the ACD (2003.4 version): ∼110 diamines in 20 clusters, ∼1300 amino acids in ∼400 clusters, and >5000 amines in ∼1500 clusters. The query used to identify the amino acids is depicted in Figure 1. Here, we found it necessary to increase the molecular weight cutoff from 175 (the default value) to 300 in order to include all of the amino acids used in the somatostatin library. The "aa_acyl" transformation was used to convert the reagents into synthons. This transform is also a member of the "Capping" transformation group and is depicted in Figure 2 as transformation seven.

Proceeding with the amino acid set to further illustrate the functionality of the RS/SA, the aa_acyl transformation was used to convert the 1315 alpha amino acids into synthons. This transformation tabulates the topological distance to chemical properties starting from the carboxylate carbon. Thus, this universe of amino acids has a cationic basic amine two bond units from that attachment point. We identified the reagents in the "Y" component of the somatostatin library on the synthon analysis page typically using aromatic centroid distances and molecular weights to arrive at the 20 amino acids. The properties of the 20 amino acids selected are shown as counts inside of diamonds in columns adjacent to the counts of the available universe of amino acids shown inside of circles. Aside from the ubiquitous cation, the final reagent selected, cyclohexylalanine, contains a cyclohexyl moiety that extends six bonds from the carboxylate and, therefore, has membership in each of the saturated hydrocarbon bins out to the six bond "distances". It is depicted in the filmstrip view with its properties encircled in the SA graphic shown in Figure 3.

Although one cannot hope to cover the chemical property space of ∼1300 reagents with 20, what becomes obvious with this type of visual feedback is how little actual coverage of property space was achieved with the reagents selected for the amino acid component of this library. From comparing the properties of the selected reagents to those of the universe, it is clear that amino acids containing polar, anionic, and cationic functionalities are not represented at all. This means that any SAR developed for the somatostatin family of receptors or other targets to which this library may be applied in the future will lack important data. An examination of the molecular weight bins shows that all but two synthons fall in the >150 categories. Further, the average molecular weight calculated for these synthons is ∼200 units, with a molecular weight of 255 for the largest reagent, as displayed in the data panel. If such an average molecular weight was common to all three components, this would fall far outside our typical target of an ∼450 average molecular weight for a targeted library. Given that the library used in this example was developed to further the SAR for the somatostatin receptor agonists, the larger size of the components might be more acceptable.

Having completed the selection of reagents, we chose to view the backup reagents before completing the "check out." This final review has been extremely useful to chemists in the past, giving them the option to ensure that the character of the library they are designing is retained in the event that it is necessary to replace a primary reagent with a backup. In some instances, a primary reagent is discarded in favor of a backup that may have been overlooked in the previous page. In this example, the first two amino acids depicted in the screen capture in Figure 4 are phenylglycine and phenylalanine as the primary reagents in the left-most column. The primary reagents are listed in order of increasing molecular weight, and the five backup reagents automatically provided for each using the NN calculation are ordered by similarity to their respective primary. At this point, the chemist might elect to have a specific enantiomer used in the library rather than the racemic amino acid and then "check out" from this page. The lists of primary and backup

reagents are then mailed to him or her. Other files are also automatically sent that are useful for constructing the virtual library or for returning to this analysis should that be necessary.

CONCLUSION

The Reagent Selector component of the intranet-based VLTK has taken what can be a time-consuming process prone to aggravation and error and converted it into one that, with a "point and click," enables the user to quickly select reagents that have a high likelihood of being available at reasonable cost in the bulk quantity needed, in good purity, and are deliverable in a timely fashion. What is unique to this implementation of a nonautomated reagent selector is the focus on the synthon: transformation of reagents, property filtering, and analysis. All of these come together in a graphical display that permits the selection of reagents on the basis of chemical property coverage rather than solely on some measure of dissimilarity or cluster membership. Further, the interactive nature of the Synthon Analysis interface ensures that the user is able to quickly assess the chemical property space covered by a reagent, compare it to the coverage of previous selections, and contrast it to the chemical diversity of the available universe of reagents. Finally, a user is not only provided with a list of the reagents they have selected for their combinatorial library design but also with a list of alternative reagents with similar properties should a desired reagent lack the availability or reactivity desired to complete the library synthesis.

The Synthon Analysis functionality continues to be developed in the context of the Reagent Selector. The graphical interface is being enhanced to allow the user to isolate a subset of reagents on the basis of their membership within multiple chemical property groups using the typical Boolean functions of "OR", "AND", and "NOT." This additional capability facilitates the selection of reagents that meet very specific criteria with regards to the functionality useful, for instance, in developing an SAR-driven library design. Outside of selecting reagents for combinatorial libraries, the synthon analysis concept is being evaluated as another means to visualize the chemical property space for sets of compounds. This could aid in the selection of representatives on the basis of particular chemical properties rather than wholly on similarity measures and provide a novel way to follow up on high-throughput screening results, for example. Finally, we are considering ways in which the synthon analysis concept could be used to visually relate biological assay results to chemical property space and relay to MRL scientists SARs in the context of R-groups off of a scaffold—the original R-group analysis concept brought full circle.

ACKNOWLEDGMENT

The authors thank Mark Mackey, who made the initial foray into designing a reagent selector at our MRL—Terlings Park, U.K. site. We also thank our collaborators in the Chemical Data department who have supplied us with many "cleaned-up" structural databases as a starting point for VLTK and RS, our group members in the Molecular Systems department who have supported our efforts and provided helpful suggestions to improve performance and the feature set, and our user community, which has borne the brunt of evolving software and found the results to be worth the wait.

Supporting Information Available: The list of 17 undesirable reagent groups set by default in the reagent selector. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES AND NOTES

(1) Blaney, J. M.; Martin, E. J. Computational approaches for combinatorial library design and molecular diversity analysis. *Curr. Opin. Chem. Biol.* **1997**, *1*, 54−59.
(2) Brown, R. D.; Hassan, M.; Waldman, M. Combinatorial library design for diversity, cost efficiency, and drug-like character. *J. Mol. Graph. Model.* **2000**, *18*, 427−437.
(3) Polinsky, A.; Feinstein, R. D.; Shi, S.; Kuki, A. LiBrain: software for automated design of exploratory and targeted combinatorial libraries of special interest. In *Molecular Diversity and Combinatorial Chemistry*; Chaiken, I. M., Janda, K. D., Eds.; American Chemical Society: Washington, DC, 1996; pp 219−232.
(4) For example, MDL's Reagent Selector, MDL Information Systems, Inc., San Leandro, CA.
(5) Jamois, E. A.; Hassan, M.; Waldman, M. Evaluation of Reagent-Based and Product-Based Strategies in the Design of Combinatorial Library Subsets. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 63−70.
(6) Martin, E. J.; Blaney, J. M.; Siani, M. A.; Spellmeyer, D. C.; Wong, A. K.; Moos, W. H. Measuring Diversity: Experimental Design of Combinatorial Libraries for Drug Discovery. *J. Med. Chem.* **1995**, *38*, 1431−1436.
(7) Brown, R. D.; Martin, Y. C. Use of Structure−Activity Data to Compare Structure-Based Clustering Methods and Descriptors for Use in Compound Selection. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 572−584.
(8) Feuston, B. P.; Chakravorty, S. J.; Conway, J. F.; Culberson, J. C.; Forbes, J.; Kraker, B.; Lennon, P. A.; Lindsley, C.; McGaughey, G. B.; Mosley, R.; Sheridan, R. P.; Valenciano, M.; Kearsley, S. K. Web enabling technology for the design, enumeration, optimization and tracking of compound libraries. *Curr. Topics Med. Chem.* **2005**, accepted for publication.
(9) Kraker, B.; Chakravorty, S. J.; Mosley, R. T.; Culberson, J. C.; Feuston, B. P.; Sheridan, R. P.; Conway, J. F.; Forbes, J. K.; Kearsley, S. K. Virtual Library Tool Kit: Comparison of functional group queries with Daylight and ISIS software, manuscript in preparation.
(10) Bush, B. L.; Sheridan, R. P. PATTY: A programmable atom type and language for automatic classification of atoms in molecular databases. *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 756−762.
(11) Miller, M. D.; Kearsley, S. K.; Underwood, D. J.; Sheridan, R. P. FLOG: a system to select quasi-flexible ligands complementary to a receptor of known three-dimensional structure. *J. Comput.-Aided Mol. Des.* **1994**, *8*, 153−174.
(12) Butina, D. Unsupervised Data Base Clustering Based on Daylight's Fingerprint and Tanimoto Similarity: A Fast and Automated Way to Cluster Small and Large Data Sets. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 747−750.
(13) Holliday, J. D.; Jelfs, S. P.; Willett, P. Calculation of Intersubstituent Similarity Using R-Group Descriptors. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 406−411.
(14) Bemis, G. W.; Murcko, M. A. Properties of Known Drugs. 2. Side Chains. *J. Med. Chem.* **1999**, *42*, 5095−5099.
(15) Gasteiger, J.; Rudolf, C.; Sadowski, J. Automatic generation of 3D-atomic coordinates for organic molecules. *Tetrahedron Comput. Methodol.* **1990**, *3*, 537−547.
(16) Rohrer, S. P.; Birzin, E. T.; Mosley, R. T.; Berk, S. C.; Hutchins, S. M.; Shen, D.-M.; Xiong, Y.; Hayes, E. C.; Parmar, R. M.; Foor, F.; Mitra, S. W.; Degrado, S. J.; Shu, M.; Klopp, J. M.; Cai, S.-J.; Blake, A.; Chan, W. W. S.; Pasternak, A.; Yang, L.; Patchett, A. A.; Smith, R. G.; Chapman, K. T.; Schaeffer, J. M. Rapid identification of subtype-selective agonists of the somatostatin receptor through combinatorial chemistry. *Science* **1998**, *282*, 737−740.
(17) Berk, S. C.; Rohrer, S. P.; Degrado, S. J.; Birzin, E. T.; Mosley, R. T.; Hutchins, S. M.; Pasternak, A.; Schaeffer, J. M.; Underwood, D. J.; Chapman, K. T. A Combinatorial Approach toward the Discovery of Non-Peptide, Subtype-Selective Somatostatin Receptor Ligands, *J. Comb. Chem.* **1999**, *1*, 388−396.

CI050090X