REALISIS: A Medicinal Chemistry-Oriented Reagent Selection, Library Design, and Profiling Platform

Aziz Yasri,*,† Didier Berthelot, Harry Gijsen, Theo Thielemans, Patrick Marichal, Michael Engels, and Jan Hoflack

Johnson & Johnson Pharmaceutical Research and Development, a division of Janssen Pharmaceutica N.V., Turnhoutseweg 30, B-23 40 Beerse, Belgium

Received April 7, 2004

REALISIS is a software system for reagent selection, library design, and profiling, developed to fit the workflow of bench chemists and medicinal chemists. Designed to be portable, the software offers a comprehensive graphical user interface and rapid, integrated functionalities required for reagent retrieval and filtering, product enumeration, and library profiling. REALISIS is component-based, consisting of four main modules: reagent searching; reagent filtering; library enumeration; and library profiling. Each module allows the chemist to access specific functionalities and diverse filtering and profiling mechanisms. By implementing the entire process of reagent selection, library design, and profiling and by integrating all the necessary functionalities for this process, REALISIS cuts the time required to design combinatorial and noncombinatorial libraries from several days to a few hours.

INTRODUCTION

The introduction of combinatorial chemistry and high-throughput screening[1−16] was expected by pharmaceutical and biotech companies to deliver significant improvements in both the efficiency and productivity of the drug discovery process. Several factors have combined to prevent the realization of those expectations. First, pharmaceutical companies are traditionally conservative and want to identify the best system before investing heavily in what are still relatively new concepts. Second, at the research chemistry level, exploration of the chemistry space tends to be undertaken either in a very restricted manner due to the limited availability of building blocks or in a broad fashion that does not take into consideration knowledge of the biological target, diversity criteria, or early pharmacokinetic characteristics. Furthermore, reagent selection is still a largely intuitive process and is sometimes limited to the use of in-house building blocks.

Reagent selection is one of the most critical steps in synthesis planning and library design. Availability, synthetic feasibility, structural or property similarity/dissimilarity, and cost are among the criteria applied by chemists to select reagents for a specific library or single compound synthesis. These criteria often determine the success or failure of lead generation or optimization programs. For reagent selection, chemists often spend hours filtering and classifying their in-house or commercial reagent lists for synthesis.

Some of the published research on computational aspects of reagent selection and library design focuses on the problem of designing diverse libraries or selecting diverse sets of compounds.[17−22] However, it is generally recognized that diversity alone is rarely sufficient in the lead optimization phase of the drug discovery process.[23−29] Chemists could improve their chances of designing more active compounds if they incorporated relevant information about the target at an early stage in the library design process.[30,43]

As well as incorporating information about the target being studied, the library design should also include some "drug-like" properties and good pharmacokinetic behavior characteristics by making use of in silico predictive ADME/tox models.[10,24,25,28,31,32] Early assessment of the library not only increases the likelihood of achieving better biological potency but also ensures that any potential hit from the assay possesses a good safety profile and represents a promising start point for lead optimization.

The main objective of our work was to design a tool that allows the medicinal chemist to achieve an appropriate degree of chemical diversity while simultaneously incorporating existing knowledge about the biological target and ensuring safety characteristics in a designed library. We called the resulting system REALISIS (Reagent And Library Selection In Silico). In addition to offering useful functionalities that support the chemist's workflow, REALISIS is designed along component-based lines, ensuring ease of use and integration with other in-house and third party commercial software.

In this paper, we outline briefly the component-based architecture of the software and describe its functionalities, which range from reagent searching to library enumeration and profiling. These functionalities are illustrated by the design of an amide library that uses primary amines and acid chlorides extracted from the ACD database. In a second paper, we will describe the integration of the REALISIS program with other in-house and third party commercial software, and we will emphasize how the integration is done and how it facilitates the access to virtual screening tools and in silico ADME/Tox predictions needed by medicinal chemists for rational library design.

* Corresponding author phone: +33 4 67 70 93 88; fax: +32 14 60 5344; e-mail: yaziz_2000@yahoo.fr.

† Present address: Molecular and Statistical Mining Consulting, 1 avenue des Venus, F-34170 Castelnau-le-lez, France.

10.1021/ci049879i CCC: $27.50
Published on Web 10/26/2004

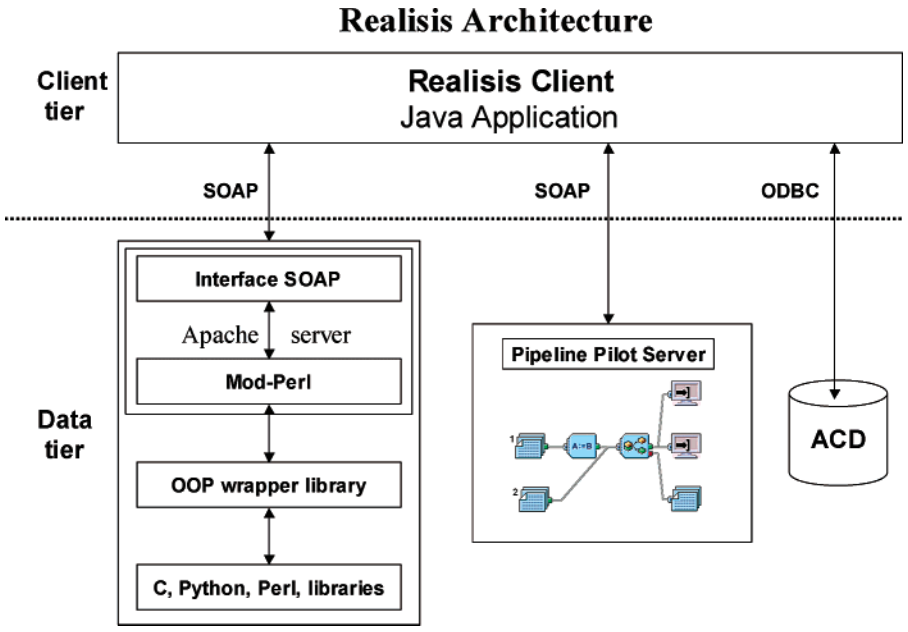

Figure 1. Top level view of REALISIS architecture. The client tier is based on a Java application using a direct connection to the data tier using SOAP and ODBC protocol. The data tier contains third party and tailored modules processing business logic. These modules are written in different programming languages (C, Python, Perl).

SYSTEM ARCHITECTURE OVERVIEW

The system is designed in a simple two-tier structure: a client tier and a data tier (Figure 1). From the client side, the Java programming language (Java 2 Platform, Standard Edition, J2SE) was used to implement the logic and the presentation of the software. The data tier groups the calculations and database servers.

Two main calculation servers were used to integrate and provide all existing computational capabilities of molecular modeling and chemo-informatics from in-house and third party commercial software: (1) Pipeline Pilot,[33] installed on an NT server, provided the client with all protocols designed by molecular modelers and chemo-informatics scientists and (2) an object-oriented wrapper library in which all in-house programs and commercial software are wrapped and exposed to other systems.

From the database side, the data tier integrates the ACD monomers database[34] by connection to Isis-Direct via ODBC protocol.

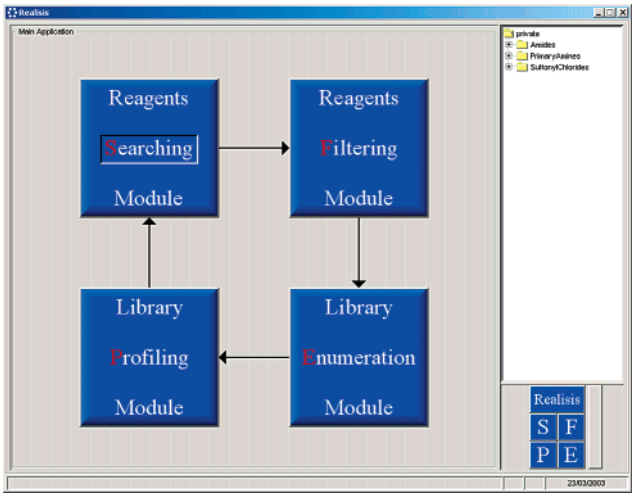

Figure 2. REALISIS main application window. The window shows the four available interconnected modules presenting the iterative process of reagent selection and library design. It also presents a hierarchical view in a tree of stored results and folders. A navigator bar is also presented in the main application allowing the user to go from a module to another.

LIBRARY DESIGN WORKFLOW, REALISIS MODULES, AND FUNCTIONALITIES

To implement the entire library design workflow, the REALISIS system was implemented in four main modules: reagent searching; reagent browsing and filtering; library enumeration; and library profiling (as shown in the program main application illustrated by Figure 2). These four modules implement the complete workflow of the chemist, from reagent retrieval to reagent purchasing, when designing a whole library or a small set of compounds.

Reagent Searching Module. In this module, the user can search and retrieve reagents (monomers) both in-house and in commercially available databases such as ACD (Figure 3). The system can search more than one reagent database simultaneously and merges the results in one unique file.

For query drawing, we have integrated a conventionally available drawing package[35] that provides a variety of drawing capabilities such as those offered by Isis-Draw[36] and Chem-Draw.[37] The searching mechanism, which consists of a substructure search used widely by chemists, is performed in a multithread fashion (more than one query at a time). Results from each query are collected in an SDF file format and stored in a private user directory provided by the software in the navigation tree. The user can also exclude any unwanted functionality or fragment from the query by drawing a substructure in the "exclude" area and can also restrict the query by stipulating molecular weight property.

This reagent searching module can import, save, and copy/paste queries drawn using other drawing software. These may vary from simple fragment queries (primary amine, acid chloride, halide, etc.) to multiple disconnected substructures (primary amine including an aromatic ring not necessarily

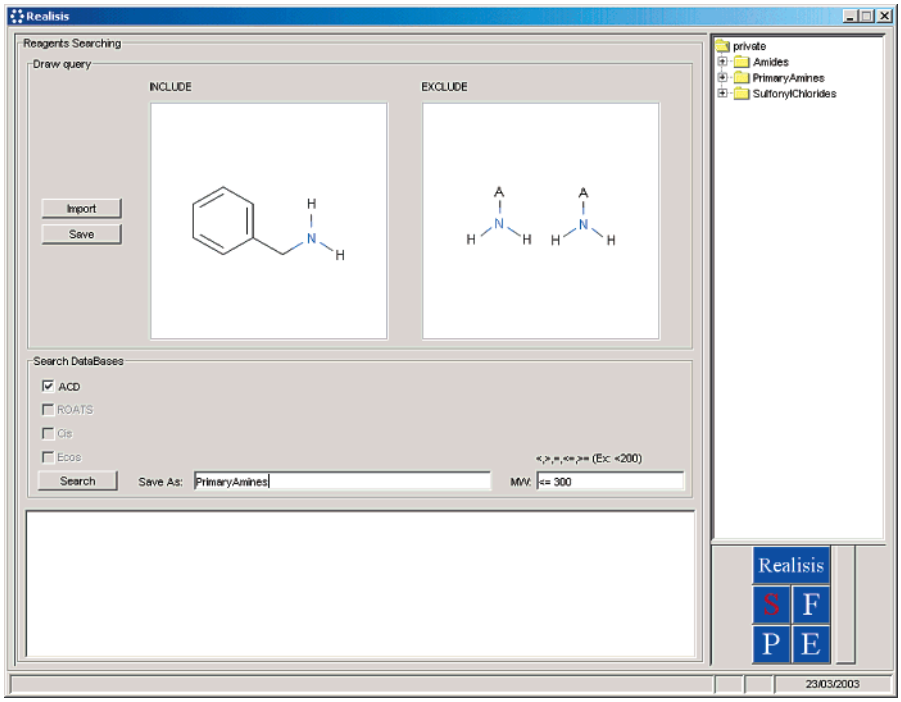

Figure 3. Reagent searching module window. The input to this module is presented in three types. (1) Substructure query drawing areas where the drawing tool is invoked. An example is given where a benzyl-amine fragment is drawn in the "include" area and two disconnected primary amine fragments are drawn in the "exclude" area. The results from this query exclude any benzyl-amine reagent containing two primary amine fragments. (2) The database area where all available reagent databases are presented for selection by the user. (3) Molecular weight property restriction. The results of the searching process are stored in an SDF file provided by the user.

connected to the nitrogen atom). The MarvinSketch drawing component integrated in this module provides the user with more sophisticated query-building using SMARTS language.[38]

Reagent Filtering Module. The reagent search may identify a huge list of monomers with undesirable lists of atoms (isotopes, metals) and fragments, especially when no restriction is applied to the query. It is important that these unwanted reagents are removed from the list before considering candidates for library enumeration or reagent purchasing. The REALISIS system provides a set of functionalities that can be used to filter out unwanted reagents by applying a range of criteria discussed below.

Chemistry Cleaning. Three kinds of chemicals are cleaned from the initial pool of reagents: (1) salts are stripped from the molecules; (2) any monomer containing an isotope is removed; and (3) any monomer showing a metal in its nonsalt part is removed.

Unwanted Fragments. A set of 35 carefully selected fragments is presented, from which users can choose. These fragments concern chemicals such as acetals, acyl-halides, aldehydes, alkyl-halides, anhydrides, azides, aziridines, oxiridines, imines, isocyanates, Michael acceptor, nitrates, nitro, peroxides, etc. The user can remove any monomer that includes unwanted fragments by making the appropriate selections from this list. To allow the maximum degree of flexibility, the system also enables users to choose additional fragments.

Duplicate Checking. Before chemists consider a list of reagents for purchase, they always check the existence of duplicates and attempt to keep a unique list of monomers. This can be achieved manually by visual inspection where the number of monomers is quite small (less than 50 reagents) but becomes time-consuming when lists run to several hundreds or thousands of reagents. REALISIS allows the user to submit small or large lists of reagents to check for duplicates, providing a unique list of monomers and identifying duplicates within a few seconds.

The "check duplicates" functionality is based on an exact match between the reagent structures. In some circumstances, the chemist may consider as duplicates reagents with different functional groups but the same core. When used in a chemical reaction, these may lead to the same product, as, for example, in the use of acid chlorides and acid bromides in an amide formation reaction. This requirement is taken into consideration when duplicate checking is performed using REALISIS.

Physicochemical Property Filtering. A set of simple physicochemical filters is provided, enabling the chemist to narrow down the number of reagents before proceeding with library enumeration. Molecular weight and flexibility index estimated by the number of single rotatable bonds can be used to eliminate any monomer which could lead to highly flexible undesired products or entities with a molecular weight of more than 700−800. Other more sophisticated filtering mechanisms (number of hydrogen bond donor and acceptor, logP, Lipinski's rule-of-five) are also available in this module. While it is more appropriate to use this kind of filter on products rather than reagents, narrowing down the number of reagents can be advantageous at this stage when large amounts of data need to be processed during subsequent steps.

Diversity Selection. Based on Pipeline Pilot fingerprint and maximum dissimilarity index, REALISIS provides a simple interface for selecting a diverse monomer subset from a large and nondiverse list. This filtering mechanism is especially useful to limit the number of reagents to an amount which can be handled during synthesis.

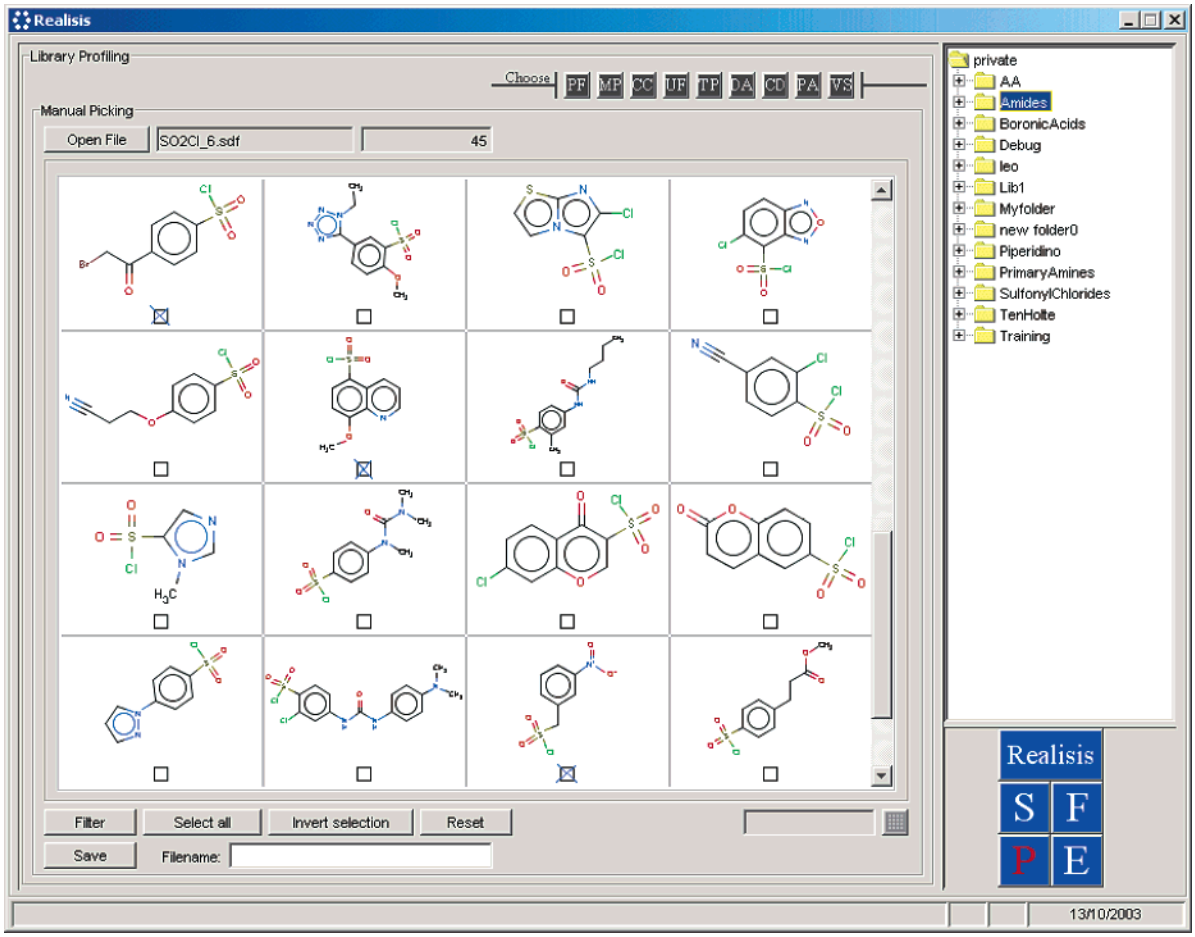

Figure 4. Reagent filtering and browsing module window. A toolbar with text tips contains shortcut buttons of all filtering mechanisms. Manual picking showed here allows the user to visualize a list of reagents and manually selects a subset of monomers.

Manual Selection. REALISIS also allows for manual reagent selection through visual inspection (Figure 4). This functionality is used to visualize the results of any filtering mechanism in the reagent filtering module or the retrieved set of reagents from the reagent searching module. It is also useful when the chemist is interested in a limited number of reagents that can be selected manually.

Library Enumeration Module. Enumeration of virtual libraries is a common task in combinatorial and noncombinatorial library design, undertaken in order to create the core product by connecting reagent cores. Drawing applications are used to create the product structures when a small number of products is being considered, but it is impractical to construct and store all possible products from larger reagent lists and to create extensive virtual libraries. In such cases there is a clear need for an automated tool to conduct a task that would be time-consuming for the chemist to perform manually.

Two main algorithms have been proposed to deal with the problem of virtual library enumeration: fragment marking or the Markush approach[39,40] (Figure 5−A) and the reaction transform approach[41] (Figure 5−B).

The first is based on the use of a Markush structure, which represents a common scaffold with variation sites labeled as R-groups. In this case, the virtual library is assembled by systematic attachment of clipped reagents to the respective variation sites. Although enumeration is reduced to a simple concatenation of the corresponding connection tables, the lists of clipped reagents must be carefully constructed from the monomers by removing the parts of the structure that are discarded during the reaction. Unfortunately, this "fragment marking" approach cannot be easily applied to reactions that involve modification of the building blocks such as the Diels−Alder reaction.[41]

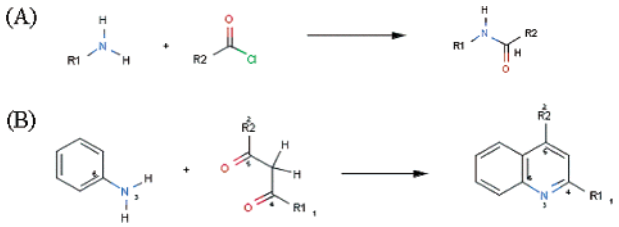

Figure 5. Schematic illustration of fragment-marking or Markush approach (A) and the reaction transform approach (B) for virtual libraries enumeration.

The second approach uses a reaction transform to encode the chemical reaction.[41,42] The transform specifies which parts of the reacting molecules undergo chemical transformations and the nature of these transformations. This approach mimics more closely the stages involved in actual chemical synthesis, does not require the presence of a common template and the generation of clipped reagents, and can be applied to a broad spectrum of chemical reactions used in combinatorial chemistry. To be applicable at a broad level, the encoding process should be able to accommodate multicomponent reactions, ring cyclizations, removal of

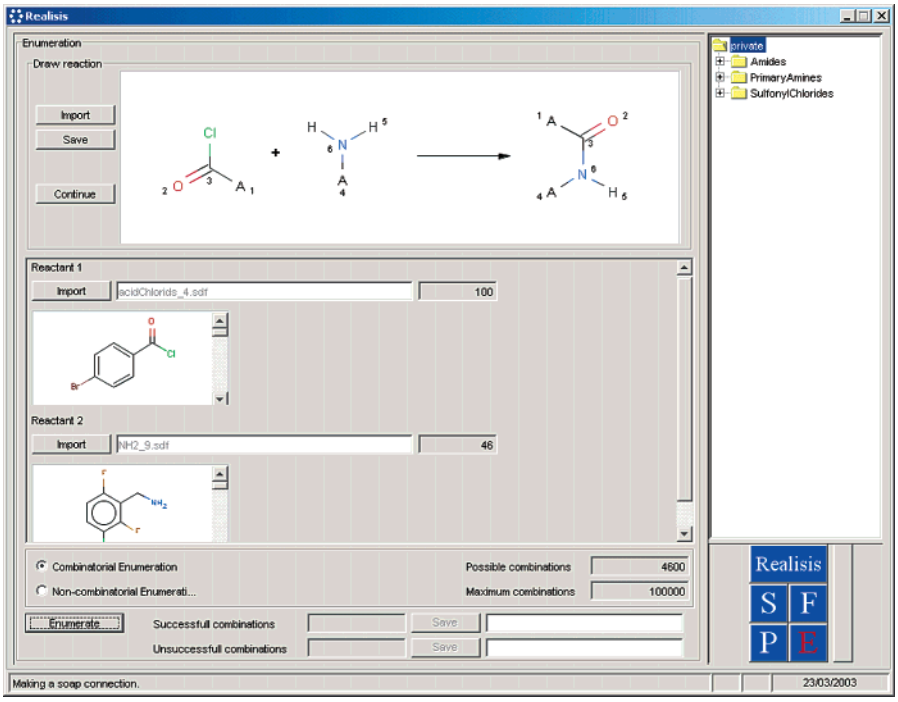

Figure 6. Library enumeration module window. This module is also used to enumerate a two-component library or multicomponent library and in cases where a simple transformation is required (protection, deprotection, etc.). Once the reaction transform is drawn or imported, the number of reagent lists required for the enumeration is set automatically. Combinatorial or noncombinatorial methods are then used to enumerate the products. Here, 100 acid chlorides and 46 primary amines are used to enumerate 4600 amide products.

protecting groups, and modification of the core structure. In addition, it should offer a means of specifying multiple products, designating stereochemistry, and differentiating the reactivity of different functional groups depending on their environment if more than one of these groups match the reaction requirements.

REALISIS uses the reaction transform approach for virtual library enumeration (Figure 6). The library enumeration module takes as input the reaction transform drawn by MarvinSketch or Isis-Draw and the list of reagents involved in the reaction. When a reagent list is uploaded into the application, the user can still browse visually through it before performing the enumeration. This module also allows the chemist to perform both combinatorial and noncombinatorial enumeration of the product. In the second scenario, the chemist decides which monomers from one list need to be combined with monomers in the other reagent list. This noncombinatorial enumeration is used most commonly in cases where small numbers of compounds have to be designed.

The enumeration procedure within REALISIS returns two categories of results to the user: (1) products resulting from successful combinations of the reagents according to the reaction transform and (2) a set of combinations for which the reaction transform does not succeed in combining the reagents. This nonsuccessful set of combinations can be used to debug the reaction or to remove those reagents whose structures do not fit into the reaction scheme (for example, where reactions involve acid bromides but where acid chlorides are provided by the user).

Library Profiling Module. After the library has been enumerated, the user can reduce library size further by applying the library profiling module. This offers chemists all the filtering mechanisms provided by the reagent filtering module (see 3.2) as well as profiling tools that can only be applied to enumerated products.

Toxicophores Removal. A set of 38 toxicophore fragments is presented in order to allow filtering out of compounds that could present a potential toxicity problem. These fragments were nominated by pharmacokinetics and medicinal chemistry experts as those encountered most frequently in highly toxic compounds. Examples include aziridines, thioepoxides, oximes, hydroquinones, etc.

Physicochemical Profiling. A list of physicochemical properties is available to the chemist for library profiling. These simple filters concern logP, MW, TPSA, HBD, and HBA. Once these descriptors have calculated, minimum and maximum values for each property are displayed to the user (Figure 7). Histograms for the calculated properties can be turned on by the user in order to profile the library. One interesting feature of this filtering module is that it allows the user to filter the library by tuning the minimum and the maximum values for each property separately or for all properties sequentially. Whenever these values are modified, the library is filtered out, and property boundaries and histograms are updated accordingly.

It is worth noting that, to offer users the flexibility to set thresholds for each property at desired levels, this filtering module does not present any rule-of-five filter. This is an important feature since, for some projects, chemists still wish to consider molecules even when they violate some of Lipinski's rules.

As with the reagent filtering module, the library profiling module allows the user to switch from one filtering mechanism to another in the order of their choice, until a library of reasonable size is obtained.

Reagent Ordering. Reagent ordering is the final step in library design. In this module the user submits a set of

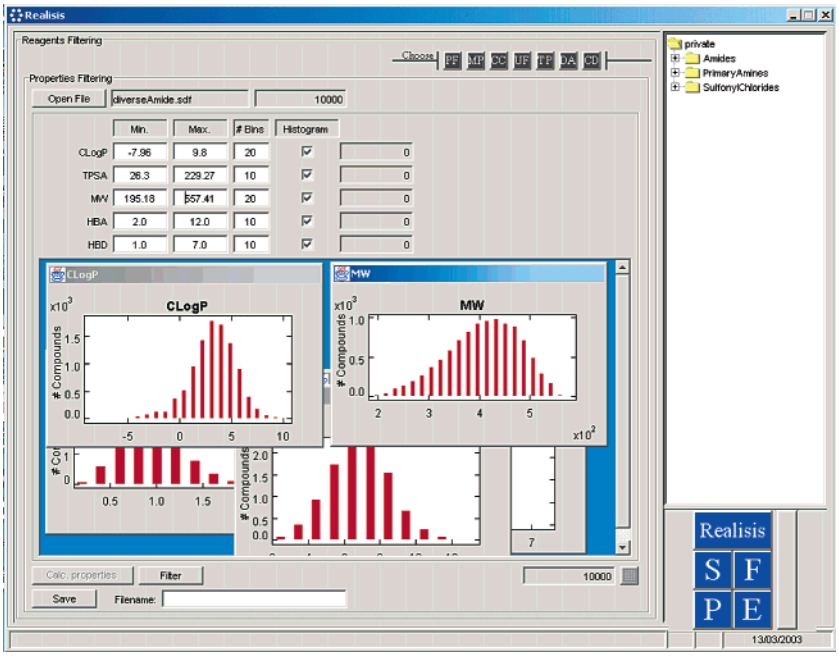

Figure 7. Library profiling module window. This module possesses a generic interface showing all the filtering and profiling tools for virtual libraries. Histogram plots are used to show graphical distributions of calculated physicochemical properties. The boundaries of each property are used as filtering tools to reduce the size of the library.

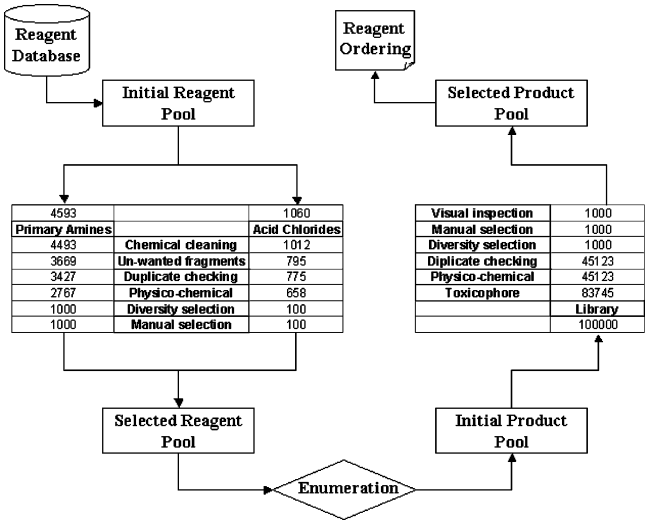

Figure 8. Schematic flowchart of reagent selection, library design, and profiling process within REALISIS. Here, the workflow is illustrated by reagent selection of required monomers for designing an amide library of 1000 compounds.

products in order to extract the reagents from which they are made. Spreadsheets for each reagent are generated following submission of the unique list of monomers, detailing names, database sources, and the frequency with which each reagent occurs in the relevant products. The spreadsheets are then used to order the reagents.

REALISIS Workflow. Figure 8 shows a flowchart detailing each step in the reagent selection and library profiling process provided by the REALISIS application. An initial and crude pool of starting materials is reduced to a final set of reagents to be used in the enumeration procedure. Chemical and physicochemical filters are applied during this process, primarily to accommodate restrictions imposed by reaction conditions (unwanted fragment removal, bifunctional groups, etc.) but also to ensure chemical space diversity. The enumerated library (initial product pool) is subjected to the same filtering scenario until a reasonable number of products is obtained.

APPLICATION TO AN AMIDE LIBRARY DESIGN

In designing an amide library, we looked for primary amines and acid chloride reagents from the ACD database. The following constraints were applied during the search: (1) the primary amines list should exclude any monomer with more than one primary amine function and (2) the acid chloride list should not contain any primary amine fragment. This ACD database search identified 1060 acid chlorides and 4593 primary amines (Figure 8).

These two initial reagent pools were then submitted to the reagent filtering module in the order shown in Figure 8 (left-hand side). Chemical cleaning of the two monomer lists returned 4493 primary amines and 1012 acid chlorides. We then applied the unwanted fragments filter, which reduced numbers to 3669 primary amines and 795 acid chlorides. Application of the check duplicates filter resulted in unique lists of 775 acid chloride and 3427 primary amine reagents. To ensure some diversity in the starting material, we then selected diverse subsets of 1000 primary amines and 100 acid chlorides. No manual selection was performed within the two sets of primary amines and acid chlorides at this stage.

The two sets of reagents required to build the amide library and which passed the reagent filtering steps (1000 primary amines and 100 acid chlorides) served to enumerate a combinatorial amide library of 100000 possible products. The enumeration procedure in REALISIS resulted in 100000 successful amide products.

The enumerated combinatorial amide library was then submitted to the library profiling module where the following filters were applied (Figure 8, right-hand side). The toxicophore fragments filter reduced compound numbers from

100000 to 83745. Subsequent application of the drug likeness filter, where ClogP < 5, MW < 500, HBD < 10, and HBA < 5 (Lipinski's rule of five) resulted in a further reduction, to 45123 compounds, all of which were retained by the check duplicates filter (not a surprising result in view of the fact that all starting material lists were unique). A set of 1000 amide compounds was then selected by applying the diversity filter, followed by final visual inspection of a few products.

CONCLUSION

In this paper we have described an efficient and rapid tool for reagent selection, library design, and profiling that fits the workflow of chemists. As part of the REALISIS design process, interviews were conducted with bench, medicinal, and computational chemists in order to canvass opinions on the functionalities and graphical user interface characteristics required of such a tool. REALISIS has been tested by a range of users who report that the system is easy to work with, thanks to the user-friendly nature of the graphical user interface and functionalities that are part of their daily workflow. Furthermore, no intensive training is required to use the system.

By implementing the entire process of reagent selection, library design, and profiling and by integrating all of the functionalities required in that process, REALISIS reduces the time spent by chemists on these tasks from several days to a few hours. Filtering and reactant property calculations take between 1 and 2 h, while the enumeration process takes an hour at most, depending on the size and complexity of the library. A further advantage of the system is that the bulk of these periods are essentially "machine time" rather than "chemist or computational chemist time". Once the enumeration process is complete, the time required for library profiling depends on the particular objectives of the chemist.

Widespread acceptance of the REALISIS software by bench chemists and medicinal chemists (70% of users from medicinal chemistry) has helped to remove part of the burden from computational chemistry experts, enabling them to focus on more complex problems. It has also gives chemists the opportunity to assess their ideas rapidly and broaden exploration of the chemistry space.

ACKNOWLEDGMENT

We wish to thank Lieven Meerpoel for the fruitful discussion on interfacing REALISIS with reagent ordering and tracking processes within Johnson & Johnson PRD. Hans Vanderzijl and Peter Bosmans wrote the Java front-end component of REALISIS, while Tom Tabruyn developed the component-based physicochemical filtering tools. We thank all Johnson & Johnson colleagues who contributed to the process of identifying user requirements that played a significant part in the development of REALISIS. We also thank the many other colleagues who have provided valuable feedback on the system, which has improved significantly its utility and applicability within Johnson & Johnson Pharmaceutical Research and Development.

REFERENCES AND NOTES

(1) Beeley, N.; Berger, A. A revolution in drug discovery. Combinatorial chemistry still needs logic to drive science forward. *BMJ [Br. Med. J.]* **2000**, *321*, 581−582.
(2) Broach, J. R.; Thorner, J. High-throughput screening for drug discovery. *Nature* **1996**, *384*, 14−16.
(3) Cox, B.; Denyer, J. C.; Binnie, A.; Donnelly, M. C.; Evans, B.; Green, D. V.; Lewis, J. A.; Mander, T. H.; Merritt, A. T.; Valler, M. J.; Watson, S. P. Application of high-throughput screening techniques to drug discovery. *Prog. Med. Chem.* **2000**, *37*, 83−133.
(4) Floyd, C. D.; Leblanc, C.; Whittaker, M. Combinatorial chemistry as a tool for drug discovery. *Prog. Med. Chem.* **1999**, *36*, 91−168.
(5) Flygare, J. A.; Sutherlin, D. P.; Brown, S. D. Combinatorial chemistry in steroid receptor drug discovery. *Methods Mol. Biol.* **2001**, *176*, 353−358.
(6) Hall, S. E. The future of combinatorial chemistry as a drug discovery paradigm. *Pharm. Res.* **1997**, *14*, 1104−1105.
(7) Hogan, J. C., Jr. Combinatorial chemistry in drug discovery. *Nat. Biotechnol.* **1997**, *15*, 328−330.
(8) Krstulovic, A. M. High-throughout screening in combinatorial chemistry for drug discovery. *J. Chromatogr. B Biomed. Sci. Appl.* **1999**, *725*, 1.
(9) Li, W. Z.; Yun, L. H. Effective paradigms in drug discovery: integration of combinatorial chemistry with rational drug design. *Yao Xue Xue Bao* **1998**, *33*, 710−716.
(10) Salemme, F. R.; Spurlino, J.; Bone, R. Serendipity meets precision: the integration of structure-based drug design and combinatorial chemistry for efficient drug discovery. *Structure* **1997**, *5*, 319−324.
(11) Seneci, P.; Miertus, S. Combinatorial chemistry and high-throughput screening in drug discovery: different strategies and formats. *Mol. Divers.* **2000**, *5*, 75−89.
(12) Shoemaker, R. H.; Scudiero, D. A.; Melillo, G.; Currens, M. J.; Monks, A. P.; Rabow, A. A.; Covell, D. G.; Sausville, E. A. Application of high-throughput, molecular-targeted screening to anticancer drug discovery. *Curr. Top. Med. Chem.* **2002**, *2*, 229−246.
(13) Tanaka, A. Combinatorial chemistry for drug discovery. *Tanpakushitsu Kakusan Koso* **2000**, *45*, 887−894.
(14) Thorpe, D. S. Forecasting roles of combinatorial chemistry in the age of genomically derived drug discovery targets. *Comb. Chem. High Throughput Screening* **2000**, *3*, 421−436.
(15) Toledo-Sherman, L. M.; Chen, D. High-throughput virtual screening for drug discovery in parallel. *Curr. Opin. Drug Discov. Devel.* **2002**, *5*, 414−421.
(16) White, R. E. High-throughput screening in drug metabolism and pharmacokinetic support of drug discovery. *Annu. Rev. Pharmacol. Toxicol.* **2000**, *40*, 133−157.
(17) Blaney, J. M.; Martin, E. J. Computational approaches for combinatorial library design and molecular diversity analysis. *Curr. Opin. Chem. Biol.* **1997**, *1*, 54−59.
(18) Graham, E. T.; Jacober, S. P.; Cardozo, M. G. A novel frequency distribution selection method for efficient plate layout of a diverse combinatorial library. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 1508−1516.
(19) Krchnak, V.; Weichsel, A. S.; Cabel, D.; Lebl, M. Linear presentation of variable side-chain spacing in a highly diverse combinatorial library. *Pept. Res.* **1995**, *8*, 198−205.
(20) Kubota, H.; Lim, J.; Depew, K. M.; Schreiber, S. L. Pathway development and pilot library realization in diversity-oriented synthesis: exploring Ferrier and Pauson-Khand reactions on a glycal template. *Chem. Biol.* **2002**, *9*, 265−276.
(21) Stavenger, R. A.; Schreiber, S. L. Asymmetric Catalysis in Diversity-Oriented Organic Synthesis: Enantioselective Synthesis of 4320 Encoded and Spatially Segregated Dihydropyrancarboxamides. *Angew. Chem., Int. Ed. Engl.* **2001**, *40*, 3417−3421.
(22) Ward, B.; Juehne, T. Combinatorial library diversity: probability assessment of library populations. *Nucleic Acids Res.* **1998**, *26*, 879−886.
(23) Andrews, K. M.; Cramer, R. D. Toward general methods of targeted library design: topomer shape similarity searching with diverse structures as queries. *J. Med. Chem.* **2000**, *43*, 1723−1740.
(24) Barn, D.; Caulfield, W.; Cowley, P.; Dickins, R.; Bakker, W. I.; McGuire, R.; Morphy, J. R.; Rankovic, Z.; Thorn, M. Design and synthesis of a maximally diverse and drug-like screening library using REM resin methodology. *J. Comb. Chem.* **2001**, *3*, 534−541.
(25) Brown, R. D.; Hassan, M.; Waldman, M. Combinatorial library design for diversity, cost efficiency, and drug-like character. *J. Mol. Graph Model* **2000**, *18*, 427−37, 537.
(26) Mason, J. S.; Cheney, D. L. Library design and virtual screening using multiple 4-point pharmacophore fingerprints. *Pac. Symp. Biocomput.* **2000**, 576−587.
(27) Mason, J. S.; Beno, B. R. Library design using BCUT chemistry-space descriptors and multiple four-point pharmacophore fingerprints: simultaneous optimization and structure-based diversity. *J. Mol. Graph Model* **2000**, *18*, 438−51, 538.
(28) Winkler, D. A.; Burden, F. R. Application of neural networks to large dataset QSAR, virtual screening, and library design. *Methods Mol. Biol.* **2002**, *201*, 325−367.

(29) Zheng, W.; Cho, S. J.; Waller, C. L.; Tropsha, A. Rational combinatorial library design. 3. Simulated annealing guided evaluation (SAGE) of molecular diversity: a novel computational tool for universal library design and database mining. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 738−746.
(30) Leach, A. R.; Bradshaw, J.; Green, D. V.; Hann, M. M.; Delany, III, J. J. Implementation of a system for reagent selection and library enumeration, profiling, and design. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 1161−1172.
(31) Dixon, S. L.; Villar, H. O. Bioactive diversity and screening library selection via affinity fingerprinting. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 1192−1203.
(32) Ghose, A. K.; Viswanadhan, V. N.; Wendoloski, J. J. A knowledge-based approach in designing combinatorial or medicinal chemistry libraries for drug discovery. 1. A qualitative and quantitative characterization of known drug databases. *J. Comb. Chem.* **1999**, *1*, 55−68.
(33) Stevenson, J. M.; Mulready, P. D. Pipeline Pilot 2.1. *J. Am. Chem. Soc.* **2003**.
(34) Available Chemical Directory, MDL Information Systems, Inc. http://www.mdli.com/pdfs/ACD2000.2.pdf 2003.
(35) MarvinSketch 2.10.5. ChemAxon Ltd. http://www.chemaxon.com/marvin 2003.
(36) Isis-Draw 2.4. MDL Information Systems, Inc. http://www.mdli.com/products/acd.html. 2003.
(37) Chem-Draw. CS ChemDraw Pro. Chemical Strucrure Drawing Program. CambridgeSoft.com Life Science Enterprise Solutions. http://www.cambridgesoft.com. 2003.
(38) Daylight theory manual, chapter 4. Daylight Chemical Information Systems Ltd. Santa Fe, NM. http://www.daylight.com/dayhtml/doc.
(39) Barnard, J. M.; Downs, G. M.; von Scholley, P.; Brown, R. D.; Barnard, J. M. Use of Markush structure analysis techniques for descriptor generation and clustering of large combinatorial libraries. *J. Mol. Graph Model.* **2000**, 452−63.
(40) Barnard, J. M. A comparison of different approaches to Markush structure handling. *J. Chem. Inf. Comput. Sci.* **1991**, *31*, 4−8.
(41) Harper, G.; Bradshaw, J.; Gittins, J. C.; Green, D. V.; Leach, A. R.; Leach, A. R.; Bradshaw, J.; Green, D. V.; Hann, M. M.; Delany, J. J. Prediction of biological activity for high-throughput screening using binary kernel discrimination. *J. Chem. Inf. Comput. Sci*. **2001**, 1295−300.
(42) Leach, A. R.; Hann, M. M. The in silico world of virtual libraries. *Drug Discovery Today* **2000**, *5*, 326−336.
(43) Tounge, A. B.; Reynolds, H. C. Abstracts, 36th Middle Atlantic Regional Meeting of the American Chemical Society, Princeton, NJ, United States, June 8−11 2003, American Chemical Society, Washington, DC CODEN: 69EBDT Conference.

CI049879I