JOURNAL OF CHEMICAL INFORMATION AND MODELING

ARTICLE

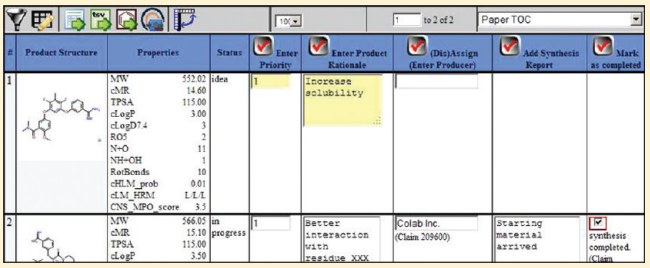
pubs.acs.org/jcim

DEGAS: Sharing and Tracking Target Compound Ideas with External Collaborators

Man-Ling Lee,* Ignacio Aliagas, Jennafer Dotson, Jianwen A. Feng, Alberto Gobbi, and Timothy Heffron

Genentech Inc., South San Francisco, California 94080, United States

ABSTRACT: To minimize the risk of failure in clinical trials, drug discovery teams must propose active and selective clinical candidates with good physicochemical properties. An additional challenge is that today drug discovery is often conducted by teams at different geographical locations. To improve the collaborative decision making on which compounds to synthesize, we have implemented DEGAS, an application which enables scientists from Genentech and from collaborating external partners to instantly access the same data. DEGAS was implemented to ensure that only the best target compounds are made and that they are made without duplicate effort. Physicochemical properties and DMPK model predictions are computed for each compound to allow the team to make informed decisions when prioritizing. The synthesis progress can be easily tracked. While developing DEGAS, ease of use was a particular goal in order to minimize the difficulty of training and supporting remote users.



INTRODUCTION

Modern small molecule drug discovery poses many challenges. Before a compound is selected for clinical studies, many properties need to be optimized, such as activity, selectivity, and safety. The process is a multiparameter optimization in which many compounds have to be made and evaluated in regards to all their properties. Each compound synthesis and testing is associated with costs in terms of time, resources, and money. Another major challenge is that today drug discovery is often conducted by geographically separated teams. Decision making is complicated by difficulties in communication and by the differences in time zones. Improving the collaborative decision process and reducing the cycle time have an impact on the likelihood of success and the cost of a drug discovery project.

A number of pharmaceutical companies have presented their compound profiling applications in the past. The features of these applications vary greatly. The most common features are:

- enumeration of combinatorial libraries[1−4]
- computation of physicochemical properties and ADME/Tox models[1−5]
- graphical data analysis[1−5]
- access to in-house databases and reagent management systems[1,2,4]

The applications let users evaluate and prioritize compounds based on their properties and chemical structure. Some provide access to in-house reagent inventories to support synthesis planning.

A complementary approach to support decision making is to assemble expert knowledge in knowledge management systems. Astra Zeneca and Actelion adopted Wiki-technology for their applications CODD[6] and ActWiki,[7] respectively. PFAKT[8] from Pfizer and ROCK[9] from Roche were implemented from scratch to meet the specific requirements for knowledge submission, revision, processing, and presentation. The knowledge captured varies widely by application: CODD and PFAKT focus on capturing compound-specific knowledge and the status of virtual compounds, while ActWiki and ROCK focus on capturing more general medicinal chemistry knowledge.

Knowledge management and compound profiling applications help to foster collaborative teamwork. However, with members of a drug discovery team operating from different sites, the aspect of instant sharing has become important. Several publications describe software applications which help sharing information and coordinating efforts. Microsoft's SharePoint[10] has become a popular framework for shared content management. This framework lets team members share documents on the web. Examples include C-ME[11] and OnePoint.[12] Collaborative Drug Discovery (CDD) has developed a proprietary web platform. It aims at bringing scientists from all over the world together using concepts from social networking.[13]

A significant number of the small molecule drug discovery projects at Genentech are collaborations with external partners in which Genentech scientists interact closely with scientists at the collaborating sites. To facilitate transparent decision making, we have introduced DEGAS[14] to the internal and external users. Target compound ideas are entered and are instantly available to all other team members. The DEGAS compound identifier provides a short unique name that can be used when referring to specific compounds in discussions. Team members at the different sites make joint decisions on which molecules should be synthesized. They also need to track synthesis progress. Storing

Special Issue: 2011 Noordwijkerhout Cheminformatics

Received: July 16, 2011
Published: November 12, 2011

target compound ideas and their calculated profiles along with team-based synthetic prioritization is a prerequisite for a transparent and well-documented decision process. Exposing a subset of Genentech's registration database allows for the verification of compound novelty and availability.

This paper describes the implementation and use of DEGAS. The application combines features from compound profiling and knowledge management into a collaborative platform. DEGAS's web-based interface was designed for ease-of-use in order to keep the effort on training and support of internal and external users low. The use of a common software platform between Genentech scientists and collaborators greatly simplifies the communication by providing all users with access to a database with target compound ideas and computed physicochemical properties. Rules to restrict data access to specific partners or projects can be added to the application's configuration. This meets the secure access requirements posed by the Genentech Informatics department for deployment outside of Genentech's firewall.

DEGAS was intentionally designed to minimize restrictions, allowing each project to adjust the application to their project-specific needs and business rules. Compound ideas can be entered as single compounds or as enumerated combinatorial libraries. Physico-chemical properties and drug metabolism and pharmacokinetics (DMPK) property predictions are automatically computed. Having determined the optimal thresholds of the properties of interest using applications independent of DEGAS, project teams can prioritize or deprioritize compounds with optimal or unfavorable in the data browser. Optionally, rationales can be entered for each target compound to facilitate team-based prioritization. The outcome of the prioritization can be stored. Synthesis of a compound can be assigned to a chemist, and reports on the status of the synthesis can be added. The status of the compounds, i.e., "idea", "in progress", "completed", or "discontinued", is deduced from the stored user inputs. The status of compounds is searchable allowing the users to easily separate the most relevant compounds for review. The search and reporting interface is highly configurable, which allows for complex queries and for reports over single or multiple projects.

Integration with Genentech's registration database enables novelty checking. If a proposed target compound exists in the registration database, then the integration with the sample database allows users to determine compound availability. The novelty check prevents duplication of work, while the computed properties and the rationales enable project teams to better assess proposed target compounds.

DEGAS COMPONENTS

The bases for DEGAS are two highly flexible software platforms, i.e., Pipeline Pilot and AEREA. Pipeline Pilot[15] is a commercial, workflow-oriented informatics platform and is used by Genentech's computational chemistry group to implement scientific programs. AEREA[16] is an open source database browser that can be customized to different data models and offers interfaces to implement data entry widgets. The accessibility of both platforms through web browsers allows a seamless integration via dynamic links. The relational database connects the two principal components (Figure 1). By leveraging existing Pipeline Pilot protocols and the AEREA infrastructure, the effort to develop DEGAS was low (0.25 FTE). Three developers were able to release the initial version to the pilot project team after two months of development. Three months after the initial release, DEGAS was revised to incorporate user feedback, and only minor changes have been made since.

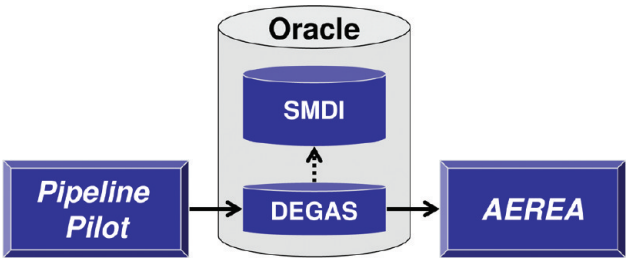


Figure 1. Database connects Pipeline Pilot and AEREA. A Pipeline Pilot protocol is used for entering compound ideas. AEREA serves as the data browser and allows users to add information for individual target compounds. The DEGAS database is linked to Genentech's registration database (SMDI) to flag known compounds.

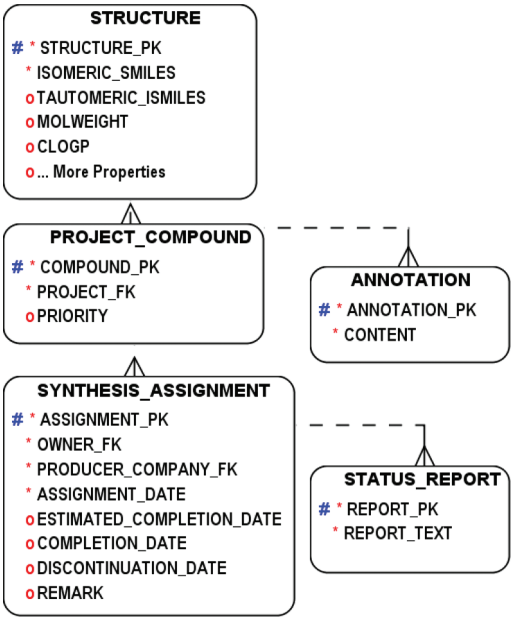


Figure 2. Main entities in the relational data model used to store the information for DEGAS. Only the columns relevant to the discussion are shown. Columns ending on "_FK" are foreign keys to other tables. Columns ending on "_PK" designate primary key columns.

Data Model. DEGAS data are stored in a relational data model (Figure 2). The main entities are the STRUCTURE and PROJECT_COMPOUND. The STRUCTURE table contains information about the chemical structures and their properties. This table includes the column TAUTOMERIC_ISMILES for storing normalized SMILES.[17] The normalized SMILES in DEGAS and Genentech's registration (SMDI) database enable the integration of the two databases through using an exact match of the normalized SMILES strings. The PROJECT_COMPOUND table contains project-specific information, thus, allowing multiple projects to profile the same structure in their context. The target rationales and other comments on project compounds are stored in the ANNOTATION table.

The one-to-many relationship between the PROJECT_COMPOUND and SYNTHESIS_ASSIGNMENT tables allows recording of the initial and any additional scale-up syntheses. The SYNTHESIS_ASSIGNMENT table contains columns for the

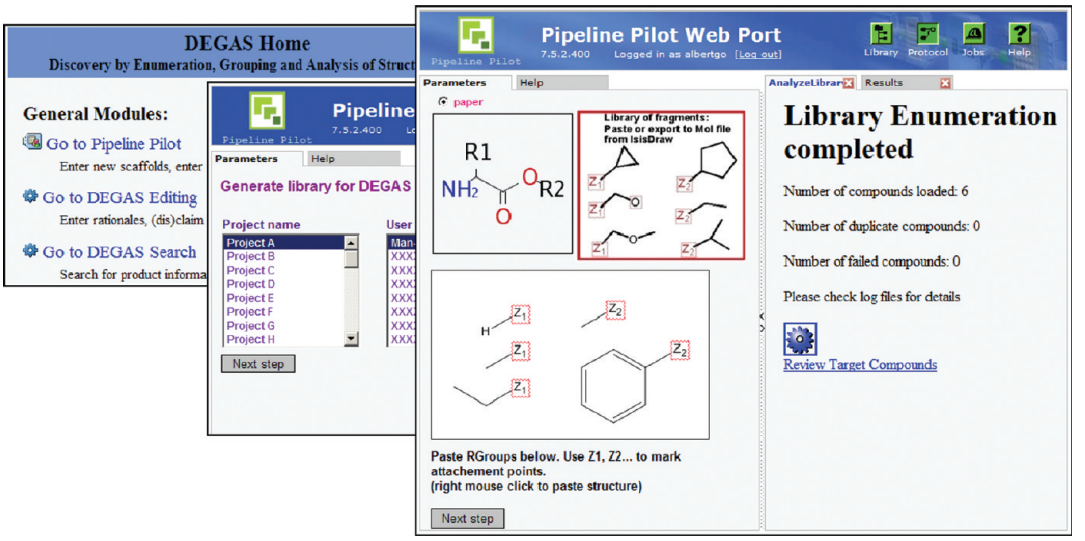


Figure 3. Main user interfaces (UI) for submission of a combinatorial library: DEGAS entry UI (left). Project and user selection UI (center). Library definition UI (right). After the enumeration, the outcome of the submission is reported to the user. A link to AEREA ("review target compounds") provides a seamless integration.

assignment owner and producer company as well as estimated completion, completion, and discontinuation dates. Comments on progress of the synthesis are stored in the STATUS_REPORT table.

The columns for estimated completion, completion, and discontinuation dates are used to determine the synthetic status of a target compound. Only one of the three columns is specified at any given time. A compound is "in progress" if the estimated completion date is given. If it is completed, the estimated completion date is removed, and the completion date is specified. If a synthesis is "discontinued", the estimated completion date is replaced by the discontinuation date. A compound without any assignment record is classified as an "idea".

The producer company can differ from the company of the assignment owner. The owner is the chemists responsible for the synthesis within the project. This feature enables the chemistry leads to track synthesis outsourced to noncollaborating contract research organizations (CRO).

Not shown in Figure 2 are the tables for computed properties. They are linked via a foreign key to the STRUCTURE_PK column in the STRUCTURE table. There is no other requirement on the format or the number of the tables for computed properties. Both the Pipeline Pilot protocol and AEREA can be easily adjusted to the addition of new properties or the removal of out-dated ones.

Pipeline Pilot. Prior to the implementation of DEGAS, protocols for library enumeration, molecular property calculation,[18−22] prediction of DMPK properties,[23] and docking were already available at Genentech. The DEGAS compound submission module was built as a Pipeline Pilot web port application to take advantage of these already well established protocols at Genentech.

A link in the DEGAS home page leads the user to a Pipeline Pilot protocol for target compound submission. The user selects a project and a user name. Allowing the selection of a username enables a chemist to submit compounds for a colleague. In the second step, the chemists can submit single compounds or a combinatorial library (Figure 3). Users can also pre-enumerate combinatorial libraries and upload the structures as sdf files.

Once target compounds are submitted, the structures are normalized and loaded into the database. The structures are normalized using the same module[17] as compounds in Genentech's registration database, thereby, ensuring consistent novelty checking. After inserting the structures into the database, Pipeline Pilot reports the outcome of the submission to the user and provides a link to the AEREA database application. At the same time, the computation of molecular properties and the DMPK property predictions are initiated as an asynchronous process. This approach has the advantage that users can start to review the compounds while the property calculations are performed. When the property calculation is completed, AEREA displays the new values immediately.

The Pipeline Pilot submission web port protocol is also accessible via a web service from PyMol[24] and Benchware3D Explorer,[25] two end-user modeling applications in use at Genentech. This allows the capturing of compound ideas as an integrated step in the design of new compounds. New compound ideas with interesting predicted binding modes can be seamlessly recorded in DEGAS. The integration employs the interfaces provided by the modeling applications. In both cases, the application saves the active molecule to the network file system and opens the Pipeline Pilot web port protocol by opening a URL in a web browser. The protocol reads the compound information from the network file system.

AEREA. AEREA[16] was designed to be configurable to any relational data model. The data model is described as a graph of connected data nodes. This follows the concept described by Hewitt et al.[26] Each data node represents a set of related fields, e.g., the fields in the PROJECT_COMPOUND table (Figure 2). Besides the DEGAS database content, the current configuration exposes Genentech's compound identifier and registration date as well as the amount of available compound sample.

AEREA provides interfaces for implementing chemical searches and data entry columns.[27] This allows AEREA to be extended for interactive data entry, enabling users to input data in the context of other information. The data entry columns can be either text or check boxes. For DEGAS, we have created a set of six columns for data entry (Figure 4). The two columns "Enter Product Rationale" and "Enter Priority" allow the team to justify the work

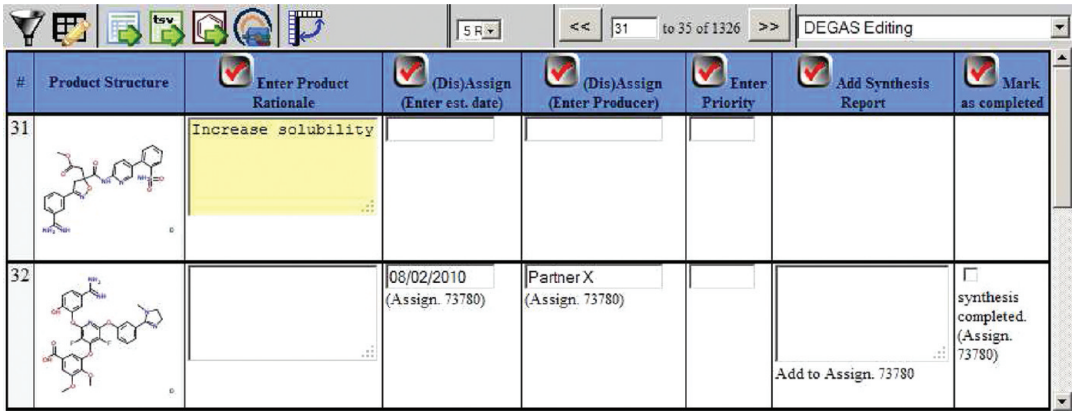


Figure 4. An example report with all columns for interactive data entry. The AEREA configuration contains instructions regarding what should be displayed in the input cell depending on the database content. For example, no text and check box are displayed in the "Add Synthesis Report" and "Mark as Completed" columns for the first compound because its synthesis has not been assigned to any chemist. To create an assignment, a chemist enters the estimated completion date in the "(Dis)Assign (enter est. date)" column. The assignment is immediately created after the submission. (Structures have been modified to protect confidential information).

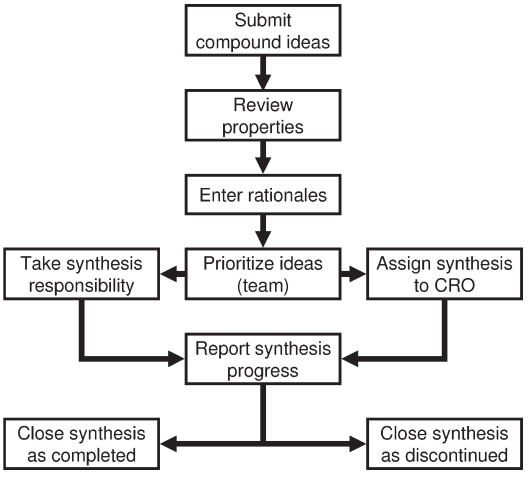


Figure 5. Example workflow. If not specifically noted in parentheses, the action is usually performed by individual chemists. Some of the non-collaborating CROs have access to DEGAS versions showing only compounds whose synthesis is assigned to them. In those cases, chemists at the CROs can only enter progress updates and close synthesis assignments.

required to synthesize the proposed target and to rank compounds, respectively. The four columns "(Dis)Assign (enter producer)", "(Dis)Assign (enter est. date)", "Add Synthesis Report" and "Mark as Completed" allow the team to indicate the progress of compound synthesis. Additionally, there are columns for deleting project compounds, synthesis assignments, and reports.

Figure 5 shows an example workflow. First the designing chemist reviews the properties of new compounds and enters rationales for whether or not they should be made. The team sets the priorities and records them in the "Enter Priority" column. Subsequently, chemists enter the estimated completion date, thereby, taking responsibility for the synthesis of the given compounds. Alternatively, chemists can specify a name into the "Producer Company" column. This captures that the chemist is responsible for the synthesis which will be done at the given CRO. If the synthesis is completed, the chemists check the box in the "Mark as Completed" column. If a synthesis is abandoned, the chemists remove the content from the "Estimated Completion Date" or "Producer" text boxes. In both cases, the date columns in the SYNTHESIS_ASSIGNMENT table are updated as described in the Data Model Section.

AEREA provides a query builder (Figure 6) and report designer.[28] The list of searchable fields and reportable columns is identical except for the data entry columns, which are only available in the report designer. Queries and report templates can be stored for public access or for private use. In general, each project has a predefined project query form and report template. The query form contains the most frequently used search constraints and can be easily modified. By selecting or deselecting the box next to a field, users can include or exclude constraints. Moreover, users can add constraints from the field list.

In the second release of DEGAS, an additional check box column was implemented, allowing users to submit compounds to the Pipeline Pilot protocol for fully automated docking. Specific docking configurations are created by the computational chemists for their project. This check box column enables the user to submit a set of compounds from DEGAS to the docking protocol. Within the protocol, the user selects a docking configuration and initiates the computation. The user and other team members can review the docking results in AEREA. The docking poses can be accessed via links from AEREA to PyMol session files.

DEGAS Access by Collaborators. DEGAS has been designed with the objective of permitting access by authorized external users. Each external user may only access the subset of data for which they have the privilege. The infrastructure must also prevent any unauthorized, accidental, or intentional access to the Genentech intranet. The DEGAS application server instance for each collaborator is, therefore, encapsulated from the Internet, from the Genentech intranet, and from DEGAS instances for other collaborators.

Figure 7 gives the schematic overview of the infrastructure that enables DEGAS access by external users. A dedicated JBoss web server[29] is created for each collaborator. The user authentication and authorization are handled by the Juniper server.[30] It checks the credential of the user in LDAP.[31] The LDAP user information describes the access privilege, and the Juniper server uses this information to route the user to the collaborator web server. The JBoss servers are installed on a Linux server located outside the

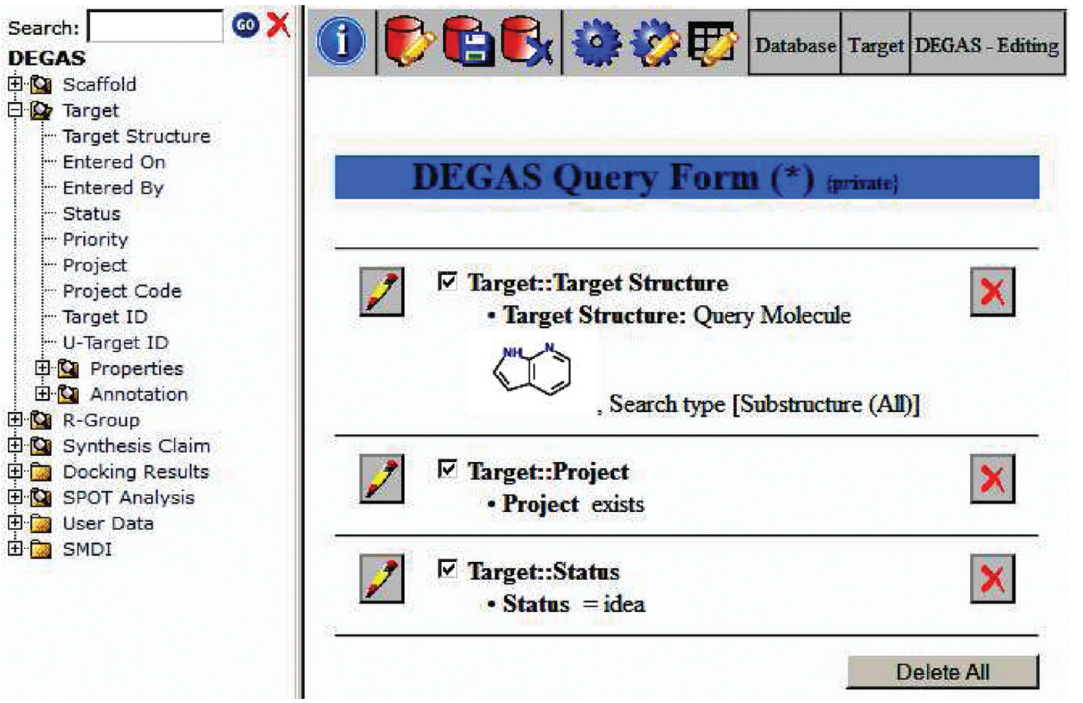


Figure 6. Default DEGAS query form. Projects modify this form to generate customized query forms. The query shown in the figure will find compounds containing the given substructure with status "idea" from any project. Shown on the left is the list of searchable fields.

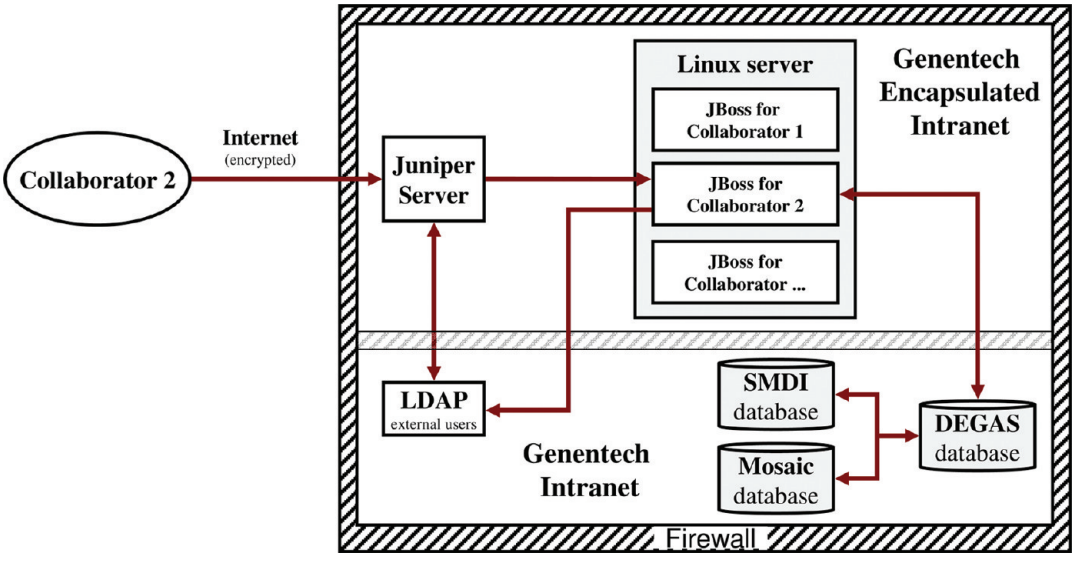


Figure 7. Schematic overview of the infrastructure enabling collaborators to access Genentech internal applications, such as DEGAS. Genentech Informatics department provides the Juniper and LDAP servers for user authentication and authorization. The Linux server has only restricted access to the Genentech intranet.

Genentech intranet but still within Genentech's fire wall. For security reasons, no application on the web servers has access to directories inside the intranet, and only specific network protocols are allowed into the intranet. Files on the web server are retrieved by programs running on computers within the intranet. Database access is restricted to the DEGAS tables and a small subset of the tables in Genentech's registration database. By using a dedicated web server for each collaborator the setup of each server can be validated, preventing accidental access to privileged data. It also facilitates the configuration of the applications on each server.

AEREA can be configured to further restrict data access. Project restrictions are added to the configuration files (Figure 8). As a result, each AEREA instance provides the given collaborator a customized view that can be easily modified. The Pipeline Pilot web port protocols, on the other hand, do not have this capability. For partners, we have replaced the top-tier Pipeline Pilot protocol for compound submission by a simple web form. The form allows

```
<table alias="prj" name="smdi_projects" required="y">
   <join type="inner" on="lprd.project_fk = prj.project_id"/>
   <constraint sql="prj.project_name IN ('PROJECT_1','PROJECT_2')"/>
</table>
```

Figure 8. A project restriction in an example AEREA configuration. A project name is only added to the constraint element if the given project team requests DEGAS access for its collaborating team.

external users to upload sdf files with new compounds and to associate them with their username and project. Upon submission, the underlying Pipeline Pilot protocol for computing the molecular and DMPK properties is executed.

DISCUSSION

Although DEGAS enables users to input their knowledge, it is primarily designed to support the daily process of drug discovery teams. All inputs in AEREA are optional, and there are only a small number of business rules imposed by the application. The project teams appreciate this flexibility because it allows each team to adjust DEGAS according to the project's scope. Usually the chemistry team leader decides with one or two team members how the team will use DEGAS and communicates the standard operating procedure (SOP) to the rest of the team.

Since its first release, DEGAS' popularity as a collaborative management application for chemistry resources has steadily increased. The popularity stems from the fact that the inputs and computed properties are immediately available to other users. Hence, all team members always work with the same up-to-date data regardless of their geographic location.

The pilot project team consisted of two equally sized subteams collaborating on compound design and prioritization. The first subteam was located at Genentech, while the second collaborating subteam was located overseas. In addition, compounds were synthesized at a noncollaborating CRO. In effect, there were three equally sized teams, across three continents. The team started to use DEGAS about one and a half years before the deadline for the clinical candidate selection. At that time, senior team members had identified that a reduction in lipophilicity should improve the overall attractiveness of synthesized compounds (solubility, plasma protein binding, and metabolic stability). Prior to using DEGAS, all chemists were required to compute physicochemical properties and report them to leaders before initiating synthetic effort. With the implementation of DEGAS, team members were asked to submit target compounds, to evaluate the computed properties, and to state the rationale for the synthesis of that compound. When assigned to the synthesis of a compound, the chemists were requested by the project lead to give the estimated completion date considering the synthetic effort. This estimate was updated periodically.

DEGAS expedited the progress of the chemistry team by reducing the time spent on generating and collecting synthetic target ideas from all chemists. Furthermore, DEGAS allowed leaders of the large distributed team real-time access to current activities. This access assured that there were no overlapping efforts at different sites, that all targets were addressing current team goals, and that the highest priority targets were being addressed. With DEGAS, the project and chemistry leaders could review anticipated compound properties and their synthetic status and could reassign chemists to work on different target compounds.

Recently, DEGAS was used for tracking the commercial compounds repurchased for the high-throughput screening (HTS) hit confirmation and structure−activity relationship (SAR) expansion process. In this case, a synthesis assignment represents a compound order. To handle the larger number of compounds, a module was implemented that simultaneously submits the compounds and creates corresponding synthesis assignments. The web interface allows users to upload SDFiles with structures, rationales, and purchase information. The rationale field allows the team to state the reason of the acquisition. The estimated delivery date is stored in the estimated completion date field.

The current implementation of DEGAS requires users to close synthesis assignments by hand, using the "Mark as Completed" column. This is too time-consuming for large compound sets. The updates have, therefore, not been performed consistently. This has reduced the usefulness of DEGAS as a tool for tracking purchase orders. It has led to the following planned improvements to implement the automated closing of purchase orders when the ordered compound is registered into Genentech's registration database and to extend the existing input columns to support bulk operations, so that large orders can be processed efficiently.

CONCLUSION

We have designed and implemented DEGAS, a flexible database application, for storing target compound ideas, profiling their properties, and managing chemistry resources. The application is installed on secure web servers and allows our partner scientists on different continents access to project specific data, thus, improving collaborative team work. The software development time was kept very short by reusing the pre-existing Pipeline Pilot protocols and by leveraging the open source software AEREA for data access. AEREA was configured to display DEGAS data and to allow users to enter data into the database. Early and late stage projects have taken advantage of the application to prioritize targets, track synthesis progress, and manage chemistry resources. DEGAS has proven to be particularly useful for large teams working at separate locations. Our next step is to extend the DEGAS functionalities for supporting projects in ordering HTS compounds for hit confirmation and follow-ups. An extension is necessary to deal with the larger number of compounds.

AUTHOR INFORMATION

Corresponding Author

*E-mail: lee.man-ling@gene.com.

ACKNOWLEDGMENT

We thank the chemists in the pilot project team for their feedback as well as Nick Skelton for his feedback on using DEGAS for tracking purchased compounds. We also acknowledge Kevin P. Clark and the Genentech Informatics team for their commitment to implementing a secure infrastructure to expose DEGAS to our many collaborators.

REFERENCES

(1) Oprea, T. I.; Gottfries, J.; Sherbukhin, V.; Svensson, P.; Kühler, T. C. Chemical information management in drug discovery: optimizing the computational and combinatorial chemistry interfaces. *J. Mol. Graph. Modell.* **2000**, *18*, 512–524.

(2) Agrafiotis, D. K.; Alex, S.; Dai, H.; Derkinderen, A.; Farnum, M.; Gates, P.; Izrailev, S.; Jaeger, E. P.; Konstant, P.; Leung, A.; Lobanov, V. S.; Marichal, P.; Martin, D.; Rassokhin, D. N.; Shemanarev, M.; Skalkin, A.; Stong, J.; Tabruyn, T.; Vermeiren, M.; Wan, J.; Xu, X. Y.; Yao, X. Advanced Biological and Chemical Discovery (ABCD): Centralizing Discovery Knowledge in an Inherently Decentralized World. *J. Chem. Inf. Model.* **2007**, *47*, 1999–2014.

(3) Leach, A. R.; Bradshaw, J.; Green, D. V. S.; Hann, M. M.; Delany, J. J. Implementation of a System for Reagent Selection and Library Enumeration, Profiling, and Design. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 1161–1172.

(4) Yasri, A.; Berthelot, D.; Gijsen, H.; Thielemans, T.; Marichal, P.; Engels, M.; Hoflack, J. REALISIS: A Medicinal Chemistry-Oriented Reagent Selection, Library Design, and Profiling Platform. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 2199–2206.

(5) *Stardrop*, version 5.0; Optibrium, Ltd.: Cambridge UK, 2011.

(6) Robb, G. Hypothesis-driven drug design using wiki-based collaborative tools. Presented at the UK-QSAR and ChemoInformatics Group Meeting (Online); Pfizer: Sandwich, U.K., May 14, 2009; http://www.ukqsar.org/slides/Grobb_Ukqsar_May09.pdf (accessed October 1, 2011).

(7) Sander, T.; Freyss, J.; von Korff, M.; Reich, J. R.; Rufener, C. OSIRIS, an Entirely in-House Developed Drug Discovery Informatics System. *J. Chem. Inf. Model.* **2009**, *49*, 232–246.

(8) Brodney, M. D.; Brosius, A. D.; Gregory, T.; Heck, S. D.; Klug-McLeod, J. L.; Poss, C. S. Project-Focused Activity and Knowledge Tracker: A Unified Data Analysis, Collaboration, and Workflow Tool for Medicinal Chemistry Project Teams. *J. Chem. Inf. Model.* **2009**, *49*, 2639–2649.

(9) Mayweg, A.; Hofer, U.; Schnider, P.; Agnetti, F.; Galley, G.; Mattei, P.; Lucas, M.; Boehm, H.-J. ROCK: the Roche medicinal chemistry knowledge application—design, use and impact. *Drug Discovery Today* **2009**, *16*, 691–696.

(10) *SharePoint*; Microsoft: Redmond, WA.

(11) Kolatkar, A.; Kennedy, K.; Halabuk, D.; Kunken, J.; Marrinucci, D.; Bethel, K.; Guzman, R.; Huckaby, T.; Kuhn, P. C-ME: A 3D Community-Based, Real-Time Collaboration Tool for Scientific Research and Training. *PLoS ONE [Online]* **2008**, *3*, e1621; doi:10.1371/journal.pone.0001621 (accessed October 13, 2011).

(12) Barber, C. G.; Haque, N.; Gardner, B. OnePoint - combining OneNote and SharePoint to facilitate knowledge transfer. *Drug Discovery Today* **2009**, *14*, 845–850.

(13) Hohman, M.; Gregory, K.; Chibale, K.; Smith, P. J.; Ekins, S.; Bunin, B. Novel web-based tools combining chemistry informatics, biology and social networks for drug discovery. *Drug Discovery Today* **2009**, *14*, 261–270.

(14) DEGAS is an acronym for: Discovery by Enumeration, Grouping, and Analysis of Structures.

(15) *Pipeline Pilot*, version 8.0; Accelrys: San Diego, CA, 2011.

(16) Lee, M.; Gobbi, A. *AEREA*; Aestel Scientific Information: San Francisco, CA, 2011. This software is open source. Contact the authors for the program code: aestelSW@gmail.com.

(17) Gobbi. A.; Lee, M. Handling of Tautomerism and Stereochemistry in Compound Registration. Accepted for publication to a appear in this issue. *J. Chem. Inf. Model.*

(18) Molecular refractivity (cMR) is computed with *CMR*, version 4.3; Biobyte Corp: Claremont, CA, 2008.

(19) cLogP, cLogD, and cpKa are computed with *MoKa*, version 1.1; Molecular Discovery: Perugia, Italy, 2009.

(20) H-bond donor/acceptor, heavy atoms, rings, charge, and other atom/bond counts are computed with the *OEChem Toolkit*, version 1.7; Openeye Scientific Software: Santa Fe, NM, 2010.

(21) TPSA is computed with a program implemented at Genentech based onErtl, P.; Rohde, B.; Selzer, P. Fast calculation of molecular polar surface area as a sum of fragment-based contributions and its application to the prediction of drug transport properties. *J. Med. Chem.* **2000**, *43*, 3714–3717.

(22) CNS MPO scores are computed with a program implemented at Genentech that is based on Wager, T. T.; Hou, X.; Verhoest, P. R.; Villalobos, A. Moving beyond Rules: The Development of a Central Nervous System Multiparameter Optimization (CNS MPO) Approach To Enable Alignment of Druglike Properties. *ACS Chem. Neurosci.* **2010**, *1*, 435–449.

(23) The DMPK property predictions are computed using SVM models that are developed at Genentech. The models are developed using Pipeline Pilot's standard components for statistics in *R* and for computing descriptors. A paper is in preparation.

(24) *PyMol*, version 1.4; Schrödinger, LLC, Cambridge, MA, 2011.

(25) *Benchware 3D Explorer*, version 2.6; Tripos International: St. Louis, MO, 2010.

(26) Hewitt, R.; Gobbi, A.; Lee, M. A Searching and Reporting System for Relational Databases Using a Graph-Based Metadata Representation. *J. Chem. Inf. Model.* **2005**, *45*, 863–869.

(27) Lee, M.; Aliagas, I.; Gobbi, A. Scientific Database Application Without Borders: Empowering The Scientists. In *Program and Abstracts of 8th International Conference on Chemical Structures*; Noordwijkerhout: The Netherlands, 2008, 70—71.

(28) The report designer UI is very similar to the query builder UI. A tree on the left allows user to select report columns. The selected columns are listed on the right and can be formatted by clicking on edit buttons.

(29) *JBoss*, version 4.2.3; JBoss Enterprise Middleware: Atlanta, GA, 2008. This application is open source: http://www.jboss.org/ (accessed October 1, 2011).

(30) Juniper Networks Instant Virtual Extranet (IVE) appliance; Juniper Networks: Sunnyvale, CA, USA, 2010.

(31) Koutsonikola, V.; Vakali, A. LDAP: framework, practices, and trends. *IEEE Internet Comput.* **2004**, *8*, 66–72.