JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

DiSCuS: An Open Platform for (Not Only) Virtual Screening Results Management

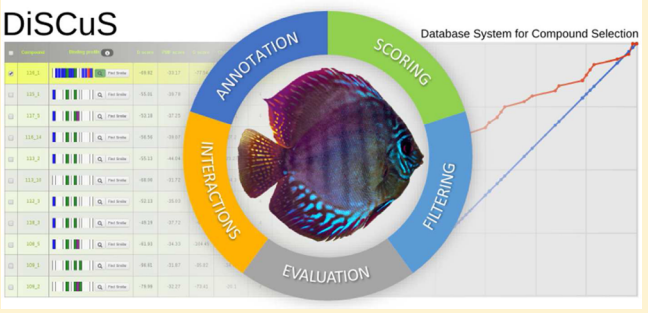
Maciej Wójcikowski,[†] Piotr Zielenkiewicz,[†,‡] and Paweł Siedlecki*[,†,‡,§]

[†]Institute of Biochemistry and Biophysics PAS, Pawińskiego 5a, 02-106 Warsaw, Poland

[‡]Department of Plant Molecular Biology, Institute of Experimental Plant Biology and Biotechnology, University of Warsaw, Miecznikowa 1, 02-096 Warsaw, Poland

Supporting Information

ABSTRACT: DiSCuS, a "Database System for Compound Selection", has been developed. The primary goal of DiSCuS is to aid researchers in the steps subsequent to generating high-throughput virtual screening (HTVS) results, such as selection of compounds for further study, purchase, or synthesis. To do so, DiSCuS provides (1) a storage facility for ligand−receptor complexes (generated with external programs), (2) a number of tools for validating these complexes, such as scoring functions, potential energy contributions, and med-chem features with ligand similarity estimates, and (3) powerful searching and filtering options with logical operators. DiSCuS supports multiple receptor targets for a single ligand, so it can be used either to evaluate different variants of an active site or for selectivity studies. DiSCuS documentation, installation instructions, and source code can be found at http://discus.ibb.waw.pl.



INTRODUCTION

Computational methods are currently present in almost every step of the drug-discovery process; one point where computational methods have had significant impact is high-throughput virtual screening (HTVS). Over the years, many applications have been developed to allow in silico screening of large data sets of compounds to known receptor structures that are both open-source/free (e.g., UCSF Dock,[1,2] Autodock,[3,4] Autodock Vina,[5] PLANTS[6]) and proprietary (Tripos Surflex-Dock,[7] Schrödinger Glide,[8−10] Gold[11−13]). Although these programs use different algorithms [Lamarckian genetic algorithm (LGA), anchor-and-grow, and ant systems, to name the most prominent ones], there is no clear advantage of using one over the other.[14,15] It is also rarely the case that the single best conformation generated by these programs is the true positive. Studies have shown that it is possible to obtain a reasonable success rate (i.e., for approximately 70% of cases), if at least 10 of the best conformations of a single ligand are taken into account.[16]

The amount of data that a HTVS campaign generates can be overwhelming. Using in silico docking methodology, one must generate many diverse conformations at the receptor active site in the hope that further validation of these complexes will yield the proper binding mode. Ligand conformation sampling is still an outstanding challenge in docking procedures,[17] and taking into account factors such as the ligands' possible protonation states (tautomers) or starting conformations, only further increases the amount of data to be analyzed. Receptor rigidity during docking is another problem that strongly influences the success of a HTVS campaign.[18,19] Methods to overcome this hindrance have emerged, generally via the usage of multiple receptor structures (so-called ensemble docking[20]) and/or consensus docking[21,22] (docking with multiple algorithms simultaneously). Such approaches have been shown to increase cross-docking accuracy, providing better tools for the initial screening and selection step and the lead-optimization process. Obviously, if one chooses to use these tools, the amount of structural data will increase accordingly.

It should be evident that a properly designed HTVS experiment will produce a vast amount of data composed of 3D ligand−receptor complexes, their accompanying "scores", and their relationships with each other. Projects that aid the user in the automation of HTVS campaigns have been developed, covering the work flow from preparing ligand databases to docking and scoring. These include commercial platforms such as Tripos Muse,[23] Schrödinger Maestro,[24] or MOE.[25] There are also available plugins, such as Knime[26] and PipelinePilot,[27] for various tasks for work flow pilots. Although these plugins are useful in procedure automation, they still leave the user with a huge library of ligand conformations that cannot be easily discarded before evaluation. To make matters worse, evaluation procedures typically involve generating more data. Various scoring functions, such as the popular ChemScore,[28] PMF,[29] or XScore,[30] provide fast but crude putative affinity estimates. Again, the usage of multiple approaches (consensus

Received: October 10, 2013

dx.doi.org/10.1021/ci400587f | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

scoring[31]) tends to be preferred by HTVS protocols. Highly useful information can be obtained by calculating the binding free energy from molecular mechanics force fields.[32] Tools that estimate ligand−receptor interactions, such as hydrogen bonds, pi-stacking rings, salt bridges, and hydrophobic contacts among others, are highly useful for the complex assessment process. Typically, it is better to collect more information about docking results than less. Overall, choosing the best compounds from a docking/scoring experiment is the most labor-intensive and error-prone task in the whole HTVS procedure.

There is a gap between the almost automatic generation of computational results and the need to manually inspect, explore, and filter the generated results. Systems that try to bridge this gap have already started to emerge, allowing easy storage and access to data. Prominent examples include VSDMIP (Virtual Screening Data Management on an Integrated Platform),[33,34] which is a Python plugin for PyMol[35] tailored toward visual inspection. It provides a great way to compare the different binding modes of a single ligand but at the same time limited ability to create user-defined subsets, carrying, for example, specific scaffolds or statistical procedures, such as receiver operating characteristic (ROC) curves or scoring function training. IVSPlat,[36] developed by Yin Xue Sun et al., is another example of a PyMol[35] plugin. Although highly useful, IVSPlat is primarily a work flow-oriented platform (i.e., receptor and ligand preparation, docking, scoring) with limited searching/filtering features. There are also various commercial database applications, such as ChemAxon JChem Base,[37] Tripos Unity,[38] and Schrödinger Canvas,[39,40] which provide good feature-to-cost ratios. Despite being fully featured products, they also have limitations, for example, there is no easy way to gather information about the binding modes of a single ligand in different receptor structures (needed for ensemble docking). These programs are also commonly paired with a specific docking program, which makes consensus docking a difficult task. They are also closed source, which hinders their development by the academic community.

All the facts mentioned above have inspired us to develop DiSCuS (Database System for Compound Selection). DiSCuS has been designed with a single purpose in mind: to provide enough information about ligand−receptor complexes (or as much as the user chooses to generate) to make the most crucial step in computational drug discovery—selecting ligands to purchase or synthesize—much easier. DiSCuS provides advanced sorting, filtering, and subset creation features as well as evaluation tools. By using a relational database, it provides the means to store and quickly access HTVS results. Additionally, DiSCuS is bundled with tools and procedures allowing various scores, interactions, and features to be defined and computed post-docking.

DiSCuS has also been designed to accompany researchers through the most of the discovery campaign by allowing the integration of assay results with the physiochemical and computational data. The system can store values (e.g., IC50) and raw data such as gel scans etc. acting as a share-point for various groups interacting within a particular project. These features make DiSCuS a data-sharing and evaluation system and are unique compared to other software developed to date.

MATERIALS AND METHODS

User Interface. The DiSCuS user interface was designed to be lightweight and multi-platform. To accomplish this goal, interaction with the system is conducted through a Web interface. Another goal was to have a minimal time delay to deliver information to the researcher despite the vast amount of data that could be stored. Many of the computations necessary for post-docking evaluation and analysis are performed during the process of importing raw HTVS results. Where it was possible, we also implemented asynchronous procedures. This approach allows the possibility to conduct procedures, compute features, or score conformations on separate machines or parallelizing such tasks. It is especially useful for minimization and other computationally expensive tasks. Because the calculation jobs are carried out in the background, the user is not haunted by the never-ending spinning hourglass and can work with the existing data while waiting for the new results to arrive.

Data Importing and Exporting. The primary method of importing docking results into DiSCuS is by uploading files through the Web interface or a Python script. Currently, DiSCuS can import docked conformations and protein models in most popular formats, such as mol2, pdb, pdbqt, and sdf. We use stock OpenBabel[41] libraries for this task, so adding support for other files is relatively straightforward. The import process can also be coupled with the researchers' own docking procedure, feeding data directly into DiSCuS's backend database, as soon as results appear out of a HTVS pipeline. This is the primary way that DiSCuS works in our computational cluster environment. Internally, ligand conformations and macro molecules (receptors) are stored as mol2 files. Data are exported by selecting subsets or individual molecules through the Web interface. User selections (subsets) can be downloaded as flat files in mol2 format or as UCSF Chimera[42] sessions if the user wishes to perform further external analysis. This latter method takes advantage of the built-in capabilities of Chimera's ViewDock tool. It also leverages its usefulness by including all the "scoring" information and interaction features calculated by DiSCuS.

Post-Docking Analysis Tools. After importing results from a HTVS experiment, DiSCuS will provide the user with a range of tools and gathered information useful for further examination of the results. As mentioned previously, post-docking analysis is one of the primary goals of DiSCuS. The system is bundled with a range of tools and methods suitable for this task, but this list is nowhere near complete. To satisfy the "power users", we provided an easy to use plugin environment, which allows the incorporation of external tools and popular proprietary software (e.g., Tripos SybylX modules) to further extend DiSCuS functionality. The DiSCuS plugin system allows the addition of numerous custom procedures or additional analysis such as scoring functions, minimization scripts, etc. The database structure required by custom plugins is created automatically, and data are available to the user as soon as post-processing is completed. Such data becomes integral to DiSCuS and can be analyzed just like data from already built-in plugins.

One of the key steps in post-docking analysis is energy minimization of ligand−receptor complexes using force fields. Docked complexes are generated by fast and rather crude methods and are not optimized for binding interactions. Utilizing short minimization procedures can be beneficial in overcoming problems and artifacts inherited from docking runs. DiSCuS is bundled with an energy minimization procedure (OpenBabel) and various force fields tailored toward small molecules, including MMFF94[43] (as default), MMFF94s,[44]

dx.doi.org/10.1021/ci400587f | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

UFF,[45] and GAFF.[46] DiSCuS will automatically provide pre- and post-minimization free energy contributors, such as electrostatic terms, van der Waal's terms, and combined terms (total terms). After the calculations have been performed, all ligand conformations will be stored in DiSCuS as energy-minimized conformations.

Docking programs may provide their own score terms for a generated protein−ligand complex. DiSCuS supports importing such estimates through the Web interface. Along with the above, an empirical scoring function, X-Score,[30] is bundled and implemented. On a test set of 30 selected PDB complexes, X-Score was able to predict binding free energies with a standard deviation of 2.2 kcal/mol.[30] Furthermore, DiSCuS is bundled with a plugin that uses the SybylX[37] environment to compute the scoring functions, empirical D_Score[1] and G_Score[12] and knowledge-based ChemScore[28] and PMF Score,[29] which are all combined in a Consensus Score (CScore). To use this feature, the user is asked to have SybylX already installed.

Docking scores, scoring functions, and free energy values are an important part of DiSCuS functionality. The RankScore tool is designed mainly as an interactive method for (1) assessing the performance of different scores to distinguish between active and inactive/decoy sets and (2) developing custom scoring procedures based on such evaluations. Such a scoring consensus can then be used for ranking novel ligands or to estimate thresholds below which ligands will be discarded from further analysis. Additionally, one can introduce interaction features such pi-stacking or metal/ion coordination to develop a new consensus. A screenshot presenting a consensus scoring (RankScore) module is available in the Supporting Information.

To allow users to gain a more detailed insight as to what is happening in their protein−ligand complex, DiSCuS is bundled with the Interaction plugin. The main purpose of the Interaction plugin is to compute the amino acid neighborhood of a ligand conformation and gather information about the possible interactions in a given complex. From this information, the tool builds a binding profile that quantitatively describes the surroundings of an individual ligand and what types of interactions can be established. Similar to the SIFt algorithm[47] described by Deng et al., such a profile is composed of electrostatic terms (donors and acceptors), a hydrophobic component (defined as aromatic ring systems and carbons not attached to hetero atoms), pi-stacking, pi-cation, and metal/ion coordination. The main difference in our approach is that we do not measure exact angles and distances to define these interactions. Instead, in our "crude" implementation, it is sufficient for interaction to be detected if an atom from a feature matching an amino acid is present in a sphere centered on the ligand atom (e.g., 3,5 Å for H-bond). Our main motivation to develop this alternate methodology was that ligand conformation sampling in docking is not exhaustive, so their placement in the active site is not ideal. Additionally, by allowing more flexible interaction terms, our procedure attempts to mimic putative active site adjustments, trying at least in some way to overcome one of the main issues of HTVS campaigns—the rigidity of receptors. "Precise" interaction terms (with distances and angles) are also supported and can be selected, for example, when the researcher is confident about the ligand−receptor complex quality. A screenshot presenting the Interactions module is available in the Supporting Information.

Binding profiles generated by the Interaction plugin are not exclusively useful for visualizing interactions. They can be highly efficient tools for similarity searches. If a user finds a ligand with a desired or interesting mode of binding, he can immediately query the system for other compounds that have similar interactions with the active site. Moreover, using the Web interface, one can define which and what types of interactions are "interesting" in a given experiment and search for ligand conformations matching such a binding profile. Finally, for each receptor, DiSCuS also provides statistics about the type and frequency of interactions between ligands and the active site amino acids. These collections can be used in various ways; in our department, we mainly utilize such data for finding compounds that bind in a unique way to the receptor or to remove ligand conformations that do not match a common binding pattern.

Plugin API. A Simple Plugin API was designed to enable external tools to easily cooperate with DiSCuS. Because of the multitude of types of computed data, output arguments must be precisely defined for DiSCuS to process them correctly. The custom plugin extends the *plugin_interface* PHP class. It consists of the following elements: (1) static $name; static $desc: name and description of a plugin. (2) public $batch_size: number of conformations to be processed simultaneously. (3) static $input: an array defining the formats for the ligands and receptor (separately). If some custom options need to be applied before format conversion, OBMol can be used in the plugin. (4) static $results: an array defining the output name, label, and type of each computed argument. Molecular and conformational features are separated from each other. (5) compute($ligands, $receptor): the public method that is the core of the plugin. This executes all routines. (5a) input: an array of ligands and the receptor string (both in formats defined in $input). (5b) output: an array of processed ligands and computed values. For each ligand name (array key), a full set of computed arguments is assigned and stored in nested arrays. The computed fields to be processed must be described with $results.

The sample plugin computing Xscore is available in the Supporting Information.

User Management. DiSCuS supports multiple users and access permissions. When we started building the system, we had different VS projects ongoing with collaborators who wanted to ensure the confidentiality of the stored data. For a system that can store sensitive data, it is necessary to have flexible access control mechanisms and user management. DiSCuS was developed to work in a multi-user environment, and it uses built-in MySQL and PHP capabilities to restrict user permissions to specific data containers and provides a simple Web interface for account management (Admin tools > Manage Users, screenshot available in the Supporting Information). Users can also create custom data subsets and assign permissions, sharing them only with chosen users. Moreover, DiSCuS utilizes permanent http links, which eliminates the requirement to save sessions. This feature also means that users with the same permissions can share their findings through direct http links, which we find very convenient and use even more frequently than subset sharing. For instance, instead of giving exact parameters for filtering, the user can just copy and paste the link to the Web site with the results and share it with others.

Features Highlight. Here, we will highlight some of the more unique and/or interesting features of the presented software. DiSCuS can be used for analyzing both simple docking experiments with a single target and for large screening

campaigns against many targets, although many of its virtues can be observed when analyzing a large amount of data. The ability to easily manipulate a large data set is to our knowledge by itself quite unique compared to its commercial and free peers. A simple task of filtering a docked library based on a single feature can become impossible to handle for those bundles when the ligand library is scaled up; for some, 100 000 is the performance limit. However, such a library is not a problem for any of the public databases of small molecules, for example, ZINC consists of almost 20 million compounds that can be filtered quite rapidly. DiSCuS is designed to handle "big data" through a relational database backend with a specifically optimized table structure, tailored for storing and manipulating ligand−receptor conformations. As a ligand-centric database, it is also perfect for experiments with many targets, such as ensemble docking analysis or broad specificity/selectivity analysis. Such features are not supported in the majority of currently available software.

DiSCuS was built as a modular system, with the idea of integrating various external tools that the authors use, both commercial and free. We strived to develop a system that would be able to send data to external tools and then integrate their output back into the database structure. In this way, DiSCuS would be enriched by many sophisticated tools already in use or currently being developed, rather than being surpassed by them. It is important to think of DiSCuS not as a replacement for, but rather as completion of other tools, or as an information hub, allowing users to cherry-pick the strongest features from various software programs and integrate them into a decision platform.

In addition to having the capability to import external data, DiSCuS provides a range of tools on its own and can be used without any external contributors. Of this range of tools, we would like to first point the reader to the Binding Profile module, which is a unique feature optimized for the computation of ligand−receptor interactions in a large-scale docking result. When activated, DiSCuS computes a wide range of physical interactions (see Materials and Methods section) and stores them as 1D strings. These can be compared among ligands (or among multiple conformations of a single ligand) to identify similar binders for filtering purposes. They can also be used to enrich ligand libraries (a quasi 3D-SAR method) based on rationally generated rather than randomly generated 3D coordinates. Binding Profiles provide swift insight on the interactions within a ligand−receptor complex and can be calculated for all the user data. This approach is useful for a wide range of applications and not only as yet another discriminating feature.

Another feature worth highlighting is the DiSCuS implementation of consensus scoring (CS) methodology. In DiSCuS, any score and ligand feature present in the database can be used in consensus scoring. This is a major improvement in comparison to CScore (Tripos SybylX), which generally has four scoring functions with fixed ranges treated as "positive". Additionally, the RankScore module, which is DiSCuS's implementation of a fully flexible consensus score, normalizes each score/feature to its relative position in the set of compounds. When bioactivity data are available (or one defines active and inactive subsets), a ROC chart can be drawn with computed AUC values. These can be instantly used within DiSCuS to measure each function's performance and allow the user to rationally customize the tailored scoring procedure by including/excluding certain features and/or applying weights (see Results and Discussion section).

A not-so-visible but highly useful feature implemented in DiSCuS is the Automatic Annotation module, which to our knowledge is not available in other similar software. Automatic Annotation links each substance that is imported into DiSCuS to publicly available databases (such PubChem, ChEMBL, and ZINC) and vendors utilizing Unichem.[48] This feature provides a fast route to pushing docking results downstream of externally stored bioassay and biochemical data. Furthermore, DiSCuS will automatically link its compounds to new bioassay results imported by researchers. In this case, one must provide a CSV file with compound names and activities, and the import procedure will update and link appropriate database tables. Additional files can be attached to each assay, for example, PDF files, images of gel scans, etc.

One of the biggest disadvantages of using DiSCuS compared to peer software is the much richer customization of 3D visualization of protein−ligand complexes in the latter. This is mainly due to the use of a Web interface for user interaction with DiSCuS. Although limited, we believe that the presentation of such complexes in JSmol,[49] combined with our highlighting mechanism showing interacting receptor residues while browsing through ligand conformations, is sufficient in most cases to make sound decisions. Nevertheless, to gain better visual insight into the nature of certain ligand−receptor complexes, an external visualization environment might be necessary. This is why we provide a UCSF Chimera plugin for downloading user-selected data directly into this molecular environment in the form of a "session". Plugins for other software, for example, PyMOL, would be something that we imagine the growing community of users would be eager to create and implement.

Last but not least, a notable feature of DiSCuS is its centralized topology. When installed on a server, there are virtually no requirements on the client side—a local PC with only a modern browser is sufficient. In contrast, Tripos SybylX or Schrödinger Maestro ask the user to install the whole software locally, and usually, all computation is also run locally or requires additional licensing costs to use other hardware. DiSCuS can be installed based on the researcher's needs, either locally or deep in air-conditioned IT cellars, and can then be accessed from any place with a network connection. In both cases, DiSCuS can provide analysis results instantaneously in most scenarios. For more information about the Web interface, installation, user documentation, and source, see the DiSCuS Web page: http://discus.ibb.waw.pl.

RESULTS AND DISCUSSION

There are many methods for improving docking performance, and many of them may be executed within DiSCuS. In this section, some case studies illustrating the application of DiSCuS in post-docking analysis workflows are presented (Figure 1). These cases consist of consensus scoring, ensemble docking, and key interactions analysis for the AmpC β-lactamase protein, PDB code: 1L2S.

Consensus Scoring and Evaluation of Scoring Function (RankScore). A full library of 48 active and six inactive compounds derived from the DUD-E data set was docked into the AMPC active site using UCSF Dock version 6.5. Docking results were uploaded into DiSCuS and underwent post-docking scoring procedures using the RankScore module. In this case, the DiSCuS workflow consisted of

dx.doi.org/10.1021/ci400587f | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

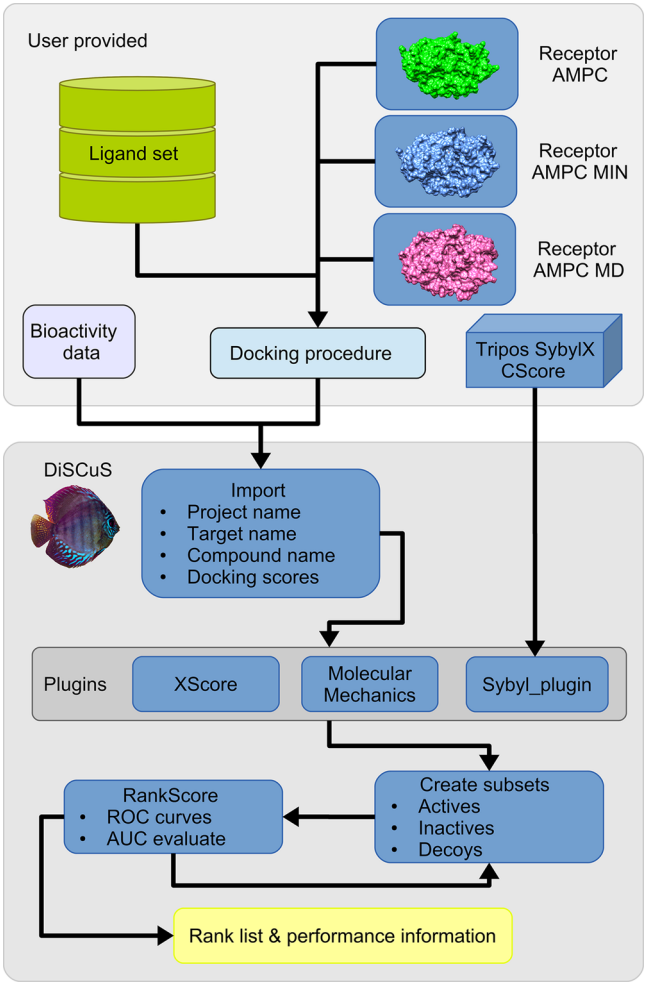


Figure 1. DiSCuS workflow for developing a new scoring scheme using RankScore and AUC values.

the XScore scoring function as well as the Sybyl_CScore plugin, which allows direct usage of the scoring functions available in Tripos Sybyl from within DiSCuS (both plugins bundled with DiSCuS). Activity data were derived from DUD-E, saved in CSV format, and merged with in silico screening results (through DiSCuS's Import plugin). Finally, two user subsets (*AMPC_active* and *AMPC_inactive*) were created based on the activity data, ending the preparation step. Subsequently, the RankScore module was used to analyze the generated docking conformations. The user subsets were applied to compute the area under curve (AUC) of the receiver operating characteristic (ROC) curve for each scoring function. These data suggested that only two scoring functions give reasonable performance, D_Score of 0.7059 and ChemScore of 0.7106. When both are taken into account in RankScore (with neutral weights of 1), the resulting AUC was 0.7746. Next, energies computed by ligand minimization in the MMFF94 force field were taken into account in the new scoring equation. Minimization procedures were also conducted within DiSCuS. The RankScore combining those three factors (D_Score, ChemScore, and MMFF94 minimization energy) resulted in an AUC of 0.8089.

To evaluate the new scoring scheme produced by RankScore, we imported the docking results of 2832 decoy compounds from the AMPC DUD-E set into DiSCuS. Our goal was to use DiSCuS to assess whether the new Scoring scheme developed in RankScore would be beneficial in discriminating decoys when mixed with active compounds. Again, we computed ROC curves for the *active subset* and *decoy subset*. The AUC for D_Score was 0.6938, ChemScore was 0.6938, and MMFF94 was 0.7872. A combined (RankScore) scoring produced an AUC of 0.8127 (Table 1). This may be a good starting point to start playing with weights or introducing other parameters such as hydrogen bonds, which is one of the advanced features of RankScore.

Table 1. ROC AUC Values and Charts Showing the Evaluation Process for Two Ligand Sets (Active vs Inactive and Active vs Decoys) Docked into the AMPC β-Lactamase Structure

Target/Score	D_Score	ChemScore	D_Score & ChemScore	MMFF94	RankScore
Active vs. Inactive	0.6627	0.6683	0.7468	0.7747	0.7821
Active vs. Decoy	0.6938	0.6616	0.7615	0.7872	0.8127

Ensemble Docking. Further improvement of the RankScore discriminative performance can be achieved thanks to DiSCuS's capability to assess data from many targets at the same time with ensemble docking. In this study, two additional instances of AMPC structure were prepared: one minimized with the MMFF94 force field (1000 steps) and another after a short 1000-step molecular dynamics run combined with another 1000 steps of MMFF94 minimization. For each of these models, steps analogous to those described in the consensus scoring were repeated. Detailed results representing the docking performance as the area under the ROC curve (AUC) are shown in Table 2. Overall, this example shows how

Table 2. AUC Values (derived from ROC curves) for Three Structures of β-Lactamase, AMPC PDB (standard structure taken from PDB), AMPC MM (subjected to a short minimization procedure), and AMPC MD (subjected to short molecular dynamics and minimization procedure) (details in text)[a]

Target\Score	D_Score	ChemScore	MMFF94	RankScore
AMPC PDB	0.6162	0.6926	0.7561	0.7740
AMPC MM	0.6936	0.7526	0.8420	0.8754
AMPC MD	0.6506	0.7414	0.7991	0.8659
Ensemble	0.7091	0.7783	0.8228	0.8867

[a]Two scoring functions are evaluated along with the total energy from the MMFF94 force field and the RankScore scoring scheme, which combines all three terms.

dx.doi.org/10.1021/ci400587f | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

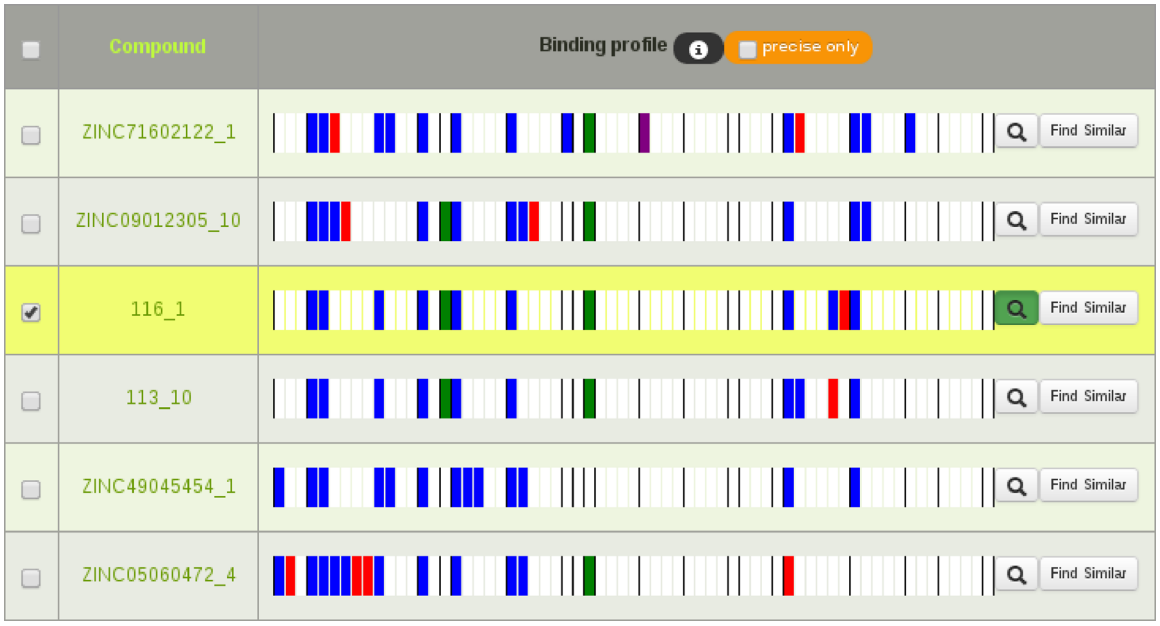


Figure 2. Binding profile from the Interaction module. The table consists of aligned binding profiles for ligand conformations with color-coded interactions (blue, H-bond donor; red, H-bond acceptor; green, hydrophobic contact; purple, pi-stacking). The highlighted binding profile is also visualized in the receptor 3D structure (not shown).

multiple receptor structures are natively supported by the system and how fast those results may lead to improved scoring schemes. Only using DiSCuS one can execute such complex analysis in few simple steps (clicks) with instant feedback about performance.

These two cases show a typical post-processing workflow for screening procedures, namely, the development of improved scoring methodologies based on docking results. As shown above, with DiSCuS's RankScore module and its ability to merge docking results from many receptors (ensemble docking), it becomes a relatively easy task to create better custom-tailored scoring schemes, introduce new terms, and quickly test various parameters or weights.

Key Interaction Analysis. Analysis of receptor−ligand interactions provides key insight into how ligands bind in the active site of receptors. Such interactions are computed automatically within DiSCuS's Interaction module. Docking results may be evaluated visually in addition to applying various scoring schemes. DiSCuS provides this type of 3D visualization functionality and extends it by providing a 1D pharmacophore-like binding profile, which is helpful when hand-picking ligands or conformations. In this case study, binding profiles have been used in a different way, namely, as a filtering tool for subset enrichment. We wanted to see whether searching for binding patterns similar to those of known ligands can enrich the overall set with active compounds. The workflow for this procedure is presented in Figure 2. Here, we used Autodock Vina to dock 48 active ligands and 2832 decoys to AMPC (ligand set directly taken from DUD-E), and the resulting complexes were imported into DiSCuS. Next, the best conformation scored by Autodock Vina for the most active ligand (named 116) was chosen, and its binding profile was analyzed (using the Molecule Summary page). It consisted of the following nine interactions: seven H-bonds and two hydrophobic contacts. It was also very close (RMSD < 1 Å considering the common scaffold) to the conformation found in the crystal structure 1L2S. This binding profile was not edited and was used "as is" to search for compounds that bind in a similar fashion (using the Interactions module). After that, the resulting subset consisted of 267 binding modes (approximately 1% of all imported conformations). Conformations were sorted by the similarity of binding profiles. In a regular HTVS campaign, this set would be the top compounds selected for evaluation, so we wanted to see what the enrichment factor (EF_1) would be for that set. While random enrichment was ∼2%, the EF for this set was 12%. It is worth noting that for AMPC, DUDE reports an EF of 8.3% (based on their own docking and scoring scheme). This case study showcases a simple yet effective and semi-automatic way to filter compounds based not on their chemical proprieties or scores but on specific interactions with the receptor. Using the Interactions module, these Binding Profiles can be selected based on an interesting compound or hand selected by the researcher from the list of residues. The resulting binding profiles can also be used to align a set of compounds in a somewhat similar interaction pattern to multiple sequence alignments.

■ CONCLUSIONS

The case studies presented above were executed within DiSCuS based exclusively on docking results and information freely available in public databases. Those considering both consensus scoring and ensemble docking lead to the conclusion that more data yields better results. DiSCuS enables researchers to store and manage large data sets and to analyze them in a simple subset-driven method regardless of the origin or type. This provides a lot of flexibility and the opportunity to merge different data sources and various algorithms in subsequent steps of virtual screening. The novel interaction analysis based on 1D binding profiles has also proven to be highly useful not only when one needs to visualize docking results but also as an important step in creating enriched libraries or even as an additional scoring procedure.

DiSCuS was designed to be extensible, so new features can be added depending on user needs. Because of its centralized

dx.doi.org/10.1021/ci400587f | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

architecture and relational database backend, DiSCuS can also be used as a storage facility for large-scale VS performed on HPC clusters. Supporting computational data with bioassay results is highly beneficial. DiSCuS takes advantage of automatic annotation features to look for corresponding information in public databases, such as PubChem or ChEMBL. Supporting data for bioassays also can be provided by the user in the form of raw values, gel scans, or PDF files with commentary. The ability to cross-link data from many docking experiments executed on various targets utilizing ensemble docking can be useful when working with selectivity/specificity issues.

Documentation, installation requirements, and related content may be found on the DiSCuS Web site: http://discus.ibb.waw.pl.

Finally, the list of features provided with DiSCuS is far from complete. We are sharing the code with the community through GitHub (see official Web site for details) in the hope that a free and open system will be useful and extended by external users. We paid a great deal of attention to utilizing free/open-source building blocks where it was possible and built a user-friendly plugin system, so extending DiSCuS functionality should be as pleasant an experience as possible.

■ ASSOCIATED CONTENT

Supporting Information

Screenshots presenting DiSCuS's interface (Interactions module, RankScore module, and User Management) and DiSCuS sample plugin computing Xscore. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*Tel.: +48 22 592 57 61. Fax: +48 22 592 21 90. E-mail: pawel@ibb.waw.pl.

Present Address

[§]P. Siedlecki: Institute of Biochemistry and Biophysics, Polish Academy of Sciences, Pawinskiego 5a 02-106, Warsaw, Poland.

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work has been supported by the Polish Ministry of Science and Higher Education (Grant No. IP2010 037470) and The National Centre for Research and Development (Grant No. PBS1/A7/9/2012).

■ ABBREVIATIONS

AMPC AmpC β-lactamase; AUC area under curve; EF enrichment factor; HTVS, VS high-throughput virtual screening; MD molecular dynamics; MM molecular mechanics; ROC receiver operating characteristic

■ REFERENCES

(1) Kuntz, I. D.; Blaney, J. M.; Oatley, S. J.; Langridge, R.; Ferrin, T. E. *J. Mol. Biol.* **1982**, *161*, 269−288.

(2) Lang, P. T.; Brozell, S. R.; Mukherjee, S.; Pettersen, E. F.; Meng, E. C.; Thomas, V.; Rizzo, R. C.; Case, D. A.; James, T. L.; Kuntz, I. D. *RNA* **2009**, *15*, 1219−1230.

(3) Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J. *J. Comput. Chem.* **1998**, *19*, 1639−1662.

(4) Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.; Olson, A. J. *J. Comput. Chem.* **2009**, *30*, 2785−2791.

(5) Trott, O.; Olson, A. J. *J. Comput. Chem.* **2010**, *31*, 455−461.

(6) *Ant Colony Optimization and Swarm Intelligence*; Dorigo, M., Gambardella, L. M., Birattari, M., Martinoli, A., Poli, R., Stützle, T., Eds.; Springer Berlin Heidelberg: Berlin, Heidelberg, 2006; Vol. *4150*.

(7) Tripos International. Surflex-Dock. http://www.tripos.com/.

(8) Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S. *J. Med. Chem.* **2004**, *47*, 1739−49.

(9) Halgren, T. A.; Murphy, R. B.; Friesner, R. A.; Beard, H. S.; Frye, L. L.; Pollard, W. T.; Banks, J. L. *J. Med. Chem.* **2004**, *47*, 1750−1759.

(10) Friesner, R. A.; Murphy, R. B.; Repasky, M. P.; Frye, L. L.; Greenwood, J. R.; Halgren, T. A.; Sanschagrin, P. C.; Mainz, D. T. *J. Med. Chem.* **2006**, *49*, 6177−6196.

(11) Verdonk, M. L.; Cole, J. C.; Hartshorn, M. J.; Murray, C. W.; Taylor, R. D. *Proteins* **2003**, *52*, 609−623.

(12) Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. *J. Mol. Biol.* **1997**, *267*, 727−748.

(13) Jones, G.; Willett, P.; Glen, R. C. *J. Mol. Biol.* **1995**, *245*, 43−53.

(14) Cross, J. B.; Thompson, D. C.; Rai, B. K.; Baber, J. C.; Fan, K. Y.; Hu, Y.; Humblet, C. *J. Chem. Inf. Model.* **2009**, *49*, 1455−1474.

(15) Cummings, M. D.; DesJarlais, R. L.; Gibbs, A. C.; Mohan, V.; Jaeger, E. P. *J. Med. Chem.* **2005**, *48*, 962−976.

(16) Plewczynski, D.; Łaźniewski, M.; Augustyniak, R.; Ginalski, K. *J. Comput. Chem.* **2011**, *32*, 742−755.

(17) Scior, T.; Bender, A.; Tresadern, G.; Medina-Franco, J. L.; Martínez-Mayorga, K.; Langer, T.; Cuanalo-Contreras, K.; Agrafiotis, D. K. *J. Chem. Inf. Model.* **2012**, *52*, 867−881.

(18) Anderson, A. C.; O'Neil, R. H.; Surti, T. S.; Stroud, R. M. *Chem. Biol.* **2001**, *8*, 445−57.

(19) Siedlecki, P.; Garcia Boy, R.; Musch, T.; Brueckner, B.; Suhai, S.; Lyko, F.; Zielenkiewicz, P. *J. Med. Chem.* **2006**, *49*, 678−83.

(20) Huang, S.-Y.; Zou, X. *Proteins* **2007**, *66*, 399−421.

(21) Plewczynski, D.; Łaźniewski, M.; von Grotthuss, M.; Rychlewski, L.; Ginalski, K. *J. Comput. Chem.* **2011**, *32*, 568−581.

(22) Houston, D. R.; Walkinshaw, M. D. *J. Chem. Inf. Model.* **2013**, *53*, 384−390.

(23) Tripos International. Tripos Muse. http://www.tripos.com/.

(24) Schrödinger LLC. Maestro Suite. https://www.schrodinger.com/.

(25) Chemical Computing Group Inc. Molecular Operating Environment (MOE) http://www.chemcomp.com.

(26) Berthold, M. R.; Cebron, N.; Dill, F.; Fatta, G. Di; Gabriel, T. R.; Georg, F.; Meinl, T.; Ohl, P.; Sieb, C.; Wiswedel, B. *Computer (Long Beach, Calif).* **2008**, *11*, 26−31.

(27) Kappler, M. A. *Curr. Opin. Drug Discov. Devel.* **2008**, *11*, 389−92.

(28) Eldridge, M. D.; Murray, C. W.; Auton, T. R.; Paolini, G. V; Mee, R. P. *J. Comput.-Aided. Mol. Des.* **1997**, *11*, 425−445.

(29) Muegge, I. *J. Med. Chem.* **2006**, *49*, 5895−5902.

(30) Wang, R.; Lai, L.; Wang, S. *J. Comput.-Aided. Mol. Des.* **2002**, *16*, 11−26.

(31) Oda, A.; Tsuchida, K.; Takakura, T.; Yamaotsu, N.; Hirono, S. *J. Chem. Inf. Model.* **2006**, *46*, 380−391.

(32) Aqvist, J.; Luzhkov, V. B.; Brandsdal, B. O. *Acc. Chem. Res.* **2002**, *35*, 358−365.

(33) Gil-Redondo, R.; Estrada, J.; Morreale, A.; Herranz, F.; Sancho, J.; Ortiz, Á. R. *J. Comput.-Aided. Mol. Des.* **2008**, *23*, 171−184.

(34) Cabrera, Á. C.; Gil-Redondo, R.; Perona, A.; Gago, F.; Morreale, A. *J. Comput.-Aided. Mol. Des.* **2011**, *25*, 813−824.

(35) *The PyMOL Molecular Graphics System*, 2010.

(36) Sun, Y. X.; Huang, Y. X.; Li, F. L.; Wang, H. Y.; Fan, C.; Bao, Y. L.; Sun, L. G.; Ma, Z. Q.; Kong, J.; Li, Y. X. *Chem. Cent. J.* **2012**, *6*, 2.
(37) ChemAxon. JChem Base. http://www.chemaxon.com (accessed December 11, 2013).
(38) Tripos International. Tripos Unity. http://www.tripos.com/.
(39) Sastry, M.; Lowrie, J. F.; Dixon, S. L.; Sherman, W. *J. Chem. Inf. Model.* **2010**, *50*, 771−784.
(40) Duan, J.; Dixon, S. L.; Lowrie, J. F.; Sherman, W. *J. Mol. Graph. Model.* **2010**, *29*, 157−170.
(41) O'Boyle, N. M.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.; Hutchison, G. R. *J. Cheminform.* **2011**, *3*, 33.
(42) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. *J. Comput. Chem.* **2004**, *25*, 1605−1612.
(43) Halgren, T. A. *J. Comput. Chem.* **1996**, *17*, 490−519.
(44) Halgren, T. A. *J. Comput. Chem.* **1999**, *20*, 720−729.
(45) Rappe, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard, W. A.; Skiff, W. M. *J. Am. Chem. Soc.* **1992**, *114*, 10024−10035.
(46) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *J. Comput. Chem.* **2004**, *25*, 1157−1174.
(47) Deng, Z.; Chuaqui, C.; Singh, J. *J. Med. Chem.* **2004**, *47*, 337−344.
(48) Chambers, J.; Davies, M.; Gaulton, A.; Hersey, A.; Velankar, S.; Petryszak, R.; Hastings, J.; Bellis, L.; McGlinchey, S.; Overington, J. P. *J. Cheminform.* **2013**, *5*, 3.
(49) JSmol. http://jsmol.sourceforge.net/.