

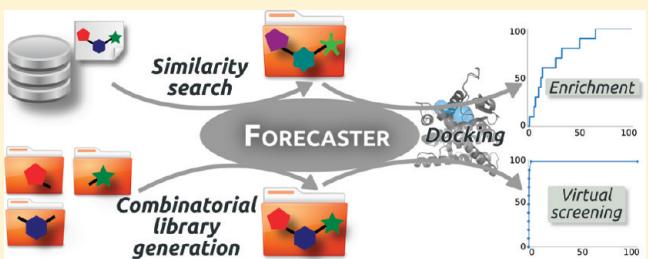
# Integrating Medicinal Chemistry, Organic/Combinatorial Chemistry, and Computational Chemistry for the Discovery of Selective Estrogen Receptor Modulators with FORECASTER, a Novel Platform for Drug Discovery

Eric Therrien, Pablo Englebienne, Andrew G. Arrowsmith, Rodrigo Mendoza-Sanchez, Christopher R. Corbeil, Nathanael Weill, Valérie Campagna-Slater, and Nicolas Moitessier\*

Department of Chemistry, McGill University, 801 Sherbrooke St W, Montreal, QC, Canada H3A 2K6

 Supporting Information

**ABSTRACT:** As part of a large medicinal chemistry program, we wish to develop novel selective estrogen receptor modulators (SERMs) as potential breast cancer treatments using a combination of experimental and computational approaches. However, one of the remaining difficulties nowadays is to fully integrate computational (i.e., virtual, theoretical) and medicinal (i.e., experimental, intuitive) chemistry to take advantage of the full potential of both. For this purpose, we have developed a Web-based platform, FORECASTER, and a number of programs (e.g., PREPARE, REACT, SELECT) with the aim of combining computational chemistry and medicinal chemistry expertise to facilitate drug discovery and development and more specifically to integrate synthesis into computer-aided drug design. In our quest for potent SERMs, this platform was used to build virtual combinatorial libraries, filter and extract a highly diverse library from the NCI database, and dock them to the estrogen receptor (ER), with all of these steps being fully automated by computational chemists for use by medicinal chemists. As a result, virtual screening of a diverse library seeded with active compounds followed by a search for analogs yielded an enrichment factor of 129, with 98% of the seeded active compounds recovered, while the screening of a designed virtual combinatorial library including known actives yielded an area under the receiver operating characteristic (AU-ROC) of 0.78. The lead optimization proved less successful, further demonstrating the challenge to simulate structure activity relationship studies.



drugs are tamoxifen and raloxifene (Figure 1).<sup>6</sup> While tamoxifen is approved by the FDA exclusively for breast cancer treatment, raloxifene is approved for both breast cancer and osteoporosis treatments.<sup>7</sup> SERMs exhibit an estrogen-like structure that allows binding to the ER, but they differ from estrogens in that they possess a basic side-chain that does not allow the receptor to adopt its active conformation, inhibiting its transcriptional activity.<sup>8</sup> The nature of the chain and its orientation relative to the ligand backbone are central to the design considerations.<sup>9</sup>

Although there is a large body of work in the development of SERMs, this area of research remains very active. In fact, in contrast to full agonists and full antagonists which act similarly in all the tissues, SERMs exhibit activities which are more tissue-dependent and represent an avenue for selective treatment of specific cancers.<sup>10</sup>

**Exploring the Chemical Space.** To date, high throughput screening (HTS) remains the primary technique to identify novel hit compounds (i.e., compounds with observable binding affinity for the target) in the pharmaceutical industry. However, although this approach has been successful, the presence of active

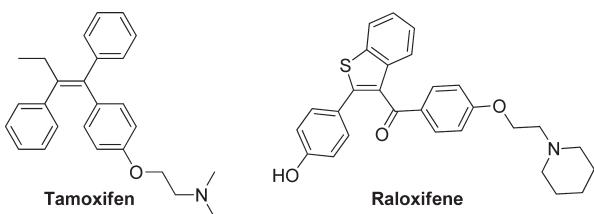
## INTRODUCTION

**Selective Estrogen Receptor Modulators.** Estrogens are the major female sex hormones and are primarily modulators of differentiation, growth, and function of the female reproductive system. In addition, they regulate bone metabolism, modulate function of the central nervous system, and control sexual behavior.<sup>1</sup> 17- $\beta$ -Estradiol (E2) is known to be the most important, dominant, and active endogenous estrogenic hormone. E2 signaling is mediated by two kinds of intracellular estrogen receptors, ER $\alpha$  and ER $\beta$ , which are members of the nuclear receptor superfamily of ligand-inducible transcription factors.<sup>2</sup> The ERs are expressed in a number of endocrine tissues, including the normal breast, uterus, bone, vagina, pituitary, and hypothalamus. They regulate gene expression by binding to DNA response elements associated with their target genes.<sup>3</sup>

The ERs have been a major target for the treatment of many diseases, including osteoporosis and estrogen-dependent breast cancers. ER $\alpha$  is expressed at low levels in normal tissues but is often overexpressed in ER $\alpha$ -positive tumors.<sup>4</sup> Selective estrogen receptor modulators (SERMs) display a selective antagonistic activity or no activity on some tissues but show estrogenic activity (i.e., agonistic) in others.<sup>5</sup> Two well-known examples of these

Received: October 6, 2011

Published: December 01, 2011



**Figure 1.** Selected FDA-approved SERMs.

compounds (hence their discovery) in a given library is not guaranteed nor is their purity. Thus, in order to ensure a successful HTS campaign, one should cover the largest possible fraction of the chemical space describing drug-like molecules (in the range of  $10^{60}$  molecules<sup>11</sup>). In practice, HTS, even when applied to a corporate database of hundreds of thousands of chemicals, still explores a tiny portion of the chemical space. Virtual screening (VS)<sup>12,13</sup> of carefully designed highly diverse libraries can help to cover a larger portion of this space.<sup>14,15</sup> Combinatorial chemistry and parallel synthesis have also been exploited to extend and/or complement the existing chemical libraries. However, while the former still produces a small fraction of the chemical space in the form of mixtures which are often difficult to deconvolute, the latter produces even smaller libraries. Specific libraries can be designed to increase the diversity with a reduced number of structures, and computational chemistry has been used to profile libraries and optimize the design of novel libraries. Finally, fragment-based drug design has appeared as a promising approach, and combining computational and experimental approaches has been seen as the optimal method.<sup>16,17</sup>

Compared to HTS, virtual screening of commercially available structures can provide results within a shorter period of time and with significantly reduced costs, especially when the biological assays are not amenable to the automation required for large compound collections.<sup>14</sup> However, even when computational chemists can potentially consider the approximately 10 million available chemicals, there is room for a significantly larger number of structures and more importantly for more diversity. A combination of HTS and VS approaches has been found to be more efficient than the individual methods.<sup>18</sup> Clearly, *de novo* design techniques would be the method of choice if the synthetic feasibility of the proposed structures was considered efficiently. Unfortunately, metrics for synthetic feasibility are currently available but are either too time-consuming to be applicable to large libraries (e.g., CAESA)<sup>19</sup> or too simplistic to be reliable.<sup>20</sup> Ideally, *de novo* design methods would rely on the large body of knowledge of organic and medicinal chemistry experts.

Herein, we propose to explore the design of target-oriented libraries combining organic chemistry expertise and *in silico* parallel synthesis. We believe that in the short term, the development of methods considering synthetic schemes defined by experimentalists is a more promising strategy than methods designing and predicting the synthetic feasibility of novel structures. In practice, these methods should include the most widely used reactions in medicinal chemistry (e.g., amide formation, reductive amination, sulphonamide formation) and/or the most reliable reactions (e.g., *click* chemistry, ring closure metathesis, Suzuki coupling). But ideally, it should be able to work with any synthetic scheme drawn in a simple molecular sketcher (e.g., ChemDraw) for use by synthetic chemists.

We report our efforts toward (1) the development of a platform for drug discovery that can be used to discover a hit and optimize its binding affinity and (2) the assessment of the platform's usability and accuracy in the context of the discovery of novel SERMs. Being involved in both computational chemistry and organic/medicinal chemistry, the main focus of our group is the integration of computational and experimental chemistry. Ideally, well-validated computational tools and/or established protocols for computer-aided drug design should be applied by medicinal chemists, as they can apply their chemical intuition as well as their expertise in drug synthesis to design novel structures.

## ■ THEORY AND IMPLEMENTATION

**Developed Software.** Over the past few years, we have developed, reported, and made available a number of programs such as ACE,<sup>21,22</sup> SMART,<sup>23,24</sup> and FITTED<sup>23,24</sup> that can predict the stereoselectivity of asymmetric reactions or dock small ligands to proteins and nucleic acids, among other tasks. However, all programs developed to date use a command line interface precluding the wide use by medicinal chemists. Thus, we have complemented these programs with an interface and additional computational tools for drug discovery as described below. Within this interface, chemists can perform a series of actions (using the different programs) on a set of data, preventing user intervention between typical processing steps. This approach is similar in spirit to Pipeline Pilot<sup>25</sup> and KNIME<sup>26</sup> with a major focus in integrating organic and medicinal chemistry knowledge. Below, we describe these programs as well as their corresponding “actions” that can be used in workflows in the FORECASTER platform. For clarity, the programs will be written with small capital letters (e.g., PROGRAMS) while the action names will be written with small capital letters, in italics, and quoted (e.g., “ACTIONS”). Some of the tasks carried out by the developed programs are not directly used for actual drug design and could be carried out with other programs (e.g., 2D to 3D conversion). However, we thought that having access to all of these tools within the same platform would significantly increase its usability by nonexperts.

**The FORECASTER Platform.** In this work, one of our major goals was to make these computational tools available to medicinal chemists in our group who are developing SERMs and to the medicinal chemistry community in general. To do so, we considered the large diversity of operating systems in common use, as members of our group regularly work on Windows, Linux, and/or Mac computers. The large number of computers is also a factor that we took into account: in general, occasional users rarely take the time to update their programs to newer versions on their personal computers; hence users within a group may have different versions. To address the operating systems variability and the difficulty in regularly updating the programs on all of the computers while developing this interface, we opted for a Web-based application as a centralized platform-independent solution. The ubiquity of Web browsers in modern operating systems greatly facilitates the distribution and maintenance of the platform, limiting the need for updates only to the application server rather than to every single workstation of our research group.

**Processing Small Molecules. Generating Good Quality Small Molecule Structures.** For several computer-aided drug design methods, 3D structures of the small molecules are required. On one hand, one can generate these 3D representations directly by building a molecule using a graphical interface such as MOE

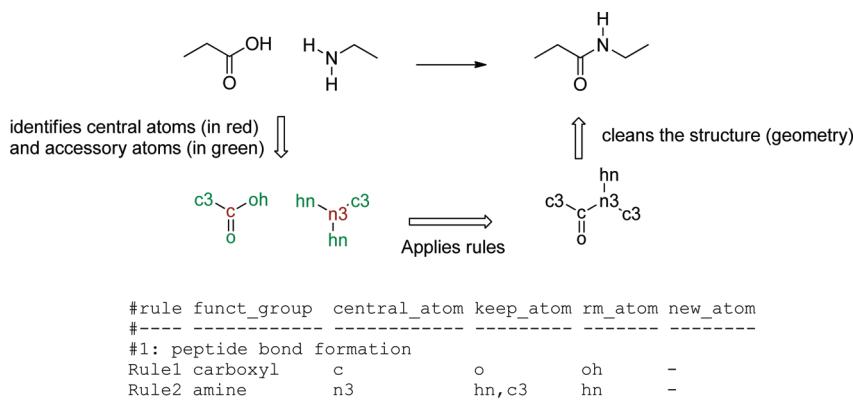


Figure 2. Functional group depiction in REACT.

(Chemical Computing Group)<sup>27</sup> or Maestro (Schrödinger).<sup>28</sup> On the other hand, one can use a 2D sketcher more popular in the organic chemistry community such as ChemDraw<sup>29</sup> and then a 2D to 3D converter. In this context, we have developed CONVERT (Conformational Optimization of Necessary Virtual Enantiomers, Rotamers and Tautomers) to convert structures from 2D-sdf format (available in ChemDraw<sup>29</sup>) into all-atom mol2 format 3D structures. When the action “CONVERT 2D TO 3D” is selected, CONVERT expects a small molecule or a library of small molecules. It then (i) defines the hybridization states for each atoms of this molecule on the basis of the provided bond orders; (ii) estimates the nitrogen pK<sub>a</sub>’s on the basis of hybridization and aromaticity; (iii) identifies possible tautomers (e.g., imidazole rings) or stereoisomers (including ammonium nitrogens); (iv) creates out-of-plane perturbation at all stereogenic centers; (v) adds hydrogen atoms; (vi) optimizes the bond lengths, angles, and torsions on the basis of geometric rules and force field parameters; (vii) carries out a conjugate gradient energy-minimization using molecular mechanics; and (viii) outputs all of the tautomers/isomers in a multi-mol2 file.

Although this procedure can be seen as straightforward, a number of molecules requiring more specific attention have been found by screening the sdf library of the Sigma-Aldrich catalog,<sup>30</sup> a common source of chemicals for medicinal chemists. First, if a Fischer projection is given, it is automatically recognized by CONVERT, which converts the 2D representation into a 3D structure, following Fischer’s rules. Second, stereogenic atoms are identified, and when the option is selected, CONVERT skips the molecule if no chirality is assigned (for example, no chirality was assigned to glucose in the sdf library of the Aldrich catalog). Third, if a salt (e.g., hydrochloride salts of amines) is given as input, the two components are identified and their atoms renumbered, and only the largest one is kept.

Three-dimensional molecular structures may be stored in different formats, and there are programs that may generate these file sometimes with subtle changes. To simplify the access to the necessary molecular information, we previously created the program SMART (Small Molecule Atom typing and Rotatable Torsion assignment),<sup>23</sup> which can read a molecule and generate pseudo-mol2 files containing many precomputed features of the molecule. SMART was initially developed to prepare ligands prior to their docking with FITTED<sup>23</sup> and is used to “SETUP LIGAND(S) FOR DOCKING” in the FORECASTER platform. In a more recent version, we modified SMART to write strings describing the ligand on the basis of descriptors (e.g., molecular weight, number of rotatable

bonds, presence of given functional groups).<sup>31</sup> This second implementation is used to “ADD DESCRIPTORS”. In the most current version, we also implemented different charging schemes such as the MMFF charges<sup>32</sup> and the electronegativity equalization method.<sup>33</sup> It can also identify specific (e.g., aldehyde, sulfonyl chloride) and generic functional groups (e.g., X, Y) for use in combinatorial chemistry; the latter is the core action of “SETUP REACTANT(S) FOR COMBINATORIAL CHEMISTRY”.

Overall, the current version of SMART (i) reads molecular structure files (sdf or mol2 formats), both single molecule or multiple molecules; (ii) creates a Z-matrix representation of the structures; (iii) identifies all of the rings; (iv) assigns bond order and atom hybridization; (v) identifies aromatic rings and resonance structures; (vi) assigns generalized Amber force field (GAFF)<sup>34</sup> atom types or dummy atoms whenever required; (vii) identifies functional groups based on a list of definition (in Supporting Information); (viii) assigns molecular descriptors; and (ix) outputs single or multi-mol2 files.

*Filtering out Undesired Molecules.* REDUCE (Recognition and Elimination by Descriptors of Undesired Chemical Entities) has been developed to read the strings generated by SMART (“ADD DESCRIPTORS”) and to filter out any undesired molecules. This function was initially implemented into FITTED.<sup>31</sup> The selection/filtering criteria are a list of user-defined descriptor rules (e.g., molecular weight <500, presence of aromatic rings, and presence of one carboxylic acid functional group). This program was later implemented as the action “FILTER BY DESCRIPTORS” in the FORECASTER platform.

*Building Combinatorial Libraries.* REACT (Rapid Enumeration by an Automated Combinatorial Tool) is another new program developed to prepare combinatorial libraries based on a defined synthetic scheme. This tool is the direct link between organic/combinatorial chemists and the other computational tools. When developing an *in silico* combinatorial chemistry tool, the major issue is the encoding of chemical reactions.<sup>35</sup> Ideally, the chemical transformations would not be limited to preset chemical reactions; instead, the user could apply the program to any reaction drawn within a simple two-dimensional file (e.g., ChemDraw format). On the basis of these premises, we have developed REACT, which, given a pair of virtual libraries of reactants and a set of rules for their conversion into products, produces a new virtual library of compounds. While this program was developed, Reactor from ChemAxon came out.<sup>36</sup> For an optimal integration in our platform and more specifically with our docking program FITTED, the representation of input and output molecules was to

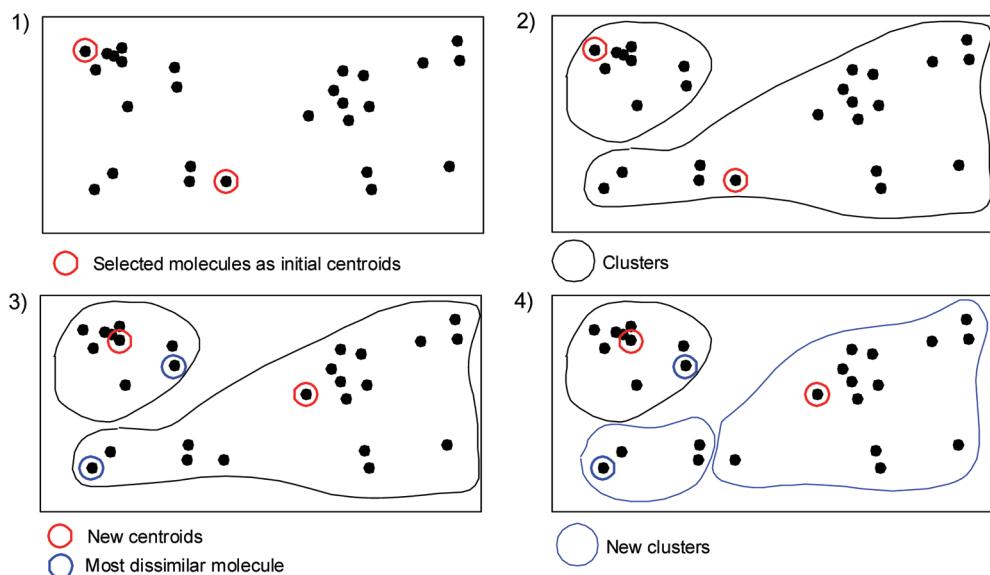


Figure 3. Clustering process used by SELECT.

be done in 3D coordinates, therefore yielding a library of compounds that would be ready for docking with FITTED or for further calculations with any other structure-based drug design tool. An action “CLEAN STRUCTURE GEOMETRY”, which includes a molecular mechanics routine for geometry optimization by conjugate gradient energy minimization, was created.

Within FORECASTER, 3D representations of chemicals can be generated from a 2D scheme using the action “CONVERT 2D TO 3D” described above. Then, the reactants must be prepared using the action “SETUP REACTANT(S) FOR COMBINATORIAL CHEMISTRY”. Finally, REACT can be used through the action “CREATE COMBINATORIAL LIBRARY”. With this action, the user selects the chemical transformation from a pulldown menu (linked to a set of rules corresponding to the given reactions) or defines a new one using the reaction manager. For an optimal use of this program by synthetic chemists, we have implemented a sketcher based on ChemWriter into the platform and an automatic rule generator which creates rules directly from a synthetic scheme (for examples of rules, see Figure S1, Supporting Information).

Within REACT, a transformation (e.g., amide bond formation) occurs between reaction centers in the reactant molecules. A reaction center (Figure 2) is defined as a functional group (e.g., carboxylic acid), using a circular fingerprint consisting of a central atom (e.g., carboxylic acid carbon) bound to accessory atoms (e.g., two carboxylic acid oxygen atoms). To carry out a transformation, REACT first identifies whether the necessary functional groups (e.g., carboxylic acid and amine) are present in the reactants. The central atom and accessory atoms are defined by GAFF<sup>34</sup> atom types or by an element or an atom name (Figure 2).

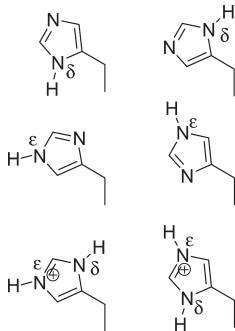
With all of these definitions, REACT reads the libraries provided for reactants 1 and 2 (or just reactant 1 in cases such as a reduction by hydrogen) and tries to match a rule to each chemical in these libraries. If a match is found, the reactants proceed with the chemical transformation. They are otherwise skipped. During the transformation, atoms to be removed (`rm_atom`) in the rules (Figure 2) are stripped, and a new bond is formed between the reactive centers (`central_atom`). At this stage, a quick optimization of the overall shape of the molecule is carried out.

A systematic rotation about the newly formed bond is then performed, and a conformation with as few steric clashes as possible is saved into an output file. This process is iterated for all of the reactant 1/reactant 2 pairs, and a combinatorial library is thus generated. The current version does not consider stereochemistry. Thus, asymmetric transformations would produce a randomly selected stereoisomer; however, asymmetric synthesis is very rarely used in combinatorial chemistry.

Organic chemists may wish to filter the reactant libraries prior to carrying out the combinatorial chemistry to consider functional group compatibility. For instance, any unprotected amino acid (featuring both amine and carboxylic acid functional groups) should not to be used in peptide coupling, as it may react with itself and polymerize. Similarly, a dicarboxylic acid should not be used, as it could react twice. This can be done using the action “FILTER BY DESCRIPTORS” prior to “SETUP REACTANT(S) FOR COMBINATORIAL CHEMISTRY”. This example reveals the interconnection of all of the actions described so far and the need to ensure the maximum compatibility of the input and output file formats required for easy use by experimentalists.

To trace the origin of the building blocks, a new field (@<TRIPOS> FRAGMENTS) has been added to the mol2 files. For instance “3 2 5 6” provides first the number of building blocks (3) and their ID in their respective reactant library (2 5 6). Thus, this string corresponds to a molecule made of three building blocks: building block #2 from the first library, #5 from the second library, and #6 from the third library (see application below). This additional field is kept throughout the other actions (this field is read and written by all of the other programs, whether they need it or not) and can therefore be read at any point in a workflow.

**Selection of Diverse Libraries and Search for Analogues.** SELECT (Selection and Extraction of Libraries Employing Clustering Techniques) has been developed to carry out two major tasks: “EXTRACT REPRESENTATIVE LIBRARY” and “SEARCH FOR ANALOGUES”. With the former action, SELECT can read a library and cluster the structures by similarity using the MACCS keys<sup>37</sup> and the Tanimoto coefficient. The clustering is carried out through a hierarchical divisive method. Two highly dissimilar molecules are



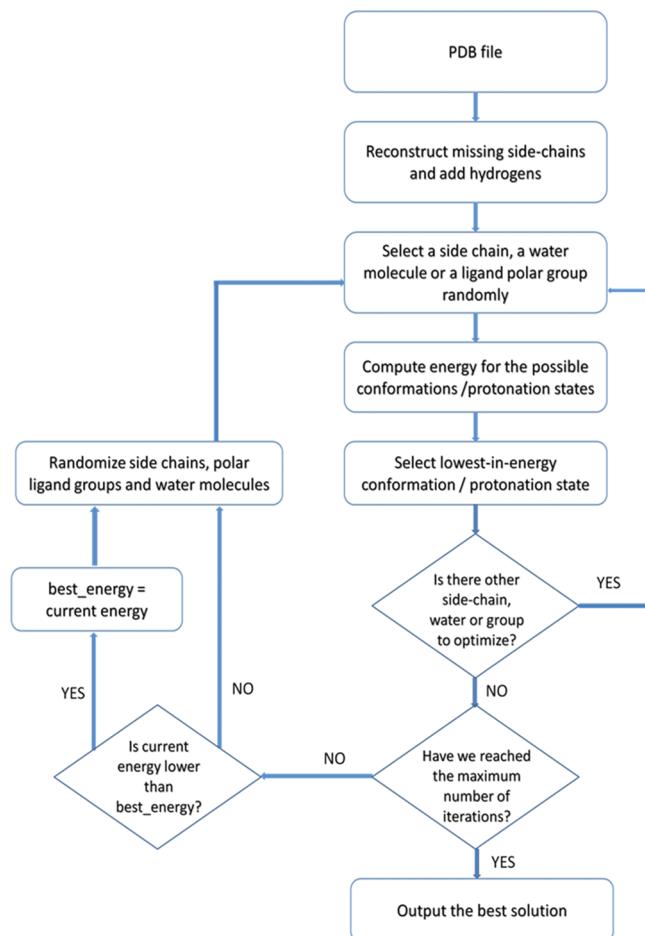
**Figure 4.** Tautomers, rotamers, and protonation states for a histidine residue.

first identified then selected as cluster centroids (first step in Figure 3). All of the other molecules are assigned to one of these two centroids (i.e., clusters) on the basis of their similarity with them (second step in Figure 3). The centroids of these two clusters are next identified together with the most dissimilar molecule for each cluster (third step in Figure 3) and the most diverse cluster selected for further splitting. The initial centroid and its most dissimilar molecule are used as centroids for the next iteration (fourth step in Figure 3). In this way, the most diverse cluster is split into two smaller clusters. This procedure (steps 3 and 4 in Figure 3) is iterated until SELECT reaches the requested number of clusters or until the diversity of all of the clusters is reduced below a user-defined level (e.g., Tanimoto coefficient of all pairs within all the clusters is greater than 0.60). As a final output, SELECT produces a representative diverse library constituted of the centroids of these clusters.

With “SEARCH FOR ANALOGUES”, SELECT can read in a structure (e.g., a hit compound) and a library (e.g., a database of available chemicals) and identify analogues to the hit(s) within this library. This is carried out by measuring the Tanimoto coefficient between the MACCS keys<sup>37</sup> of the hit(s) and each molecule of the library.

With both actions, SELECT outputs a library of the selected molecules being either a library of cluster centroids in “EXTRACT REPRESENTATIVE LIBRARY” or a focused library of analogues with “SEARCH FOR ANALOGUES”.

**Processing Macromolecules. Converting PDB Files into Good Quality mol2 Files.** As Protonate3D (Chemical Computing Group)<sup>38</sup> and the Protein Preparation Wizard from Schrödinger,<sup>39</sup> PREPARE (Protein Rotamer Elaboration and Protonation based on Accurate Residue Energy) has been developed specifically to carefully prepare all-atom protein files from PDB files. Over the years, feedback from FITTED users revealed that many failures were due to improperly prepared protein mol2 files from PDB files. For instance, missing atoms in residues, misassigned protonation states, unrecognized changes in the PDB format made by some platforms, and unoptimized hydrogen positions have impacts on the accuracy of our docking program. Due to the similarity in the electronic density around the NH<sub>2</sub> and O from the amide group in the side chains of asparagine and glutamine residues, X-ray crystallography cannot clearly assign the rotamers on these two side chains. Similarly, the histidine imidazole rings can adopt two rotameric conformations as well as three protonation states that could match the electronic density (Figure 4). The reconstructed side chains can also adopt several conformations (limited to the ones present in the conformational library) while water molecules can have multiple orientations.

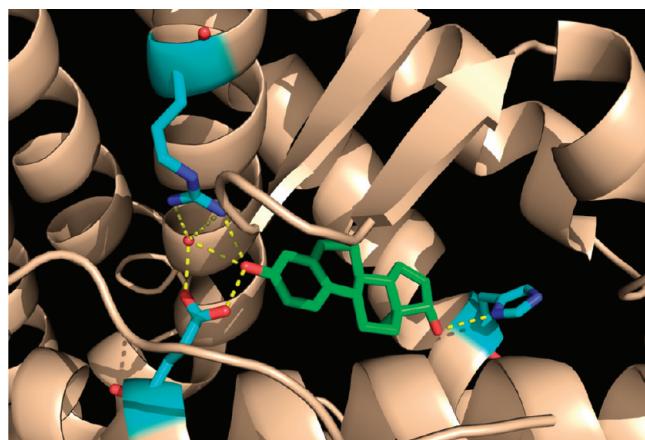


**Figure 5.** Flowchart of the procedure used by PREPARE to optimize PDB structures.

Alternative side chain conformations are also sometimes given in the PDB file. All of this structural information is often ignored or misused.

PREPARE was developed to handle all of these potential issues associated with the conversion of a PDB file into a usable high quality mol2 file. An iterative optimization algorithm attempts to identify the best combination of rotamers (e.g., serine hydroxyl groups, ligand ammonium groups), protonation states (e.g., histidine residues), orientation of water molecules, and reconstructed side chain conformations. The algorithm optimizes these degrees of freedom in a random sequence. A group (e.g., ligand functional group or side-chain) is randomly selected, and all of the conformations (orientation in the case of water) and/or protonation states (e.g., six conformations for histidine, Figure 4) are considered (Figure 5). The conformations of the reconstructed side-chain are investigated considering a conformation library as defined by Lovell et al.<sup>40</sup> For each conformational change, the potential energy of the newly constructed conformation is computed with the GAFF force field and compared to the previous conformations. The final conformation is archived, and all of the degrees of freedom are randomized before the process reiterates. By default, the process stops after 10 iterations. Additional iterations may be necessary if several water molecules or flexible residues (e.g., reconstructed side-chains) are in close proximity and can interact with each other in several alternative complex networks.

In order to optimize the orientation of the water molecules, 100 evenly distributed points on a sphere around each oxygen atom are precomputed. Whenever a water molecule is selected

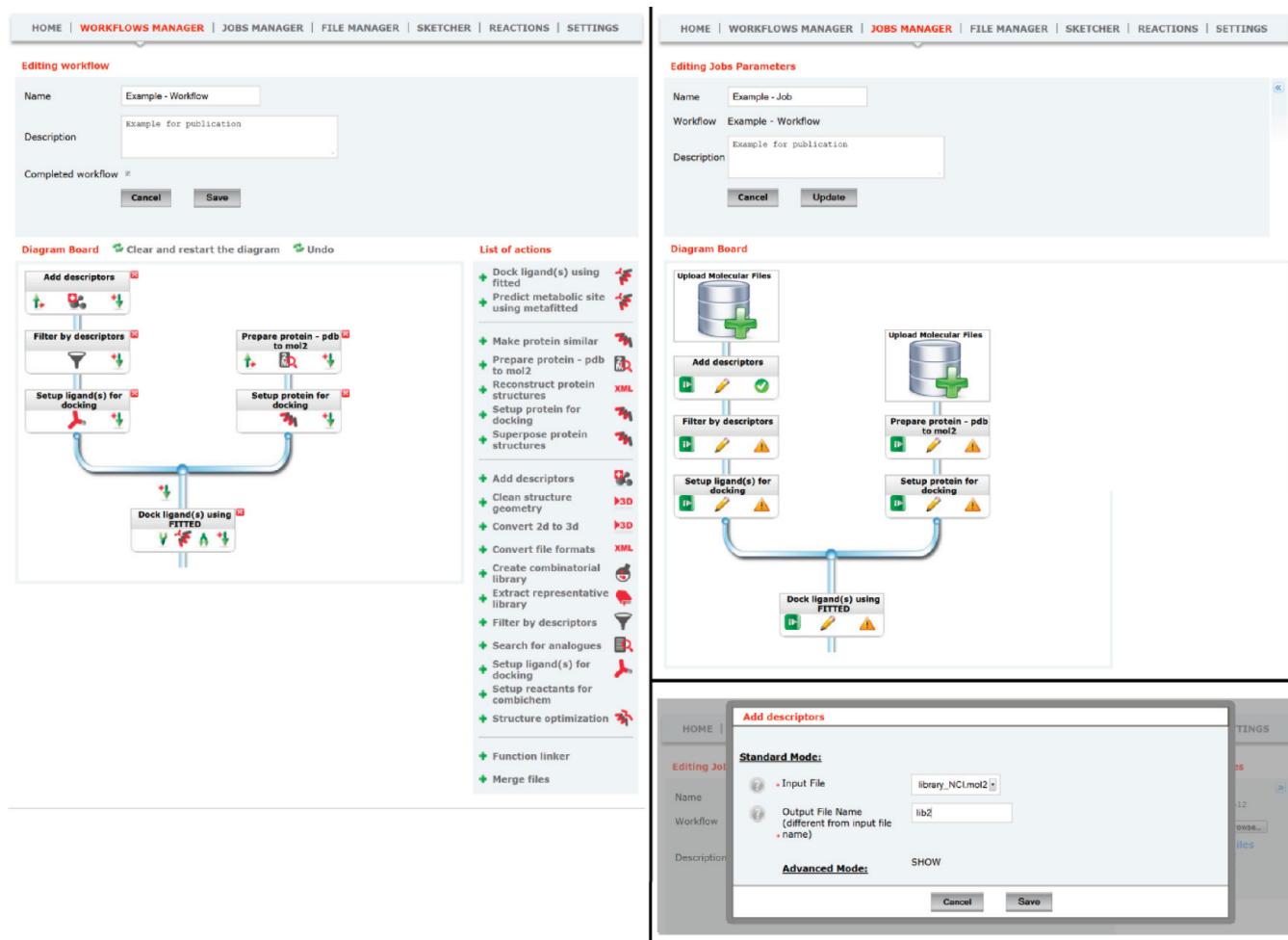


**Figure 6.** Crystal structure of ER bound to 17 $\beta$ -estradiol (PDB code: 1ERE).

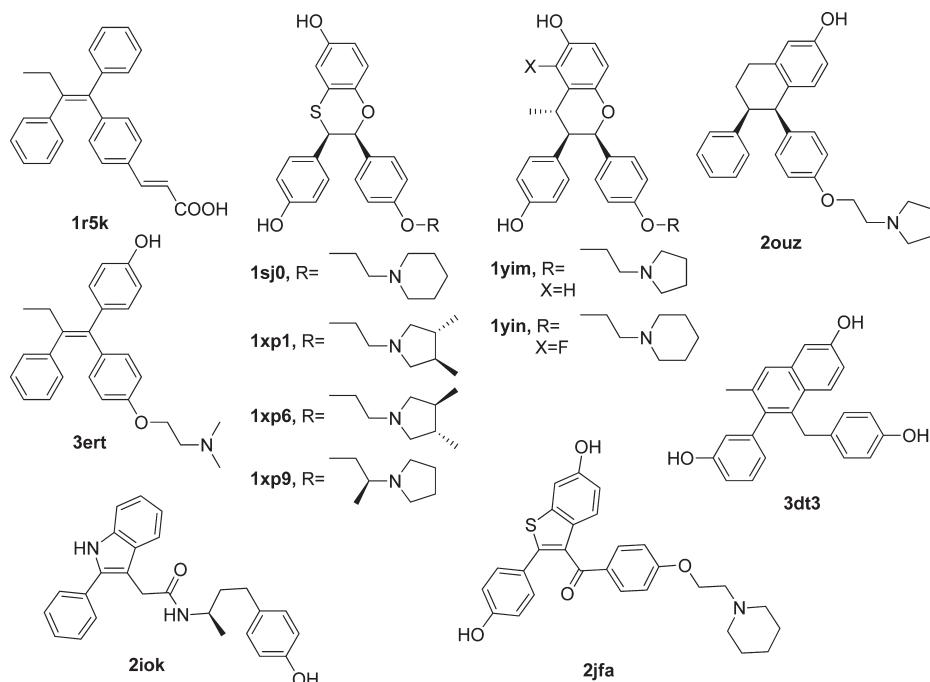
for optimization, a hydrogen probe is placed on each of these hundred points, and its best six locations are saved. Then, the second hydrogen is added onto these six alternative water molecules and rotated by 18 steps of 20°. The lowest in energy of these 108 orientations is kept. Similarly, the polar groups (e.g., serine hydroxyl groups) are rotated by steps of 20°, and the conformation with the lowest energy is kept.

Although all of the steps are fully automated, the user can also instruct the program to keep or remove a cofactor (recognized by PREPARE as a ligand) as part of the protein or to protonate a carboxylic acid (e.g., aspartic acid in HIV-1 protease catalytic site). It can also force a given hybridization for a ligand atom that might be distorted in the original PDB file, hence resulting in a misassigned hybridization state.

When the action “PREPARE PROTEIN - PDB TO MOL2” is selected, PREPARE (i) reads a PDB file, (ii) separates the protein from the ligand, (iii) reconstructs missing side chains using a conformational library,<sup>40</sup> (iv) adds hydrogen atoms to the protein and water molecules, (v) identifies alternative conformations, (vi) identifies alternative protonation states (e.g., histidine, Figure 4), (vii) optimizes all of these degrees of freedom in the presence of



**Figure 7.** Left panel: Creating workflows using FORECASTER. The actions listed on the right menu can be picked and assigned to boxes. The green arrows are used to add boxes to the workflow. Right panel: Using a workflow in Job manager. Within each box, the green tick indicates that the necessary keywords are given while the yellow exclamation mark indicates that parameters should be set by clicking on the pencil (i.e., name of input file). The bottom panel shows the menu appearing when clicking on the pencil of the “ADD DESCRIPTORS” action. The question mark buttons are linked to help. The arrows in green boxes indicate that this action will be performed while a “pause” button indicates that the workflow stops before this action (see text).



**Figure 8.** Ligand structures used for cross-docking experiments.

the ligand, and (viii) outputs the best conformations, according to the computed energy.

When docking with FITTED to flexible proteins, multiple protein conformations need to be provided. If one uses different PDB files, the sequence must be identical and the structures superimposed. When the action “*MAKE SIMILAR*” is selected, PREPARE extracts the primary sequences from the PDB files and first carries out a sequence alignment, which is next used to tether binding site  $\alpha$  carbons for superposition. This sequence alignment is also used to identify the differences between the primary sequences of the input structures. Deletion and mutation are then carried out to ensure that all of the structures differ only by the Cartesian coordinates (i.e., conformations). If a deletion and/or mutation is performed within the binding site, a warning message is issued. This action is to be used prior to “*PREPARE PROTEIN - PDB TO MOL2*”.

**Setting up Protein Files for Docking.** As soon as the PDB structures are optimized and converted into a mol2 file (“*PREPARE PROTEIN - PDB TO MOL2*”), they can be processed by PROCESS (PROtein Conformational Ensemble System Setup) for their use with our docking program FITTED. This action is labeled as “*SETUP PROTEIN FOR DOCKING*”. PROCESS can take more than one file as an input (several protein conformations can be used by FITTED to simulate protein flexibility upon docking) and prepare them all at once for docking. PROCESS also prepares the cavity files and identifies the interaction sites, all of this information being used by FITTED for optimal docking as described previously.<sup>23,24,31</sup>

**Docking with FITTED 3.0 (Flexibility Induced Through Targeted Evolutionary Description).** As soon as the ligand(s) and protein files are ready (with “*SETUP LIGAND(S) FOR DOCKING*” and “*SETUP PROTEIN FOR DOCKING*”), they can be processed by FITTED (“*DOCK LIGAND(S) USING FITTED*”). Our docking program, FITTED,<sup>23,24,31</sup> and its associated scoring function, RankScore,<sup>41,42</sup> have been described in great detail previously, and only a brief description is given here. This docking program

uses a hybrid matching algorithm/genetic algorithm to dock the ligand(s) to the protein (or nucleic acid<sup>43</sup>) binding site. When creating the initial population of poses, a set of filters (shape, interaction site match, force field energy) is used to select the best poses for further evolution. When the ligands are docked, a summary file is generated which includes the molecule name, fragment identity (as defined by REACT, see above), the scores, and potential energies.

## ■ RESULTS AND DISCUSSION

**Computer-Aided Discovery of Estrogen Receptor Modulators.** A variety of computational methods, including different docking programs and QSAR models, have proven to be successful in the identification process of new ER antagonists.<sup>44–49</sup> The first crystal structure determination of the ER ligand-binding domain (LBD)<sup>50</sup> provided insight into the mode of action of ER ligands at the molecular level (Figure 6). Currently, more than 60 crystal structures of both ER $\alpha$  and ER $\beta$  bound to a variety of agonists and antagonists are available in the protein data bank (PDB).<sup>51</sup> This collection of structural information reveals the presence of a well-defined binding cavity and crucial ligand–protein interactions inside the binding pocket. Additionally, binding affinity data for many known agonists and antagonists are available (see for example the ChEMBL database: <https://www.ebi.ac.uk/chembl/>). These studies have definitely contributed to the ER being one of the most successful examined targets for VS as a lead identification method.<sup>52</sup> ER $\alpha$  was first used as a VS target by Baxter and co-workers for the validation of PRO LEADS, a Tabu search-based flexible docking method.<sup>53</sup> This report was followed by Bissantz and co-workers, who evaluated the proficiency of three docking programs in combination with seven different scoring functions to discriminate 10 known actives from a random database of 990 ligands;<sup>54</sup> they concluded that all three docking methods could differentiate between true hits and random ligands.

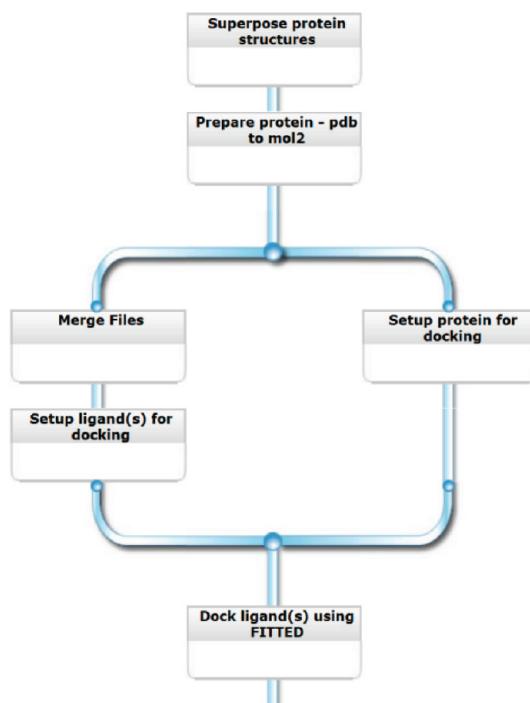


Figure 9. Workflow for cross-docking.

The DockCrunch project performed large-scale VS to identify novel drug leads.<sup>55</sup> Using ER $\alpha$  as a target in both its agonist and antagonist conformations, they screened a prefiltered library of 1.1 million commercially available compounds enriched with 20 known actives. When applying PRO\_LEADS as the docking algorithm and ChemScore as the scoring function, they not only obtained a good discrimination between agonists and antagonists but they also discovered a set of 37 new ligands with high binding affinity for the receptor.

**Integrating Computational Chemistry and Medicinal Chemistry.** When developing the FORECASTER platform, our major goal was the integration of tasks requiring expertise in computational chemistry and organic/medicinal chemistry. While the former could set parameters for a given calculation, develop a protocol based on several tasks, and integrate a new program, the latter could run routine tasks such as docking of a compound on a validated protein structure or other tasks that can be carried out by medicinal chemists who can take advantage of their expertise in synthesis (is the molecule feasible?) and medicinal chemistry (is the molecule potentially toxic, water-soluble, etc?) to prioritize the molecules to investigate. Our group both develops programs and applies them to guide the synthesis of molecules, and we have found the integration of both areas of expertise to be beneficial for the entire research group. In the current research project, medicinal/synthetic chemists in our group were the major testers.

To improve the usability of all of the developed software, these programs were implemented in the Web-based interface FORECASTER in the form of actions and the keywords used by these programs in the form of options. These actions can be easily assembled by computational chemists into workflows (Figure 7 left panel), and the flow of files from one action to another is taken care of by the platform. This workflow can then be applied to selected molecules and targets by medicinal chemists (Figure 7 right panel). ChemWriter (version 1), a Java applet,<sup>56</sup> has also

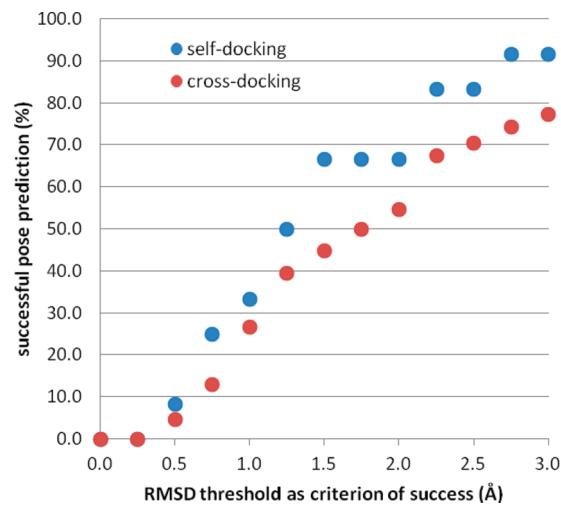


Figure 10. Docking accuracy as a function of RMSD threshold for the docking of ER antagonists with FITTED.

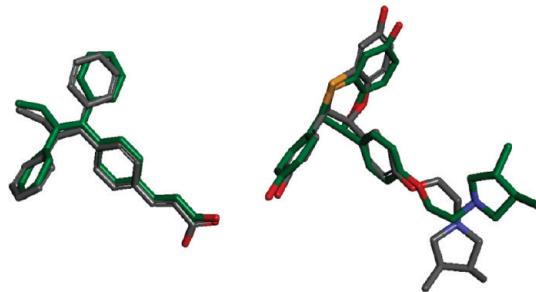
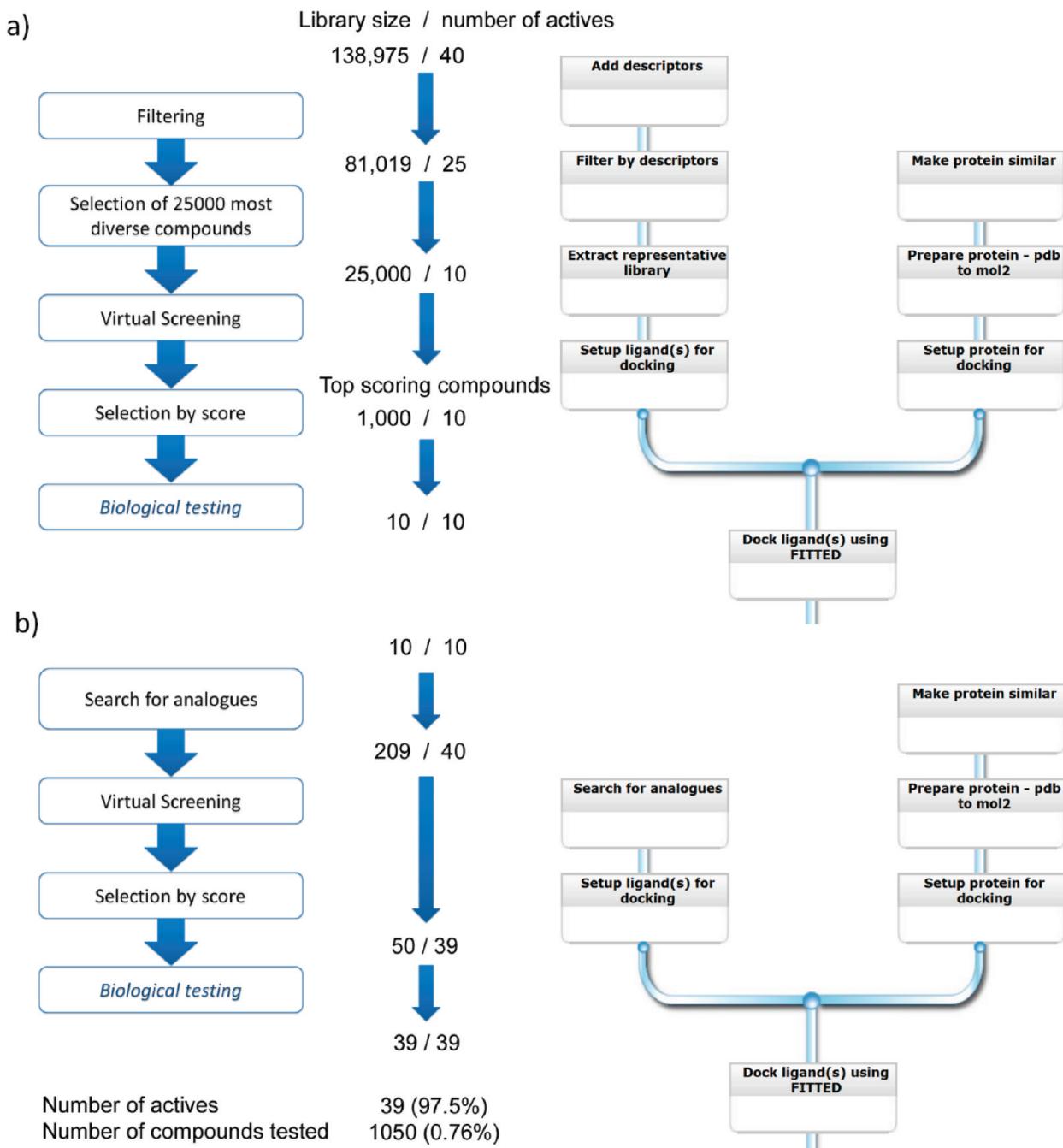


Figure 11. Docked poses (gray) and X-ray binding modes (green) of 1r5k (left) and 1xp1 (right) ligands when docked to the 1R5K receptor structure. 1r5k: RMSD = 0.60 Å. 1xp1: RMSD = 2.75 Å.

been implemented in order to allow the user to draw new molecules directly in the workflow or create new synthetic schemes for combinatorial chemistry under the REACTIONS tab. A number of “help” buttons have been implemented to ensure an optimal use. Computational chemists or server administrators can also add new actions (i.e., new plugins) and handle user accounts and files in the SETTINGS section (last tab). The option to add new actions was essential to allowing the easy integration of any other program. When adding an action, the user specifies the type of commands (command line with arguments or keyword file), the expected keywords, their type (i.e., file, string or integer) and their default values, the path to the executable, and a description of this action. The platform takes this information to create an action and the corresponding window of parameters.

**Current Status of This Platform.** As discussed in the Theory and Implementation section, a Web-based interface significantly reduces the burden associated with frequent updates on a large network of computers as well as the variety of environments (e.g., old –32 bits— vs more recent –64 bits— Windows environments, various Linux distributions, Mac OS) as updates have to be done on a single server. This is a major advantage for use by large teams of experimentalists. However, the development of a Web-based interface also faces issues that, for some of them, are yet to be addressed. First, we have found that the visual aspect is

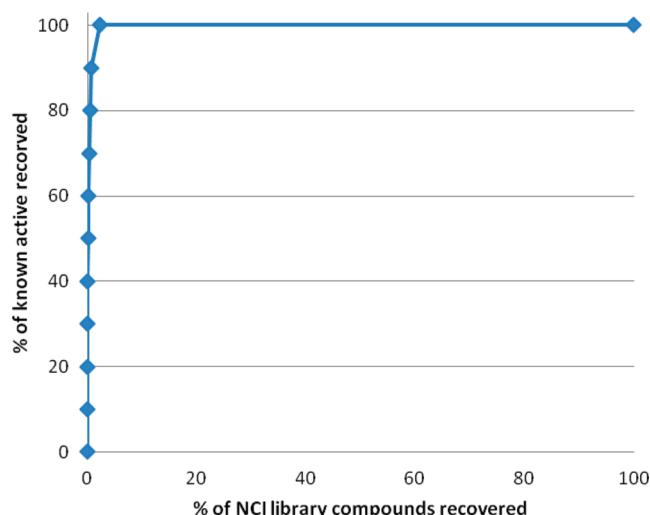


**Figure 12.** Designed protocols (left) and FORECASTER workflows to identify active compounds from a library and its application to the NCI library seeded with known bioactive compounds. (a) Filtering, selection, and docking-based virtual screening of the library; (b) search for analogues of identified hits.

somewhat affected when using a browser version different from the one we used during the development of this interface. Second, as of now, the number of simultaneous users has never exceeded four, and our server has been sufficient to run the simultaneous jobs. However, an interface to popular queuing/distributed computing systems (such as SunGridEngine, TORQUE, Condor) should be developed for larger teams or supercomputers. Third, another limitation of this Web interface is the use of Jmol as a visualizer. The limited functionalities of this applet make the visualization of poses difficult. As a fifth limitation, currently, the analysis of the results of the calculations (e.g., ROC computation, scores of hits) requires the use of external

programs such as spreadsheets, although this type of program is traditionally available on most computers. Finally, the generation of large files or of a large number of files in the same directory leads to difficulty in either opening or searching files. A more advanced file manager would be required.

**Cross-Docking of ER Modulators.** Several crystal structures of SERMs cocrystallized with ER exist and were first used to evaluate the ability of our programs to predict the binding modes and to test the workflow functionalities. We selected 12 structures from the PDB, focusing on antagonists (1R5K, 1SJ0, 1XP1, 1XP6, 1XP9, 1YIM, 1YIN, 2JFA, 2I0K, 2OUZ, 3DT3, and 3ERT, Figure 8). From now on, the ligands and receptor structures

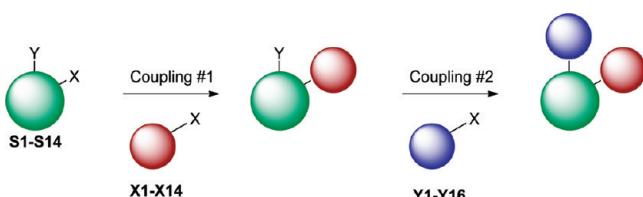


**Figure 13.** Receiver operating characteristic curve for the virtual screening of 24 990 diverse NCI compounds and 10 known actives. The NCI library compounds are considered inactive.

will be referred to using lower case and upper case PDB codes, respectively. Hydroxytamoxifen and raloxifen have been cocrystallized with ER more than once (2BJ4 and 2JF9 as well as 3ERT, 1ERR, and 2JFA, respectively), but only one structure for each was considered. The workflow shown in Figure 9 was built. Through this workflow, the PDB receptor structures were superimposed by automatic identification of the primary sequences of each of the receptors while the binding site was located using the cocrystallized ligands to focus the superimposition on the binding site residues. The ligands and receptors from the PDB files were then split, and hydrogen atoms were added and optimized (second box in the workflow). In the left branch, ligands were then combined into a single file and prepared for docking. In the right branch, the receptor structures were prepared for docking and given as input to the docking program FITTED. When the cross-docking mode is selected with FITTED, the program reads the many target structures but uses them separately in rigid-protein mode. Thus, all 12 ligands given in Figure 8 were individually docked to the 12 receptor structures, and the deviations to the observed binding modes (RMSD) were computed.

Traditionally, a threshold of RMSD = 2.0 Å is used to define whether a pose is correctly predicted or not. In our case, we are adding error by cross-docking on superimposed protein structures. This superimposition is never perfect as the receptor structure varies significantly and RMSD between the  $\alpha$  carbons of the 12 structures can be as large as 1.1 Å. Thus, we selected a threshold of RMSD = 2.25 Å for cross-docking and 2.0 Å for self-docking as described previously.<sup>24</sup> When these criteria are selected, the ER antagonists are docked back (i.e., self-docked) correctly 67% of the time, while they are cross-docked successfully also in 67% of the cases (Figure 10). This accuracy is higher than observed previously with other receptors and proteins.<sup>31</sup>

In addition to RMSDs, visual inspection was carried out for docked poses greater than 2 Å. For instance, when docking ligands such as 1xp1 with an ammonium-containing side-chain to the receptor structure 1RSK (in which the ligand 1r5k has a carboxylate on the side chain), RMSDs on the order of 2–3 Å were observed. In these cases, the core was well docked, but the



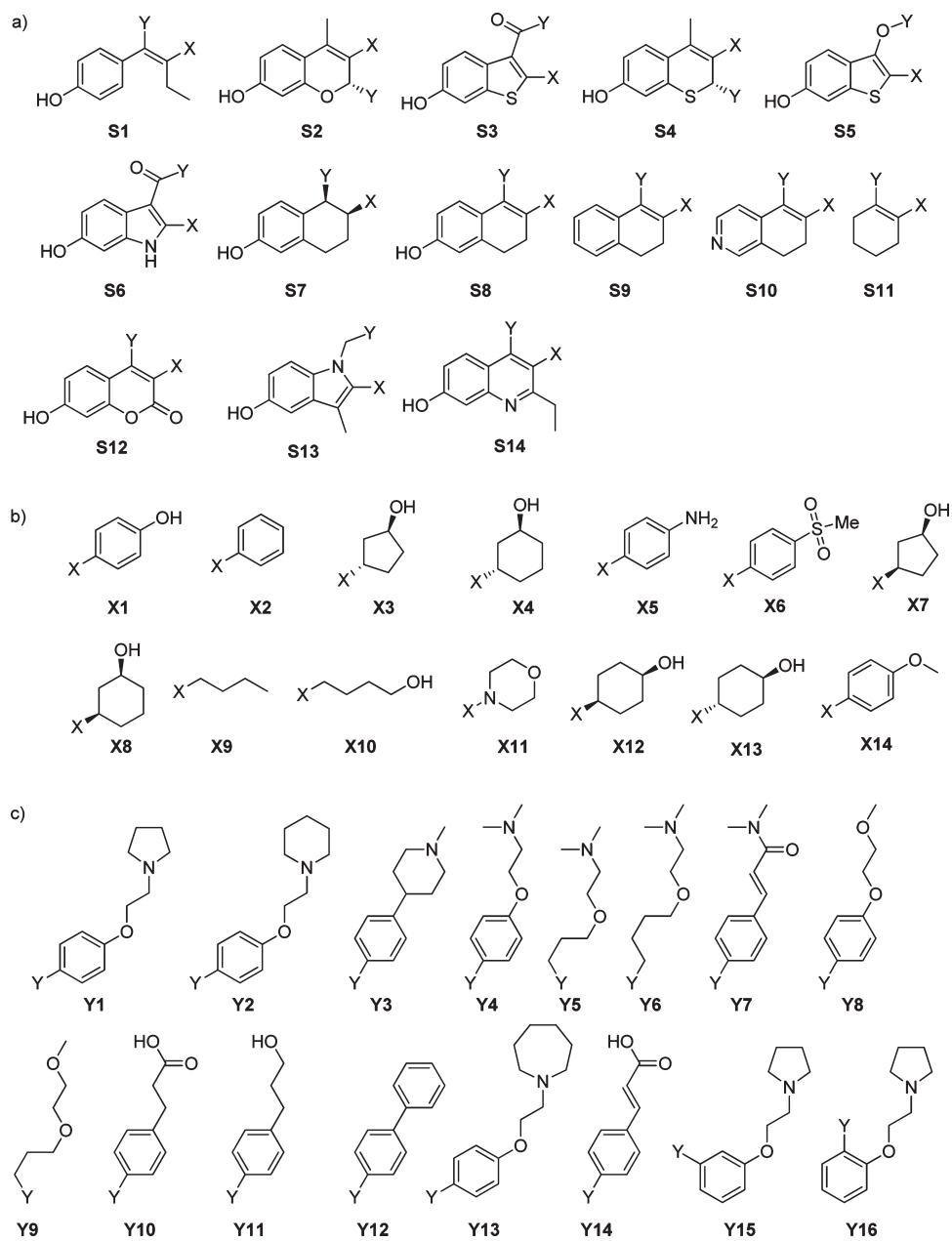
**Figure 14.** Virtual synthesis of a combinatorial library.

side chain was misplaced (Figure 11). 2iok and 2jfa feature similar motifs (Figure 8) but bind very differently. Interestingly, the unexpected observed binding mode of 2iok was not retrieved by FITTED, which predicted a binding mode for 2iok very similar to that of 2jfa, resulting in an RMSD of 6.82 Å. Overall, the workflow ran uneventfully, and the pose prediction was accurate.

**Application to the Screening of a Large Database on a Flexible ER Structure.** We are involved in the development of SERMs and, as a validation study, we looked at realistic scenarios for the discovery of novel entities. In a standard drug discovery program, the first step is to identify hits. In this context, computational methods can be used to prioritize compounds for biological testing and/or search for analogues of existing hits (me-too drugs). A first protocol for the identification of potential ER antagonists was designed (Figure 12). The corresponding workflow was developed (Figure 12a) for filtering and screening the NCI database (available from the ZINC database<sup>57</sup>) on a flexible estrogen receptor. In order to assess the accuracy of the protocol and programs, 40 known antagonists (given as Supporting Information) were added to this database for a total of 138 975 compounds. This workflow iterated through the database, added descriptors to all of the molecules, and filtered them for physical properties and undesired functional groups. Among the selected filters are the following: molecular weight  $\in [250,650]$ , net charge  $\in [-1,+1]$ , and number of rotatable bonds  $\in [0,7]$ . Molecules featuring known reactive or undesired functional groups including aldehyde, azide, boronate, and acyl halide were also removed. At this stage, 81 019 (58%) compounds were kept, including 25 actives, with several of the 40 actives being filtered out for their flexibility (number of rotatable bonds). To increase the speed of the protocol, a subset of representative structures was selected. For this purpose, the remaining molecules were clustered (using SELECT up to a maximum of 25 000 clusters or a minimum Tanimoto coefficient of 0.80 for all of the clusters). The first criterion was fulfilled, and the centroids of each of the 25 000 clusters were extracted to build a representative library. At this stage, 10 active compounds were conserved. These 25 000 compounds were next prepared for docking. On the right branch of the workflow illustrated in Figure 12, three PDB files of ER $\alpha$  cocrystallized with three antagonists were given as input. Their sequences were aligned and the structures superimposed, and the resulting structures were set for docking with FITTED.

Instead of completing the workflow on a desktop computer, the option to pause the workflow at any step in the workflow was selected right before the docking, and the files generated by the workflow were transferred onto a large computer cluster. The 25 000 compounds were docked in just a few hours. The area under the receiver operating characteristic curves (AU-ROC) has often been used as a metric, to define whether a program can discriminate active from inactive compounds.<sup>58,59</sup> If this value is 1.0, the discrimination is optimal; if it is 0.0, the discrimination is also optimal but inverted (all of the actives are predicted as

Chart 1. Selected Building Blocks



inactive and vice versa). If the value is 0.5, the discrimination is random. In this study, the fairly low number of active compounds (10) does not provide a truly representative study of the program accuracy. Nevertheless, analysis of the docking data revealed that all of these 10 known actives were ranked within the top 2.5% with nine in the top 0.75%, and a nearly perfect AU-ROC of 0.995 was calculated (Figure 13). This revealed the excellent accuracy of our programs in identifying active ER antagonists among diverse structures.

At this stage, if 1000 compounds of this final library of 25 000 compounds were tested for biological activity, at least 10 known active compounds would have been identified. The active entities consisted of 0.029% of the initial library (40 out of 138,975 compounds) and 0.04% of the more diverse library (10 out of 25 000 compounds). After the protocol, 1% of the

1000 compounds selected were known actives, yielding enrichments as high as 34.5 and 25, respectively. In this scenario, 25% of the 40 actives would have been found by testing only 0.4% of the library.

**Application to the Lead Optimization.** As soon as hits are identified, structure–activity relationship studies are traditionally carried out. In this context, one can search for analogues of these hits prior to synthesizing new entities. In the previous stage, we identified 10 hits following the described scenario. A second workflow was designed and is illustrated in Figure 12b. This workflow relies on the “SEARCH FOR ANALOGUES” action to look for analogues of these 10 uncovered hits. The search was performed on the initial NCI database seeded with 40 actives. It is clear that, in a standard chemical library, the number of actives may be lower than 40 and these actives would be weaker.

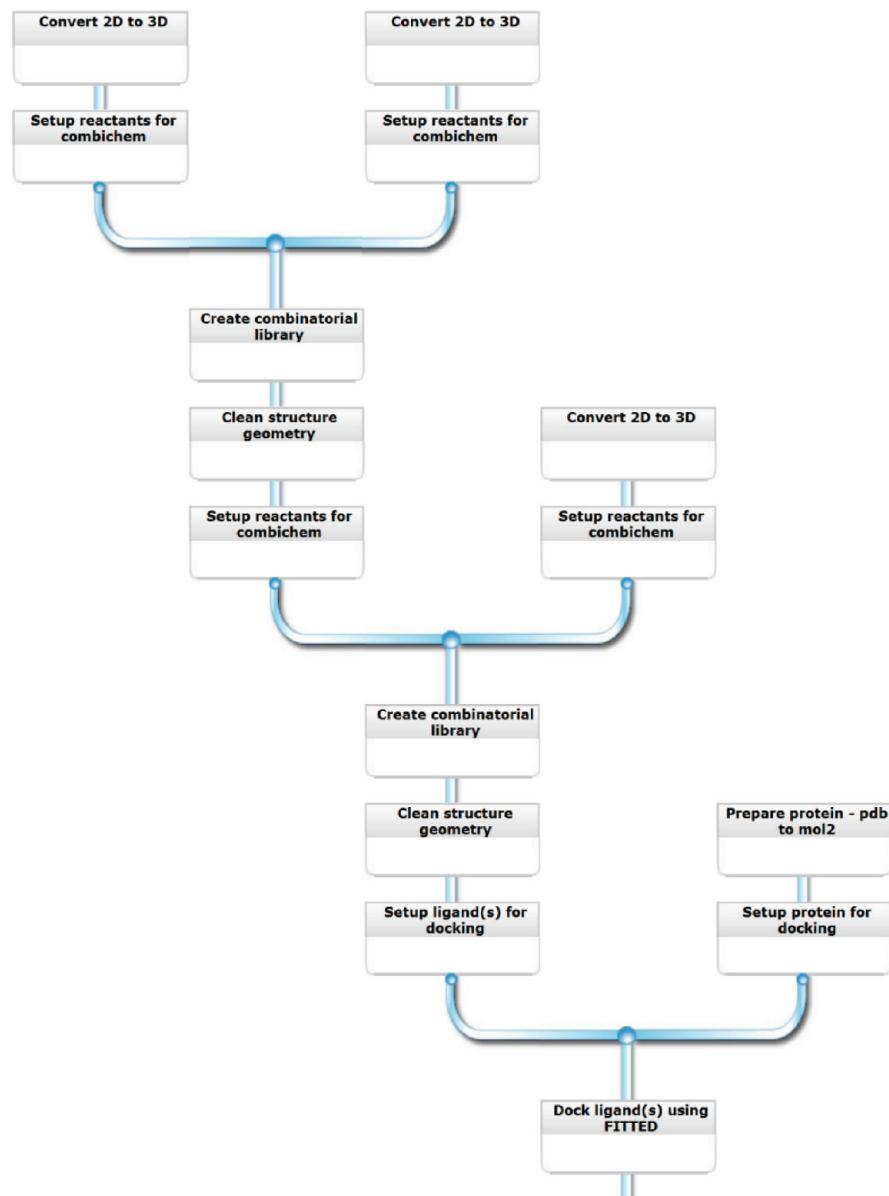
However, one can apply this search to much larger libraries (millions of compounds), as this workflow requires the docking of a few hundreds of potential ligands.

**Table 1. Known Antagonists**

	X	scaffold	Y	ID (bitstring)
acolbifene	X1	S2	Y2	1 2 2
raloxifene	X1	S3	Y2	1 3 2
arzoxifene	X1	S5	Y2	1 5 2
pipendoxifene	X1	S13	Y2	1 13 2
bazedoxifene	X1	S13	Y13	1 13 13
4-OH-tamoxifen	X2	S1	Y4	2 1 4
4-OH-CB5638	X2	S1	Y14	2 1 14
lasofoxifene	X2	S7	Y1	2 7 1
SP500263	X2	S12	Y2	2 12 2
quinoline GSK	X2	S14	Y14	2 14 14

When the 10 hits were used, application of this action with a Tanimoto coefficient of 0.60 identified all of the 40 known actives (including the 10 used as input) together with another 169 structures (Figure 12b). These 209 structures were then docked with FITTED, scored, and ranked by scores. Thirty-eight of the top 40 are known actives (including the previously identified 10 hits). In a real case scenario, the 50 top ranked compounds (excluding the previous 10 hits which have been tested previously) could have been selected for testing. If this was the case, the 50 tested compounds would include 29 of the remaining 30 seeded active compounds. In fact, when the results from the first screening and this search for analogues are combined, 1050 compounds would have been biologically assayed and (at least) 39 would be active compounds for an enrichment factor of 129 compared to the original library.

**Application to the Design of a Combinatorial Library of Potential SERMs.** When hits are found, one can also design a library of analogues and test them for their potency (Figure 14).



**Figure 15.** Workflow used to create and dock a combinatorial library made of three building blocks.

To simulate this approach, we have designed a library of potential SERMs made of the building blocks shown in Chart 1. A library of scaffolds (**S1–S14**) and two libraries of side chains (**X1–X14** and **Y1–Y16**) were drawn and saved as sdf files. As with the workflows shown above, these libraries were designed to include known binders (Table 1). A workflow was built (Figure 15) which creates the combinatorial library of 3136 compounds through two chemical transformations and prepares it for docking. In the extreme right branch of the workflow, a PDB file is given and prepared for docking.

The entire protocol was carried out on a workstation using a single core. The final result file was read, and a spreadsheet table of bitstrings and scores was created to analyze and rank all of the compounds by docking scores (Figure 16). Nevertheless, not only did the entire protocol proceed until the end with no need for interactions with the user, but also the AU-ROC of 0.78 indicated a significantly discriminative process. It is worth mentioning that when considering the structures made through this combinatorial process, many of them may actually be active. Thus, this discrimination, in contrast to the hit discovery discussed above, is between known strong actives and other compounds that are most likely at least weakly active.

**Combinatorial Synthesis Based on a Synthetic Scheme.** To further test the ability of this set of programs to discriminate between weak and strong binders, the same workflow was used to simulate the optimization of a lead compound. In 1997, Grese et al. reported an exhaustive structure activity relationship study around raloxifene.<sup>60</sup> The synthesis, illustrated in Figure 17, starts from a library of compounds with the general structure **1** which is subjected to Friedel-Crafts acylation with **2** to afford the corresponding compounds with the general structure **3**. The addition of Grignard reagents (**4**) to compounds **3** provided

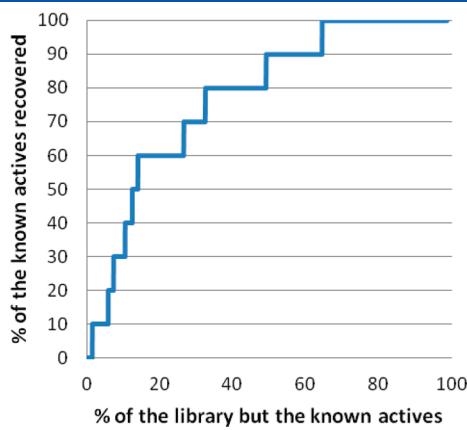


Figure 16. Receiving operator characteristic curve for the docking of a combinatorial library of 3136 compounds including 10 known actives.

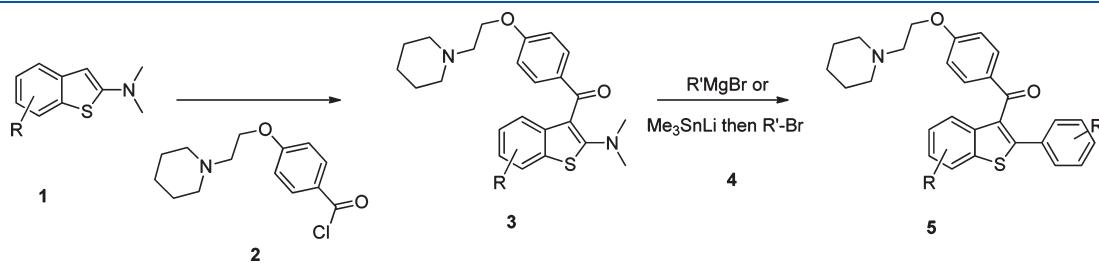


Figure 17. Synthesis of raloxifene analogues.

analogues with the structure **5**. Alternatively, **5** could be obtained by the conversion of **3** into the corresponding stanny derivatives and Stille coupling with a variety of alkyl or aryl bromide of type **4**.

This synthetic scheme was used with the same workflow as above (Figure 15) and with these two chemical reactions given to the platform. We have drawn a library of 23 compounds with structure **1** and 49 compounds with structure **4**. This led to a library of 1127 compounds with biological activity data (competitive binding and cell-based assays) available for 77 of these.<sup>60</sup> Accurately predicting the ranking of these 77 compounds is very challenging as most of them exhibit similar activity, and the discrimination between active molecules remains a challenge for state-of-the-art scoring functions.<sup>42</sup> Once more, the workflow ran uneventfully, and the predicted ranking of these 77 structures (identified by their bitstring) was compared to the observed rankings. Expectedly, the accuracy in this lead optimization study was lower than that observed in the previous study with a correlation as low as  $r^2 = 0.16$  (Figure 18).

## CONCLUSIONS

Aiming to integrate medicinal chemistry and computational molecular design, we have developed FORECASTER, a set of tools for the virtual screening of small molecules binding to biomacromolecules (proteins, receptors, and nucleic acids), based on a flexible-target docking program, FITTED. These tools can prepare the protein and ligand files for docking (PREPARE, PROCESS, CONVERT, and SMART), as well as generate combinatorial libraries of ligands (REACT) and extract diverse or focused libraries from large molecular libraries (REDUCE, SELECT). All of these tools have

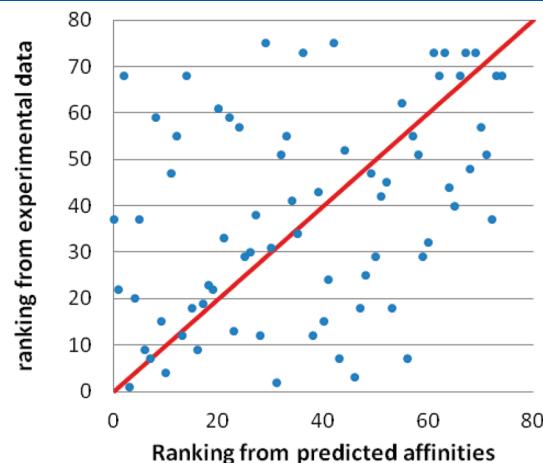


Figure 18. Observed vs predicted rankings for 77 known raloxifene analogues.

been combined into a platform with a Web-based user interface, which removes the requirement of editing text files and dealing with command-line applications that keep many potential users away from these tools. The use of REACT, which builds combinatorial libraries from user-defined chemical schemes, requires expertise in organic chemistry, and a sketcher was implemented into a rules generator. A chemist can draw a chemical transformation and automatically generate the rules, which can in turn be used to prepare combinatorial libraries.

Application of this platform to drug discovery scenarios and more specifically to the discovery of SERMs demonstrated its usability and its accuracy. A first application yielded libraries with a very high enrichment rate of potential antagonists, shown by the high recovery of seeded actives. This first result validated the platform as a predictive and user-friendly tool for the discovery of new therapeutic entities. More “hit-to-lead” scenarios including combinatorial synthesis of analogues were also tested and put forward known active structures. Virtual screening on much larger libraries, design, and biological evaluation of potential SERMs can now start. This platform is available for free for academic use upon request at <http://fitted.ca>.

## ■ EXPERIMENTAL SECTION

**Development of Programs.** The interface to FORECASTER was written in Ruby-On-Rails and tested on Firefox (version 3 and higher), Safari (version 5 and higher), and Internet Explorer (version 7 and higher), while all of the programs (SMART, SELECT, REACT, PREPARE, PROCESS, CONVERT, REDUCE, and FITTED) were written in C++. In order to facilitate the integration of functions, all of these programs but FITTED were next combined into a single piece of software. All of these programs are available upon request at <http://fitted.ca> and are available for Windows, Mac OSX, and Linux operating systems. The necessary ChemWriter applet requires a separate license.

**Application.** The workflows were created using the Web interface FORECASTER and ran on an Intel Core 2 Quad CPU Q6700. The data (text output files) were analyzed using Excel datasheets, while the 2D structures were drawn using either ChemDraw or ChemWriter implemented in FORECASTER. The default parameters were used with all of the programs. Detailed keyword files are given as Supporting Information.

## ■ ASSOCIATED CONTENT

**S Supporting Information.** Example of reaction rules and the 40 active structures in mol2 format are given. This material is available free of charge via the Internet at <http://pubs.acs.org>.

## ■ AUTHOR INFORMATION

### Corresponding Author

\*E-mail: [nicolas.moitessier@mcgill.ca](mailto:nicolas.moitessier@mcgill.ca).

## ■ ACKNOWLEDGMENT

We thank “Le Ministère du Développement Économique, de l’Innovation et de l’Exportation du Québec” through the program “Soutien à la maturisation technologique”, AstraZeneca R&D Montreal, NSERC, and CIHR for funding. While performing the work, P.E. was supported by a scholarship from CIHR (Strategic Training Initiative in Chemical Biology) and a J. W.

McConnell Memorial Fellowship from McGill University. V.C.-S. acknowledges the Drug Development Training Program (DDTP—CIHR) for a fellowship. N.W. thanks Le Groupe de Recherche Axé sur la Structure des Protéines (GRASP—FRSQ). R.M.S. is supported by a scholarship from CIHR (Strategic Training Initiative in Drug Discovery and Development) and a CONACyT (Consejo Nacional de Ciencia y Tecnología) graduate studies scholarship. N.M. thanks the Faculty of Science for the Fessenden Professorship in Innovation. We acknowledge generous allocations of computer time from Calcul Québec.

## ■ REFERENCES

- (1) Couse, J. F.; Korach, K. S. Estrogen receptor null mice: what have we learned and where will they lead us? *Endocrine Rev.* **1999**, *20*, 359.
- (2) Sanchez, R.; Nguyen, D.; Rocha, W.; White, J. H.; Mader, S. Diversity in the mechanisms of gene regulation by estrogen receptors. *Bioessays* **2002**, *24*, 244.
- (3) Kos, M.; Reid, G.; Denger, S.; Gannon, F. Minireview: genomic organization of the human ERalpha gene promoter region. *Mol. Endocrinol.* **2001**, *15*, 2057.
- (4) Shaw, J. A.; Udukang, K.; Mosquera, J. M.; Chauhan, H.; Jones, J.; Walker, R. A. Oestrogen receptors alpha and beta differ in normal human breast and breast carcinomas. *J. Pathol.* **2002**, *198*, 450.
- (5) Miller, C. P. SERMs: evolutionary chemistry, revolutionary biology. *Curr. Pharm. Des.* **2002**, *8*, 2089.
- (6) O'Regan, R. M.; Jordan, V. C. Tamoxifen to raloxifene and beyond. *Semin. Oncol.* **2001**, *260*.
- (7) Pickar, J. H.; MacNeil, T.; Ohleth, K. SERMs: Progress and future perspectives. *Maturitas* **2010**, *67*, 129.
- (8) Jordan, C. J. Antiestrogens and Selective Estrogen Receptor Modulators as Multifunctional Medicines. 1. Receptor Interactions. *J. Med. Chem.* **2003**, *46*, 1081.
- (9) Wang, T.; You, Q.; Huang, F. S.; Xiang, H. Recent advances in selective estrogen receptor modulators for breast cancer. *Mini. Rev. Med. Chem.* **2009**, *9*, 1191.
- (10) Riggs, B. L.; Hartmann, L. C. Selective Estrogen-Receptor Modulators - Mechanisms of Action and Application to Clinical Practice. *New Engl. J. Med.* **2003**, *348*, 618.
- (11) Bohacek, R. S.; McMullan, C.; Guida, W. C. The art and practice of structure-based drug design: a molecular modelling perspective. *Med. Res. Rev.* **1996**, *16*, 3.
- (12) Koppen, H. Virtual screening - what does it give us? *Curr. Opin. Drug Discovery Dev.* **2009**, *12*, 397.
- (13) Tuccinardi, T. Docking-based virtual screening: Recent developments. *Comb. Chem. High Throughput Screening* **2009**, *12*, 303.
- (14) Ferreira, R. S.; Simeonov, A.; JadHAV, A.; Eidam, O.; Mott, B. T.; Keiser, M. J.; McKerrow, J. H.; Maloney, D. J.; Irwin, J. J.; Shoichet, B. K. Complementarity between a docking and a high-throughput screen in discovering new cruzain inhibitors. *J. Med. Chem.* **2010**, *53*, 4891.
- (15) Shoichet, B. K. Virtual screening of chemical libraries. *Nature* **2004**, *432*.
- (16) Schulz, M. N.; Hubbard, R. E. Recent progress in fragment-based lead discovery. *Curr. Opin. Pharmacol.* **2009**, *9*, 615.
- (17) Loving, K.; Alberts, I.; Sherman, W. Computational approaches for fragment-based and de novo design. *Curr. Top. Med. Chem.* **2010**, *10*, 14.
- (18) Keseru, G. M.; Makara, G. M. Hit discovery and hit-to-lead approaches. *Drug Discovery Today* **2006**, *11*, 741.
- (19) Gillet, V.; Myatt, G.; Zsoldos, Z.; Johnson, A. SPROUT, HIPPO and CAESA: Tools for de novo structure generation and estimation of synthetic accessibility. *Perspect. Drug Discovery Des.* **1995**, *3*, 34.
- (20) Gasteiger, J. De novo design and synthetic accessibility. *J. Comput.-Aided Mol. Des.* **2007**, *21*, 307.
- (21) Corbeil, C. R.; Thielges, S.; Schwartzentruber, J. A.; Moitessier, N. Toward a computational tool predicting the stereochemical outcome of asymmetric reactions: Development and application of a rapid and

- accurate program based on organic principles. *Angew. Chem., Int. Ed.* **2008**, *47*, 2635.
- (22) Weill, N.; Corbeil, C. R.; De Schutter, J. W.; Moitessier, N. Toward a computational tool predicting the stereochemical outcome of asymmetric reactions: Development of the molecular mechanics-based program ACE and application to asymmetric epoxidation reactions. *J. Comput. Chem.* **2011**, *32*, 2878.
- (23) Corbeil, C. R.; Englebienne, P.; Moitessier, N. Docking ligands into flexible and solvated macromolecules. 1. Development and validation of FITTED 1.0. *J. Chem. Inf. Model.* **2007**, *47*, 435.
- (24) Corbeil, C. R.; Moitessier, N. Docking ligands into flexible and solvated macromolecules. 3. Impact of input ligand conformation, protein flexibility, and water molecules on the accuracy of docking programs. *J. Chem. Inf. Model.* **2009**, *49*, 997.
- (25) Pipeline Pilot, 8.0 ed.; SciTegic Inc: San Diego, CA, 2009.
- (26) Berthold, M. R.; Cebron, N.; Dill, F.; Gabriel, T. R.; Kotter, T.; Meini, T.; Ohl, P.; Sieb, C.; Thiel, K.; Wiswedel, B. In *Studies in Classification, Data Analysis, and Knowledge Organization (GfKL 2007)*, 8.0 ed.; Springer: New York, 2007.
- (27) MOE; Chemical Computing Group: Montreal, Quebec, Canada, 2009.
- (28) Maestro, 8.0 ed.; Schrödinger, LLC: Portland, OR, 2007.
- (29) ChemDraw Ultra, 12 ed.; CambridgeSoft: Cambridge, MA, 2010.
- (30) Sigma-Aldrich. www.sigma-aldrich.com (accessed Dec. 2011).
- (31) Corbeil, C. R.; Englebienne, P.; Yannopoulos, C. G.; Chan, L.; Das, S. K.; Bilimoria, D.; L'Heureux, L.; Moitessier, N. Docking ligands into flexible and solvated macromolecules. 2. Development and application of FITTED 1.5 to the virtual screening of potential HCV polymerase inhibitors. *J. Chem. Inf. Model.* **2008**, *48*, 902.
- (32) Halgren, T. A. Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94. *J. Comput. Chem.* **1996**, *17*, 490.
- (33) Oda, A.; Hirono, S. Geometry-dependent atomic charge calculations using charge equilibration method with empirical two-center Coulombic terms. *THEOCHEM* **2003**, *634*, 159.
- (34) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general Amber force field. *J. Comput. Chem.* **2004**, *25*, 1157.
- (35) Holliday, G. L.; Murray-Rust, P.; Rzepa, H. S. Chemical Markup, XML, and the World Wide Web. 6. CMLReact, an XML Vocabulary for Chemical Reactions. *J. Chem. Inf. Model.* **2005**, *46*, 145.
- (36) Reactor; ChemAxon Ltd: Budapest, Hungary. Contact ChemAxon for pricing information: www.chemaxon.com (accessed Dec. 2011). Bode, J. W. *J. Am. Chem. Soc.* **2004**, *126*, 15317.
- (37) Durant, J. L.; Leland, B. A.; Henry, D. R.; Nourse, J. G. Reoptimization of MDL keys for use in drug discovery. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 1273.
- (38) Labute, P. Protonate 3D: Assignment of Macromolecular Protonation State and Geometry. <http://www.chemcomp.com/journal/proton.htm> (accessed Dec. 2011).
- (39) Schrodinger, Protein Preparation Wizard 2011. <http://www.schrodinger.com/supportdocs/18/16/> (accessed Dec. 2011).
- (40) Lovell, S. C.; Word, J. M.; Richardson, J. S.; Richardson, D. C. The penultimate rotamer library. *Proteins: Struct., Funct., Genet.* **2000**, *40*, 389.
- (41) Moitessier, N.; Therrien, E.; Hanessian, S. A method for induced-fit docking, scoring, and ranking of flexible ligands. Application to peptidic and pseudopeptidic  $\beta$ -secretase (BACE 1) inhibitors. *J. Med. Chem.* **2006**, *49*, 5885.
- (42) Englebienne, P.; Moitessier, N. Docking ligands into flexible and solvated macromolecules. 5. Force-field-based prediction of binding affinities of ligands to proteins. *J. Chem. Inf. Model.* **2009**, *49*, 2564.
- (43) Kieltyka, R.; Englebienne, P.; Moitessier, N.; Sleiman, H. Quantifying interactions between G-quadruplex DNA and transition-metal complexes. *Methods Mol. Biol. (Clifton, N.J.)* **2010**, *608*, 223.
- (44) Yang, W. H.; Wang, Z. Y.; Liu, H. L.; Yu, H. X. Exploring the binding features of polybrominated diphenyl ethers as estrogen receptor antagonists: Docking studies. *SAR QSAR Environ. Res.* **2010**, *21*, 351.
- (45) Liu, H.; Papa, E.; Gramatica, P. Evaluation and QSAR modeling on multiple endpoints of estrogen activity based on different bioassays. *Chemosphere* **2008**, *70*, 1889.
- (46) Celik, L.; Lund, J. D. D.; Schiøtt, B. Exploring interactions of endocrine-disrupting compounds with different conformations of the human estrogen receptor  $\alpha$  ligand binding domain: A molecular docking study. *Chem. Res. Toxicol.* **2008**, *21*, 2195.
- (47) Koh, M.; Park, S. B. Computer-aided design and synthesis of tetra-aryl-substituted alkenes and their bioevaluation as a selective modulator of estrogen-related receptor  $\gamma$ . *Mol. Divers.* **2010**, *1*.
- (48) Singh, U. S.; Shankar, R.; Yadav, G. P.; Kharkwal, G.; Dwivedi, A.; Keshri, G.; Singh, M. M.; Moulik, P. R.; Hajela, K. Synthesis and structure guided evaluation of estrogen agonist and antagonist activities of some new tetrazolyl indole derivatives. *Eur. J. Med. Chem.* **2008**, *43*, 2149.
- (49) Dong, X.; Hilliard, S. G.; Zheng, W. Structure-based quantitative structure-activity relationship modeling of estrogen receptor  $\beta$ -ligands. *Fut. Med. Chem.* **2011**, *3*, 933.
- (50) Brzozowski, A. M.; Pike, A. C. W.; Dauter, Z.; Hubbard, R. E.; Bonn, T.; Engström, O.; Öhman, L.; Greene, G. L.; Gustafsson, J. A.; Carlquist, M. Molecular basis of agonism and antagonism in the oestrogen receptor. *Nature* **1997**, *389*, 753.
- (51) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235.
- (52) Knox, A. J. S.; Yang, Y.; Lloyd, D. G.; Meegan, M. J. Virtual screening of the estrogen receptor. *Expert Opin. Drug Discovery* **2008**, *3*, 853.
- (53) Baxter, C. A.; Murray, C. W.; Waszkowycz, B.; Li, J.; Sykes, R. A.; Bone, R. G. A.; Perkins, T. D. J.; Wyllie, W. New approach to molecular docking and its application to virtual screening of chemical databases. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 254.
- (54) Bissantz, C.; Folkers, G.; Rognan, D. Protein-based virtual screening of chemical databases. 1. Evaluation of different docking/scoring combinations. *J. Med. Chem.* **2000**, *43*, 4759.
- (55) Waszkowycz, B.; Perkins, T. D. J.; Sykes, R. A.; Li, J. Large-scale virtual screening for discovering leads in the postgenomic era. *IBM Systems J.* **2001**, *40*, 360.
- (56) ChemWriter, 1.0 ed.; Metamolecular: La Jolla, CA, 2010.
- (57) Irwin, J. J.; Shoichet, B. K. ZINC - A Free Database of Commercially Available Compounds for Virtual Screening. *J. Chem. Inf. Model.* **2005**, *45*, 177.
- (58) Nicholls, A. What do we know and when do we know it? *J. Comput.-Aided Mol. Des.* **2008**, *22*, 239.
- (59) Hawkins, P. C. D.; Warren, G. L.; Skillman, A. G.; Nicholls, A. How to do an evaluation: Pitfalls and traps. *J. Comput.-Aided Mol. Des.* **2008**, *22*, 179.
- (60) Giese, T. A.; Cho, S.; Finley, D. R.; Godfrey, A. G.; Jones, C. D.; Lugar, C. W.; Martin, M. J.; Matsumoto, K.; Pennington, L. D.; Winter, M. A.; Adrian, M. D.; Cole, H. W.; Magee, D. E.; Phillips, D. L.; Rowley, E. R.; Short, L. L.; Glasebrook, A. L.; Bryant, H. U. Structure-Activity Relationships of Selective Estrogen Receptor Modulators: Modifications to the 2-Arylbenzothiophene Core of Raloxifene. *J. Med. Chem.* **1997**, *40*, 146.