Published in final edited form as:

Eur J Med Chem. 2011 May ; 46(5): 1512–1523. doi:10.1016/j.ejmech.2011.01.069.

Discovery of novel SERCA inhibitors by virtual screening of a large compound library

Christopher Elam^a, Michael Lape^a, Joel Deye^a, Jodie Zultowsky^a, David T. Stanton^b, and Stefan Paula^{a,*}

^aDepartment of Chemistry, Northern Kentucky University, Highland Heights, KY 41099, U.S.A.

^bMiami Valley Innovation Center, Procter & Gamble, Cincinnati, OH 45252, U.S.A.

Abstract

Two screening protocols based on recursive partitioning and computational ligand docking methodologies, respectively, were employed for virtual screens of a compound library with 345,000 entries for novel inhibitors of the enzyme sarco/endoplasmic reticulum calcium ATPase (SERCA), a potential target for cancer chemotherapy. A total of 72 compounds that were predicted to be potential inhibitors of SERCA were tested in bioassays and 17 displayed inhibitory potencies at concentrations below 100 μ M. The majority of these inhibitors were composed of two phenyl rings tethered to each other by a short link of one to three atoms. Putative interactions between SERCA and the inhibitors were identified by inspection of docking-predicted poses and some of the structural features required for effective SERCA inhibition were determined by analysis of the classification pattern employed by the recursive partitioning models.

Keywords

hydroquinones; virtual screening; recursive partitioning; ligand docking; calcium pump; enzyme inhibition; structure-activity relationship

1. Introduction

Due to their significant value as research tools and their medicinal potential as an emerging new class of anti-prostate cancer agents, the discovery and design of novel inhibitors of sarco/endoplasmic reticulum calcium ATPase (SERCA¹) have become areas of considerable interest in the recent past [1–6]. Among several structurally unrelated classes of inhibitors, the natural product thapsigargin (TG) and some of its derivatives have found the most widespread use because of their high potencies and specificity for SERCA [7–10]. Since its total synthesis from relatively simple starting materials requires many steps and thus suffers

© 2011 Elsevier Masson SAS. All rights reserved.

*corresponding author: Natural Sciences Center, Nunn Drive, Northern Kentucky University, Highland Heights, KY 41099-1905, U.S.A., Tel.: 859-572-6552, FAX: 859-572-5162, paulas1@nku.edu.

¹Abbreviations: SERCA: sarco/endoplasmic reticulum calcium ATPase; BHQ: di-*tert*-butylhydroquinone; SAR: structure-activity relationship; VS: virtual screening; RP: recursive partitioning; QSAR: quantitative structure-activity relationship; GRI: Genome Research Institute; GOLD: genetic optimization for ligand docking; RMSD: root mean square deviation; NADH: nicotinamide adenine dinucleotide; ATP: adenosine triphosphate; PK: pyruvate kinase; LDH: lactate dehydrogenase; MD: molecular dynamics

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

from low yields [11,12], TG is a rather expensive agent whose main source remains the Mediterranean plant *Thapsia garganica* from which it can be extracted. Therefore, searches for alternative SERCA inhibitors are ongoing and, so far, they have resulted in the discovery of a sizeable repertoire of inhibitors with good potencies. Examples include the fungal metabolite cyclopiazonic acid [13–16], terpenolides [17], the antifungal drug clotrimazole [18–20], derivatives of thiuronium benzene [21–24], the flame retardant tetrabromobisphenol [25,26], curcumin [27,28], and di-1,5-*tert*-butylhydroquinone (BHQ) [29–31]. Despite targeting the same enzyme, SERCA, these inhibitors display a considerable variability in chemical structure, inhibitory potency, and physical properties. Compared to most inhibitors, BHQ and some of its analogs are of particular interest because of their structural simplicity, the associated low cost, and their efficacy and potency towards SERCA [1,29,32].

Since their discovery as effective SERCA inhibitors, several structure-activity relationship (SAR) studies of BHQ analogs have been conducted that elucidated the structural features responsible for inhibitory potency [1,6,29,32]. In addition, high resolution X-ray crystal structures of BHQ in complex with SERCA have become available [33,34]. These have provided a detailed characterization of the inhibitor binding site at the molecular level and identified the amino acids engaged in interactions with BHQ. In essence, these studies showed that BHQ binding is mediated by two hydrogen bonds between the two BHQ hydroxyl groups and Asp59 and Pro308 as well as by hydrophobic interactions between the non-polar *tert*-butyl groups of the inhibitor and several non-polar amino acid residues lining the binding pocket.

As for any enzyme inhibitor target, virtual screening (VS) methods can be employed in the search for new inhibitors of SERCA. VS is a computational technique for the rapid prediction of a particular properties of compounds contained in large libraries [35,36]. Since it is faster and significantly less expensive than experimental high throughput screening, VS nowadays plays a central role in many drug discovery programs. If the 3D structure of the target receptor is known from crystallography or NMR studies, structure-based VS *via* ligand docking is often the method of choice. Docking routines predict the binding pose of a ligand in the receptor binding site and compute the binding affinity using scoring functions [37]. In the absence of a 3D receptor structure, ligand-based VS methods such as quantitative structure-activity relationship (QSAR) modeling or pharmacophore development can establish models capable of predicting bioactivities [38–40]. Unlike structure-based VS, ligand-based VS requires activity data for a sufficiently large set (often 30 or more) of structurally related training compounds. Whereas the applicability of ligand-based VS is often limited to molecules that bear some structural resemblance to those in the training set, its advantage is its high speed of execution that allows the search of sizeable libraries in a matter of hours. Examples for the successful application of structure-based VS include the *in silico* identification of epidermal growth factor receptor inhibitors with anti-proliferative activity against cancer cells [41], the search for small-molecule inhibitors of the SARS virus [42], and the discovery of human xylulose reductase inhibitors for the treatment of complications from diabetes [43]. Ligand-based VS methodologies have been instrumental in the discovery of carbonic anhydrase [44] and renin inhibitors [45] as well as in the search for inhibitors of the vascular endothelial growth factor receptor kinase [45].

In an effort to expand the current repertoire of hydroquinone-based SERCA inhibitors, we recently developed a VS protocol and applied it to the “Cactus” compound collection of 260,000 entries maintained by the National Cancer Institute [6]. The protocol started with a similarity search that reduced the number of compounds to those that were structurally related to the parent compound BHQ. Those were then computationally docked into the BHQ-binding site of SERCA and rank-ordered according to their docking scores. The

effectiveness of the protocol was assessed in subsequent bioassays of the top-ranked compounds that led to the discovery of 19 novel inhibitors, all of which inhibited the enzyme at concentrations below 50 μ M. Motivated by the quite favorable hit rate of this particular screening method (33%), we sought to apply it to other compound collections as well. Simultaneously, we explored alternative VS protocols that involved recursive partitioning (RP) and that are not dependent on structure-based design methodologies.

Among the various VS methodologies that have been employed for drug discovery in the past, RP is a relatively new approach. Generally speaking, RP is a statistical method that establishes selection rules to classify objects with similar properties into groups. RP has found widespread use in medical diagnostic tests, but it is also suitable for screening purposes in drug discovery [46,47]. In the latter case, library compounds are the objects which are grouped into classes with comparable bioactivities and chemical structures, which are expressed numerically in the form of classical chemical descriptors. Unlike docking, RP does not require knowledge of the 3D structure of the binding site, but needs a reasonably large set of training compounds with known potencies for the establishment of selection rules. Once the latter are defined, the contents of much larger compound collections can be classified in a straightforward and rapid manner. In fact, the speed of its execution is believed to be a major strength of RP compared to some other methods. Because of their intuitive nature, RP-generated classification trees can also assist the development and interpretation of SARs. Moreover, RP has the distinct advantage of incorporating information on inactive compounds into its selection rules, a feature that is rather difficult to realize in traditional QSAR modeling. Whereas QSAR-, docking-, or pharmacophore-based methods are well-established approaches in drug discovery, RP as a relatively new method has only been employed in a limited number of cases. Examples for the successful application of RP techniques include the analysis and categorization of monoamine oxidase inhibitors, potassium channel blockers, and CYP450 inhibitors [48–50] as well as the discrimination between water soluble and insoluble compounds [51].

In this study, we screened a large compound collection of 345,000 compounds employing both our previously described similarity search/docking protocol as well as a newly developed search procedure that incorporated a classification by RP. The compound library that was the target of this study had been originally assembled by Procter & Gamble Pharmaceuticals and is currently maintained by the Genome Research Institute (GRI) at the University of Cincinnati. This collection includes a significant number of compounds selected based on chemical diversity to span “drug-space” as well as several directed libraries of compounds that contains inhibitors of specific biological targets such as kinases. After the completion of the screens, we selected between 30 and 40 compounds identified as potential SERCA inhibitors by each of the two protocols and measured their potencies in bioassays. The most active inhibitors that had been selected by the docking-based protocol were analyzed by close examination of their docking-predicted binding poses to find crucial ligand/receptor interactions. Likewise, structural features critical for activity of the compounds chosen by the RP-based protocol were determined by analysis of molecular descriptors used in the RP classification schemes.

2. Results

2.1. Development and application of an RP-based selection model

As outlined in Fig. 1, the RP-based VS protocol – from now on referred to as Protocol 1 – consisted of three steps that involved a physical property filter, an RP consensus selection step, and a substructure search. Although the term “RP-based” might suggest otherwise, it should be noted that each one of the three steps of the protocol served a unique purpose and was therefore considered to be of equal importance for the performance of the screen. The

first step was implemented to eliminate those compounds whose physical properties would likely prevent them from being effective SERCA inhibitors. For this purpose, the octanol-water partition coefficient (log P), the intrinsic aqueous solubility (log WS), and the molecular weight (MW) of 88 previously characterized BHQ-related compounds [1,6,32] (53 “actives” and 35 “inactives”, see Supplementary Materials for structures and activities) were calculated. Statistical analysis of these three parameters (Table 1) revealed that active compounds were in general somewhat more hydrophobic (higher log P values) than inactive ones and that they also were slightly less soluble in water (lower log WS values). Thus, only compounds in the library with log P values between 1.5 and 4.5 and log WS values between -4.0 and -1.5 were considered. In addition, a molecular weight filter ranging from 150 – 350 g/mol was applied to account for the finite size of the inhibitor binding pocket of SERCA. Taken together, the three selection criteria in the filtering step reduced the number of compounds under consideration from 345,000 to 40,577.

For the development of RP models capable of classifying compounds according to their inhibitory potencies against SERCA, the 88 compounds mentioned above were used as a training set. They were grouped into bins according to their potencies using two types of binning – binary binning, in which compounds were simply classified as “active” or “inactive”, and range binning, in which each of the seven bins covered 0.5 log units on the activity scale starting at a value of 0.5 (including one bin for “inactives”, Table 2). Next, conformation-independent descriptors were calculated with the program winMolconn based on the modeled molecular structures of the compounds and RP models were developed for both types of binning (Model 1 for binary binning and Model 2 for range binning) using the statistical software JMP. In order to enhance the activity separation of Model 2, an additional model with six bins (Model 3) was created using only compounds previously classified as “actives” by Model 1 (see Supplementary Materials for selection rules of the three models). Since some terminal “leaves” of the obtained RP classification trees were mixed and contained compounds from more than one activity bin, the activity of such a mixed leaf was calculated as the average of the bin numbers of the compounds assigned to it and then rounded to the closest integer. The quality of Model 3 is illustrated in Fig. 2, which shows the correlation between experimentally observed and model-predicted bin allocations ($r^2 = 0.75$). Most of the compounds were correctly assigned and were therefore found on the diagonal of the graph. Only a few inhibitors were placed in incorrect bins by the model, but even in these cases, most of the model-predicted bin assignments differed from the experimental one by no more than one bin and, more importantly, no outliers were encountered.

Since the main focus of the screen was the discovery of novel SERCA inhibitors with the highest possible potencies for subsequent testing, we attempted to minimize the number of false positives in the test candidate pool by implementing a consensus selection procedure that included all three RP models. In order to be in the consensus pool, a compound had to be classified as “active” by Model 1 while simultaneously belonging to one of the four most active bins of Models 2 and 3 (Fig. 3). This step reduced the number of compounds under consideration from 40,577 to 8,503 (Fig. 1).

The purpose of the last step of Protocol 1 was the elimination of compounds that passed the previous selection criteria but had little structural similarity to the known SERCA inhibitors in the training set. Since the vast majority of actives in the training set displayed a phenol moiety, a phenol substructure search was implemented, which further reduced the number of compounds to a final total of 429. After considering the activity predictions for these compounds according to Model 3, a selection of 39 compounds were requested from GRI that included the 8 compounds in the highest occupied activity bin (Bin 2 since Bin 1 was

empty) and 31 randomly chosen ones from the second highest occupied activity bin. All 39 requested compounds were received for experimental testing in inhibition assays.

2.2. The docking-based screen

Our previously developed docking-based screening protocol [6] – designated Protocol 2 – and its individual steps are outlined in the flow chart of Fig. 1. The similarity search implemented in the first step served to eliminate compounds with little structural similarity to the parent compound BHQ. This was accomplished by using a Tanimoto similarity index cut-off value of 65% and reduced the number of compounds for subsequent docking to a manageable size (419) that did not require extensive computation times. The structures of these compounds were then docked into the BHQ binding site of SERCA with the program GOLD (scoring function: ChemScore) as described previously [6,32]. GOLD is a commonly used docking routine that operates with a genetic search algorithm to flexibly dock ligands into a predefined binding site and evaluates the fitness of poses using a variety of scoring functions [52,53]. To ensure reproducibility, only those runs producing a consensus were considered further (303 compounds). By definition, a consensus was obtained if at least half of 20 independent docking repeats conducted under identical conditions yielded the same result (within 2 Å RMSD). Among the 303 compounds with a consensus, there were 37 molecules whose highest-ranking consensus pose scored higher than that of BHQ. Samples of these compounds were requested from GRI and 33 were received for experimental testing.

2.3. Experimentally determined inhibitory potencies

The inhibitory potencies of compounds were determined as described previously using a spectroscopic assay that measured the rate of NADH oxidation coupled to SERCA-catalyzed ATP hydrolysis [1,32]. To ensure that apparent SERCA inhibition was not caused by interference with the auxiliary enzymes pyruvate kinase (PK) or lactate dehydrogenase (LDH), compounds testing positive in the coupled assay were also subjected to a phosphate liberation assay. Even though it was more time-consuming and required more sample materials, the latter assay did not rely on the use of PK and LDH and detected directly the liberation of inorganic phosphate from ATP by spectroscopically monitoring its reaction with the dye malachite green [54]. All compounds testing positive in the coupled assay also tested positive in the phosphate liberation assay, indicating that there were no false positives caused by the inhibition of PK or LDH.

The results summarized in Table 3 show that 17 among the 72 compounds tested had IC₅₀ values below 100 μM. The majority of them (12) had been selected by Protocol 2 whereas 5 had been identified by Protocol 1. The identities and structures of inactive molecules are listed in the Supplementary Materials section.

3. Discussion

3.1. SERCA/inhibitor interactions as predicted by docking

In order to focus on the more potent SERCA inhibitors, the following discussion will be restricted to compounds with IC₅₀ values below 50 μM (“potent inhibitors” in Fig. 4). Inspection of their chemical structures shows that the 12 inhibitors discovered by Protocol 2 belong to two general classes: derivatives of bisphenol and remote analogs of BHQ with a single, central phenyl ring. The first group is the largest and comprised of molecules that possess two aromatic rings connected by a central “hinge” composed of one to three carbon atoms (**2**, **3**, **4**, **6**, **7**, **8**, **9**, and **13**) whereas **1**, **5**, **11**, **10**, and **12** can be assigned to the second group. One of the strengths of computational docking is its straightforward visualization of the predicted intermolecular interactions that take place between a ligand and the amino

acids in the binding site. Inspection of each molecule's top-ranked docking pose in the consensus reveals a common binding mode assumed by most (except for **6**) of the bisphenol analogs. In comparison to the binding pose of BHQ, the space normally occupied by the central phenyl moiety of BHQ accommodates the "hinges" connecting the two aromatic rings (illustrated for **7** in Fig. 5). The aromatic rings, on the other hand, reside in areas in which the *tert*-butyl groups of BHQ are located. This observation is in agreement with binding modes assumed by structurally related compounds characterized in an earlier study [6]. Not surprisingly, compounds **5** and **11**, whose structures resemble more closely that of BHQ, are predicted to bind to SERCA in orientations that place their central phenyl ring on top of that of BHQ (illustrated for **5** in Fig. 5) and their substituents in the areas occupied by the *tert*-butyl groups of BHQ. Unique poses were assumed by compounds **1**, **6**, **10**, and **12** (not shown). It should be noted that **1** resembles the alkylphenols whose endocrine disrupting properties have been attributed to SERCA inhibition [55].

Consistent with earlier reports on other SERCA inhibitors [32], the breakdown of the ChemScore scoring function into individual contributions reveals a combination of hydrogen bonds and hydrophobic contacts as the major forces responsible for the binding of the inhibitors investigated in this study. Hydrogen bonds with the inhibitors' hydroxyl groups involve predominantly the backbone carbonyl oxygen atom of Pro308 and the side chain carboxyl oxygen atoms of Asp59 and Asp254 (Fig. 5, lower panels). Most of the hydrophobic contacts are mediated by the non-polar moieties of the amino acids Leu61, Val62, Leu65, Leu253, Leu256, Ile307, Pro308, Glu309, Leu311, and Pro312 that line the surface of the binding site and give it its overall hydrophobic character (Fig. 5, upper panels).

3.2. Interpretation of the descriptors employed by the RP models

As demonstrated by others [48,50], the nature of the descriptors employed in RP classification trees can reveal structural requirements that make a compound active – without prior knowledge of the receptor's 3D structure. Since this type of analysis can become ambiguous if the trees are complex and highly branched, we examined the tree of Model 1, which had the simplest structure (Fig. 6; see Supplementary Materials for trees generated by Models 2 and 3). The descriptor employed in the first split of the tree is *n4pag12*, the count of the number of paths of length 4 that join nodes of degree 1 and 2. If *n4pag12* takes a value of smaller than one, the compound is directly classified as inactive in a terminal leaf. Fig. 7 illustrates the meaning of this particular descriptor in the case of BHQ. In the more general case of hydroquinones with two *tert*-butyl groups, *n4pag12* reflects the substitution pattern at the phenyl ring. For this descriptor to take a non-zero value, the unsubstituted position at the ring needs to be in between a hydroxyl and a *tert*-butyl group, which is the case for BHQ but not for some of its isomers with lower potencies.

Like the first split, the second one also generates a terminal leaf that is inactive. The descriptor relevant for this split is *Sssss4*, which is defined as the sum to the electrotopological state indices of all quaternary carbon atoms in a molecule [56,57]. The majority of the quaternary carbon atoms present in the training set are part of *tert*-butyl substituents on a phenyl ring, which underscores the need for this type of bulky, non-polar inhibitor moieties. More importantly, *Sssss4* emphasizes the nature of the environment in which this key feature finds itself within the molecule. Electrotopological state atom descriptors such as *Sssss4* take into account the intrinsic nature of the atom in question, in this case the quaternary carbon atom, and also the influence of the neighboring atoms. An examination of the structures assigned to this inactive leaf shows that they all possess at least two quaternary carbon atoms, just like BHQ. However, unlike BHQ, they lack strong hydrogen-bond acceptors at the phenyl ring. For example, the inactive 1,4-di-*tert*-butylbenzene has no substituents other than hydrogen atoms in addition to its two *tert*-butyl

groups. Another straightforward example is the inactive 2,5-di-*tert*-butylaniline, a molecule with one additional amine substituent whose lone pair at the nitrogen atom is a rather weak hydrogen bond acceptor. Both of these materials are inactive, but replacing one of the hydrogen atoms in 1,4-di-*tert*-butylbenzene or the single amine in 2,5-di-*tert*-butylaniline with a hydroxyl group (e.g., 2,5-di-*tert*-butylphenol) yields an active compound with a remarkable IC_{50} value of 6.81 μM . This observation suggests that it is the combination of hydrophobic groups in close proximity to strong hydrogen-bond acceptors such as hydroxyl groups that is crucial for bioactivity. The implied pattern of hydrogen-bonding and hydrophobic substituents as captured by the descriptor Sssss4 is compatible with the structure of SERCA's BHQ binding site that reveals the hydrogen-bond forming amino acids (Pro308 and Asp59) nearby two hydrophobic pockets (Fig. 5). Lastly, it should be noted that in both the first and the second split of the tree, compounds are eliminated because they are predicted to be inactive, a feature that is not easily implemented in traditional QSAR models that primarily focus on properties that make a compound active.

Another noteworthy descriptor used by Model 1 is Qv, which is indicative of the overall polarity of a molecule [57]. This descriptor is used in the third split and reoccurs in the fifth. It suggests that the narrow window of polarity for Qv between 1.1 and 0.7 is a requisite for potency. Again, this notion is consistent with crystallographic information that shows the need for a somewhat polar ligand to achieve complementarity with the polarity pattern of the binding site. The polarity of BHQ, for example, is primarily due to the presence of the polar hydrogen-bonding hydroxyl groups and partially offset by the non-polar *tert*-butyl groups.

3.3. Future use of the screening protocols

With regard to the overall hit rate defined as the number of compounds with detectable potencies below 100 μM divided by the number of compounds tested, Protocol 2 appears to be faring better ($12/33 = 36\%$) than Protocol 1 ($5/39 = 13\%$). This finding is not entirely unexpected since the availability of a high resolution 3D structure of the receptor facilitating the use of structure-based docking methodologies is in general considered to be a significant advantage whereas the strength of ligand-based approaches such as RP is their broader applicability. None of the compounds selected by Protocol 1 were considered "active" by Protocol 2 because they failed the similarity criterion, did not produce a docking consensus, or received a score lower than that of BHQ. On the other hand, seven of the compounds selected by Protocol 2 for testing were also classified as "active" by Protocol 1 (GRI numbers: 290792, 415530, 455028, 513898, 283992, 305506, and 402297), but escaped the random selection process of molecules present in the second highest occupied activity bin. With the exception of these seven materials, there was no further overlap between the two pools of test compounds, suggesting that the two screening protocols seem to focus on different compound properties. In a perfect but unlikely scenario, both protocols would have selected the same pool of inhibitors, all of which would have been active. This suggests that – despite their quite remarkable performance in the present study that lead to the identification of a total of 17 novel molecules with IC_{50} values below 100 μM from a compound library with 345,000 entries – both protocols can be improved further.

In the case of Protocol 1, the RP model's performance is dependent on the size and diversity of the training set. Ideally, the structural diversity of the training set would be representative of the entire library, but given the tremendous size difference between the two pools, this is a difficult goal to achieve. In any event, the newly acquired experimental data obtained in this study can be included in an expanded training set for the future development of more comprehensive RP models. Ultimately, this would lead to an iterative VS approach in which several small compound sets are selected and screened so that each selection benefits from incorporating the results of the previous iteration. Another point for improvement may be the loosening of the molecular weight restriction in the prefilter to allow the inclusion of

larger compounds. Indeed, some of the inhibitors identified by Protocol 2 which does not employ a MW filter are larger than the cutoff value of 350 g/mol. Evidently, some of the larger inhibitors can interact favorably with residues in proximity of the binding cavity and should thus not be excluded based on their size.

As far as Protocol 2 is concerned, the rather stringent similarity filter in the first step could be modified to include a larger number of compounds for subsequent docking. Whereas omitting the filter altogether and docking the entire library would be unrealistic because of limitations in computational resources, using a somewhat lower molecular similarity index cutoff to include several hundred more compounds in the next step appears to be a reasonable strategy. Moreover, the implementation of a flexible binding site during docking could be attempted. This would permit the simulation of conformational selection, a phenomenon causing the binding site to adopt a conformation in the presence of an inhibitor that is different from its conformation in the inhibitor-free state. Local binding site flexibility can be achieved by certain docking programs such as the latest releases of AutoDock [58] and GOLD, which allow specified amino acid side chains to flex while placing the inhibitor in the binding site. However, protocols involving flexible binding sites need to be carefully chosen and validated because a previous study has shown that such protocols can sometimes lead to erroneous solutions caused by unrealistically large distortions of SERCA's BHQ binding site [59]. Limitations of current flexible docking tools relate to larger, global conformational changes involving the movement of backbone atoms or entire secondary structural elements that cannot be simulated properly. As an alternative, a preliminary ligand pose obtained by rigid docking could be further optimized by molecular dynamics (MD) simulations and then rescored. Whereas MD simulations are ideal for the thorough study of a small number of selected inhibitors, their high demands on computer time constitute a considerable obstacle for large-scale virtual screens of sizeable compound libraries.

4. Experimental and Computational Methods

4.1. Modeling of molecular structures

Molecular models of the compounds in the training set had been created in three earlier studies [1,6,32] and were used without further manipulations. These structures had been modeled in Sybyl (versions 7.1 and 8.0; Tripos, St. Louis, MO) and their conformational energies had been minimized by molecular mechanics (MMFF94s force field, MMFF94 charges, $\epsilon(r) = 4$, termination at 0.01 kcal/(mol Å)) using a conjugate gradient method. Models of the 3D structures of GRI library molecules were obtained electronically from the curator. Since neither Protocol 1 nor Protocol 2 required optimized 3D conformations, all structures were used for VS without further manipulations.

4.2. Calculation of molecular descriptors and physical properties

The topological descriptors were computed with the program winMolconn (version 1.0.1.3; Hall Associates Consulting, Quincy, MA). The descriptors covered a broad set of indices including molecular connectivity indices, atom and bond-based electrotopological state indices, and counts of functional group types. The log P and log WS parameters were calculated with the programs CSlogP and CSlogWS, respectively, from ChemSilico LLC (Tewksbury, MA).

4.3. Similarity and substructure filters

The Tanimoto index-based similarity search preceding computational docking (Protocol 2) was conducted in ChemFinder (version 9.0.1; CambridgeSoft, Cambridge, MA), using a

65% cut-off value with regard to BHQ. The substructure search that was part of Protocol 1 was performed with ChemBioFinder (version 12.0; CambridgeSoft, Cambridge, MA).

4.4. Recursive partitioning

After grouping the molecules into classes (“bins”) according to their activities (two classes in the case of Model 1, seven classes in the case of Model 2, and six classes in the case of Model 3), RP classification trees (Fig. 6 and Supplementary Materials) were created with the program JMP (version 7.0.2; SAS; Cary, NC). Final leaves of the classification trees containing compounds belonging to more than one bin were assigned average bin values that were rounded to the closest integer.

4.5. Molecular docking

The crystal structure of the SERCA/BHQ available from the Protein Databank (2AGV) [33] was prepared for docking as described previously by removing all non-protein entities, adding hydrogen atoms, and conducting an energy minimization with the AMBER force field ($\epsilon(r) = 4$) while keeping the positions of the heavy atoms invariable [32]. For docking, the program GOLD (Genetic Optimisation for Ligand Docking, Cambridge Crystallographic Data Centre, UK) in conjunction with the ChemScore scoring function was used [52,53,60,61]. The genetic algorithm was executed at the default settings and the docking sphere had a radius of 15 Å centered on the position normally occupied the C-2 phenyl carbon (atom number 15395) of BHQ in the crystal structure. Each molecule was docked 20 times under identical conditions. An in-house program examined the output of CCDC’s RMS analysis utility (executed as part of the GOLD run) to determine if 10 or more of the poses were clustered within 2.0 Å of each other. If this was the case, a majority consensus was declared and the results were analyzed further.

4.6. SERCA activity inhibition assays

SERCA microsomes were prepared from rabbit hind leg tissue according to a standard procedure [32,62]. All reagents required for the ATPase activity assays were obtained from Fisher Scientific (Pittsburgh, PA), with the exception of the enzymes PK and LDH, which were received from Sigma-Aldrich (St. Louis, MO). Small samples of potential inhibitors dissolved in DMSO at 10 mM were kindly provided by Dr. William Seibel at GRI (University of Cincinnati). Due to limitations in the amount of sample material, all samples were used in bioassays without further characterization or purification.

The SERCA-catalyzed rate of ATP hydrolysis coupled to the oxidation of NADH by the action of the enzymes PK and LDH was measured spectroscopically at a wavelength of 340 nm [6,32]. Samples (225 µL) containing 6.67 µg SERCA, 100 mM KCl, 5 mM MgCl₂, 0.5 mM EGTA, 0.7 mM CaCl₂, 4.5 µM calcimycin, 1.5 mM phosphoenolpyruvate, 10.7 µL of a PK/LDH mixture (600–1000 units/mL and 900–1400 units/mL, respectively), 20 mM Trizma (pH: 7.5), and varying concentrations of inhibitor were prepared in 96-well polystyrene plates. After triggering the reaction by the addition of 11 µL of a 5 mM ATP solution, rates were measured with a SpectraMax 190 microplate reader (Molecular Devices, Sunnyvale, CA) for five minutes. Inhibitory potency was reported as the IC₅₀ value, the inhibitor concentration which reduced SERCA activity by 50%. IC₅₀ values were obtained by fitting the measured rates versus inhibitor concentration to a three parameter logistic equation. Reported results are the averages of at least three independent repeats.

All compounds inhibiting the rate of NADH oxidation were tested in an independent phosphate liberation assay that directly monitored the SERCA-catalyzed production of inorganic phosphate via its chromogenic reaction with the dye malachite green. In a 96-well polystyrene plate, 10 µL of a 0.1 mg protein/mL SERCA solution in 10 mM EGTA were

mixed with 102.5 μL of 0.1 M KCl, 5 mM MgCl_2 , 0.5 mM EGTA, 4.5 μM calcimycin, 0.7 mM CaCl_2 , and 20 mM Trizma (pH 7.5). 11 μL of 5 mM ATP was added to initiate the reaction, which was then stopped after delay times between 1 and 10 minutes by the addition of a solution of malachite green (3 mM), sodium molybdate (10 mM), Triton-X-100 (0.05%), and 0.7 M HCl. The absorbance of each well was measured at 630 nm with a plate reader and reaction rates were obtained by linear regression of absorbance values versus time. Since the sole purpose of this second assay was to rule out inhibitor interference with PK or LDH in the coupled assay, it was only conducted at two or three inhibitor concentrations [6].

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by grants from the Kentucky Biomedical Research Infrastructure Network (P20RR016481-09), Research Corporation (Award 6843), and the National Institutes of Health (1R15GM084431-01) to S.P. We thank Dr. William Seibel at the Genome Research Institute at the University of Cincinnati for making sample materials available to us. We are grateful to Dr. Robert Kempton and Dr. Manori Jayasinghe for critically reading the manuscript.

References

1. Paula S, Abell J, Deye J, Elam C, Lape M, Purnell J, Ratliff R, Sebastian K, Zultowsky J, Kempton RJ. Design, synthesis, and biological evaluation of hydroquinone derivatives as novel inhibitors of the sarco/endoplasmic reticulum calcium ATPase. *Bioorganic & medicinal chemistry*. 2009; 17:6613–6619. [PubMed: 19699645]
2. Treiman M, Caspersen C, Christensen SB. A tool coming of age: thapsigargin as an inhibitor of sarco-endoplasmic reticulum Ca^{2+} -ATPases. *Trends Pharmacol Sci*. 1998; 19:131–135. [PubMed: 9612087]
3. Denmeade SR, Isaacs JT. The SERCA pump as a therapeutic target: making a "smart bomb" for prostate cancer. *Cancer Biol Ther*. 2005; 4:14–22. [PubMed: 15662118]
4. Denmeade SR, Jakobsen CM, Janssen S, Khan SR, Garrett ES, Lilja H, Christensen SB, Isaacs JT. Prostate-specific antigen-activated thapsigargin prodrug as targeted therapy for prostate cancer. *J Natl Cancer Inst*. 2003; 95:990–1000. [PubMed: 12837835]
5. Isaacs J. New strategies of the medical treatment of prostate cancer. *BJU International*. 2005; 96:35–40. [PubMed: 16359437]
6. Deye J, Elam C, Lape M, Ratliff R, Evans K, Paula S. Structure-based virtual screening for novel inhibitors of the sarco/endoplasmic reticulum calcium ATPase and their experimental evaluation. *Bioorganic & medicinal chemistry*. 2009; 17:1353–1360. [PubMed: 19117760]
7. Christensen SB, Andersen A, Poulsen JC, Treiman M. Derivatives of thapsigargin as probes of its binding site on endoplasmic reticulum Ca^{2+} ATPase. Stereoselectivity and important functional groups. *FEBS Lett*. 1993; 335:345–348. [PubMed: 8262181]
8. Nielsen SF, Thastrup O, Pedersen R, Olsen CE, Christensen SB. Structure-activity relationships of analogues of thapsigargin modified at O-11 and O-12. *J Med Chem*. 1995; 38:272–276. [PubMed: 7830270]
9. Norup E, Smitt UW, Christensen SB. The potencies of thapsigargin and analogues as activators of rat peritoneal mast cells. *Planta Med*. 1986; 52:251–255. [PubMed: 2429340]
10. Andersen A, Treiman M, Poulsen JC, Cornett C, Moldt P, Olsen CE, Christensen SB. Ca^{2+} -ATPase inhibitory activity of a locked analogue of thapsigargin. *Bioorganic & medicinal chemistry letters*. 1994; 4:657–660.
11. Ball M, Andrews SP, Wierschem F, Cleator E, Smith MD, Ley SV. Total synthesis of thapsigargin, a potent SERCA pump inhibitor. *Org Lett*. 2007; 9:663–666. [PubMed: 17256950]

12. Ley SV, Antonello A, Balskus EP, Booth DT, Christensen SB, Cleator E, Gold H, Hogenauer K, Hunger U, Myers RM, Oliver SF, Simic O, Smith MD, Sohoel H, Woolford AJ. Synthesis of the thapsigargin. *Proc Natl Acad Sci U S A*. 2004; 101:12073–12078. [PubMed: 15226504]
13. Goeger DE, Riley RT, Dorner JW, Cole RJ. Cyclopiazonic acid inhibition of the Ca^{2+} -transport ATPase in rat skeletal muscle sarcoplasmic reticulum vesicles. *Biochem Pharmacol*. 1988; 37:978–981. [PubMed: 2964239]
14. Plenge-Tellechea F, Soler F, Fernandez-Belda F. On the inhibition mechanism of sarcoplasmic or endoplasmic reticulum Ca^{2+} -ATPases by cyclopiazonic acid. *J Biol Chem*. 1997; 272:2794–2800. [PubMed: 9006919]
15. Seidler NW, Jona I, Vegh M, Martonosi A. Cyclopiazonic acid is a specific inhibitor of the Ca^{2+} -ATPase of sarcoplasmic reticulum. *J Biol Chem*. 1989; 264:17816–17823. [PubMed: 2530215]
16. Soler F, Plenge-Tellechea F, Fortea I, Fernandez-Belda F. Cyclopiazonic acid effect on Ca^{2+} -dependent conformational states of the sarcoplasmic reticulum ATPase. Implication for the enzyme turnover. *Biochemistry*. 1998; 37:4266–4274. [PubMed: 9521749]
17. Appendino G, Prosperini S, Valdivia C, Ballero M, Colombano G, Billington RA, Genazzani AA, Sterner O. SERCA-inhibiting activity of C-19 terpenolides from *Thapsia garganica* and their possible biogenesis. *J Nat Prod*. 2005; 68:1213–1217. [PubMed: 16124763]
18. Bartolommei G, Tadini-Buoninsegni F, Hua S, Moncelli MR, Inesi G, Guidelli R. Clotrimazole inhibits the Ca^{2+} -ATPase (SERCA) by interfering with Ca^{2+} binding and favoring the E2 conformation. *J Biol Chem*. 2006; 281:9547–9551. [PubMed: 16452481]
19. Benzaquen LR, Brugnara C, Byers HR, Gatton-Celli S, Halperin JA. Clotrimazole inhibits cell proliferation in vitro and in vivo. *Nat Med*. 1995; 1:534–540. [PubMed: 7585119]
20. Snajdrova L, Xu A, Narayanan N. Clotrimazole, an antimycotic drug, inhibits the sarcoplasmic reticulum calcium pump and contractile function in heart muscle. *J Biol Chem*. 1998; 273:28032–28039. [PubMed: 9774419]
21. Berman MC, Karlsh SJ. Interaction of an aromatic dibromoisothiuronium derivative with the Ca^{2+} -ATPase of skeletal muscle sarcoplasmic reticulum. *Biochemistry*. 2003; 42:3556–3566. [PubMed: 12653560]
22. Hoving S, Bar-Shimon M, Tijmes JJ, Goldshleger R, Tal DM, Karlsh SJ. Novel aromatic isothiuronium derivatives which act as high affinity competitive antagonists of alkali metal cations on Na/K-ATPase. *J Biol Chem*. 1995; 270:29788–29793. [PubMed: 8530371]
23. Hua S, Xu C, Ma H, Inesi G. Interference with phosphoenzyme isomerization and inhibition of the sarco-endoplasmic reticulum Ca^{2+} ATPase by 1,3-dibromo-2,4,6-tris(methylisothiuronium) benzene. *J Biol Chem*. 2005; 280:17579–17583. [PubMed: 15746094]
24. Inesi G, Hua S, Xu C, Ma H, Seth M, Prasad AM, Sumbilla C. Studies of Ca^{2+} ATPase (SERCA) inhibition. *J Bioenerg Biomembr*. 2005; 37:365–368. [PubMed: 16691466]
25. Ogunbayo OA, Lai PF, Connolly TJ, Michelangeli F. Tetrabromobisphenol A (TBBPA), induces cell death in TM4 Sertoli cells by modulating Ca^{2+} transport proteins and causing dysregulation of Ca^{2+} homeostasis. *Toxicol In Vitro*. 2008; 22:943–952. [PubMed: 18329244]
26. Ogunbayo OA, Michelangeli F. The widely utilized brominated flame retardant tetrabromobisphenol A (TBBPA) is a potent inhibitor of the SERCA Ca^{2+} pump. *Biochem J*. 2007; 408:407–415. [PubMed: 17784851]
27. Bilmen JG, Khan SZ, Javed MH, Michelangeli F. Inhibition of the SERCA Ca^{2+} pumps by curcumin. Curcumin putatively stabilizes the interaction between the nucleotide-binding and phosphorylation domains in the absence of ATP. *Eur J Biochem*. 2001; 268:6318–6327. [PubMed: 11733029]
28. Wootton LL, Michelangeli F. The effects of the phenylalanine 256 to valine mutation on the sensitivity of sarcoplasmic/endoplasmic reticulum Ca^{2+} ATPase (SERCA) Ca^{2+} pump isoforms 1, 2, and 3 to thapsigargin and other inhibitors. *J Biol Chem*. 2006; 281:6970–6976. [PubMed: 16410239]
29. Khan YM, Wictome M, East JM, Lee AG. Interactions of dihydroxybenzenes with the Ca^{2+} -ATPase: separate binding sites for dihydroxybenzenes and sesquiterpene lactones. *Biochemistry*. 1995; 34:14385–14393. [PubMed: 7578042]

30. Moore GA, McConkey DJ, Kass GE, O'Brien PJ, Orrenius S. 2,5-Di(*tert*-butyl)-1,4-benzohydroquinone - a novel inhibitor of liver microsomal Ca^{2+} sequestration. *FEBS Lett.* 1987; 224:331–336. [PubMed: 2961610]
31. Wictome M, Holub M, East JM, Lee AG. The importance of the hydroxyl moieties for inhibition of the Ca^{2+} -ATPase by trilobolide and 2,5-di(*tert*-butyl)-1,4-benzohydroquinone. *Biochem Biophys Res Commun.* 1994; 199:916–921. [PubMed: 8135840]
32. Lape M, Elam C, Versluis M, Kempton R, Paula S. Molecular determinants of sarco/endoplasmic reticulum calcium ATPase inhibition by hydroquinone-based compounds. *Proteins.* 2008; 70:639–649. [PubMed: 17879345]
33. Obara K, Miyashita N, Xu C, Toyoshima I, Sugita Y, Inesi G, Toyoshima C. Structural role of countertransport revealed in Ca^{2+} pump crystal structure in the absence of Ca^{2+} *Proc Natl Acad Sci U S A.* 2005; 102:14489–14496. [PubMed: 16150713]
34. Takahashi M, Kondou Y, Toyoshima C. Interdomain communication in calcium pump as revealed in the crystal structures with transmembrane inhibitors. *Proc Natl Acad Sci U S A.* 2007; 104:5800–5805. [PubMed: 17389383]
35. Seifert MHJ, Wolf K, Vitt D. Virtual high-throughput *in silico* screening. *Biosilico.* 2003; 1:143–149.
36. Walters WP, Stahl MT, Murcko MA. Virtual screening - an overview. *Drug Discovery Today.* 1998; 3:160–178.
37. Cavasotto CN, Orry AJ. Ligand docking and structure-based virtual screening in drug discovery. *Current topics in medicinal chemistry.* 2007; 7:1006–1014. [PubMed: 17508934]
38. Sun H. Pharmacophore-based virtual screening. *Current medicinal chemistry.* 2008; 15:1018–1024. [PubMed: 18393859]
39. Reddy AS, Pati SP, Kumar PP, Pradeep HN, Sastry GN. Virtual screening in drug discovery -a computational perspective. *Current protein & peptide science.* 2007; 8:329–351. [PubMed: 17696867]
40. Stahura FL, Bajorath J. New methodologies for ligand-based virtual screening. *Current pharmaceutical design.* 2005; 11:1189–1202. [PubMed: 15853666]
41. Cavasotto CN, Ortiz MA, Abagyan RA, Piedrafita FJ. In silico identification of novel EGFR inhibitors with antiproliferative activity against cancer cells. *Bioorganic & medicinal chemistry letters.* 2006; 16:1969–1974. [PubMed: 16413185]
42. Dooley AJ, Shindo N, Taggart B, Park JG, Pang YP. From genome to drug lead: identification of a small-molecule inhibitor of the SARS virus. *Bioorganic & medicinal chemistry letters.* 2006; 16:830–833. [PubMed: 16325400]
43. Carbone V, Ishikura S, Hara A, El-Kabbani O. Structure-based discovery of human L-xylulose reductase inhibitors from database screening and molecular docking. *Bioorganic & medicinal chemistry.* 2005; 13:301–312. [PubMed: 15598553]
44. Gupta SP. Quantitative structure-activity relationships of carbonic anhydrase inhibitors. *Progress in drug research.* *Fortschritte der Arzneimittelforschung.* 2003; 60:171–204.
45. Gupta SP. Quantitative structure-activity relationships of renin inhibitors. *Mini reviews in medicinal chemistry.* 2003; 3:315–321. [PubMed: 12678825]
46. Rusinko A 3rd, Farman MW, Lambert CG, Brown PL, Young SS. Analysis of a large structure/biological activity data set using recursive partitioning. *J Chem Inf Comput Sci.* 1999; 39:1017–1026. [PubMed: 10614024]
47. Rusinko A 3rd, Young SS, Drewry DH, Gerritz SW. Optimization of focused chemical libraries using recursive partitioning. *Comb Chem High Throughput Screen.* 2002; 5:125–133. [PubMed: 11966421]
48. Choi I, Kim SY, Kim H, Kang NS, Bae MA, Yoo SE, Jung J, No KT. Classification models for CYP450 3A4 inhibitors and non-inhibitors. *Eur J Med Chem.* 2009; 44:2354–2360. [PubMed: 19026469]
49. Dubus E, Ijjaali I, Petitot F, Michel A. In silico classification of HERG channel blockers: a knowledge-based strategy. *ChemMedChem.* 2006; 1:622–630. [PubMed: 16892402]

50. Jones DR, Ekins S, Li L, Hall SD. Computational approaches that predict metabolic intermediate complex formation with CYP3A4 (+b5). *Drug Metab Dispos.* 2007; 35:1466–1475. [PubMed: 17537872]
51. Lamanna C, Bellini M, Padova A, Westerberg G, Maccari L. Straightforward recursive partitioning model for discarding insoluble compounds in the drug discovery process. *J Med Chem.* 2008; 51:2891–2897. [PubMed: 18419111]
52. Jones G, Willett P, Glen RC. Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation. *J Mol Biol.* 1995; 245:43–53. [PubMed: 7823319]
53. Jones G, Willett P, Glen RC, Leach AR, Taylor R. Development and validation of a genetic algorithm for flexible docking. *J Mol Biol.* 1997; 267:727–748. [PubMed: 9126849]
54. Van Veldhoven P, Mannaerts G. Inorganic and organic phosphate measurements in the nanomolar range. *Analytical Biochemistry.* 1987; 161:45–48. [PubMed: 3578786]
55. Michelangeli F, Ogunbayo OA, Wootton LL, Lai PF, Al-Mousa F, Harris RM, Waring RH, Kirk CJ. Endocrine disrupting alkylphenols: structural requirements for their adverse effects on Ca^{2+} pumps, Ca^{2+} homeostasis & Sertoli TM4 cell viability. *Chem Biol Interact.* 2008; 176:220–226. [PubMed: 18773882]
56. Kier LB, Hall LH. An electrotopological-state index for atoms in molecules. *Pharm Res.* 1990; 7:801–807. [PubMed: 2235877]
57. Kier, LB.; Hall, LH. *Molecular Structure Description: The Electrotopological State.* New York: Academic Press; 1999.
58. Goodsell DS, Olson AJ. Automated docking of substrates to proteins by simulated annealing. *Proteins.* 1990; 8:195–202. [PubMed: 2281083]
59. Lape M, Elam C, Paula S. Comparison of current docking tools for the simulation of inhibitor binding by the transmembrane domain of the sarco/endoplasmic reticulum calcium ATPase. *Biophys Chem.* 2010; 150:88–97. [PubMed: 20167416]
60. Baxter CA, Murray CW, Clark DE, Westhead DR, Eldridge MD. Flexible docking using Tabu search and an empirical estimate of binding affinity. *Proteins.* 1998; 33:367–382. [PubMed: 9829696]
61. Eldridge MD, Murray CW, Auton TR, Paolini GV, Mee RP. Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *J Comput Aided Mol Des.* 1997; 11:425–445. [PubMed: 9385547]
62. MacLennan DH. Purification and properties of an adenosine triphosphatase from sarcoplasmic reticulum. *J Biol Chem.* 1970; 245:4508–4518. [PubMed: 4250726]

Figure 1.
Flow chart describing the selection process for the RP-based (Protocol 1) and the docking-based (Protocol 2) VS protocols.

Figure 2. Correlation between experimentally observed and RP model-predicted bin occupancies (Model 3; active compounds only). Bars of the same color belong to the same predicted activity bin.

Figure 3.

Consensus selection method. Selected compounds as indicated by the red asterisk had to be members of one of the marked bins in each of the three models.

Figure 4.

Chemical structure and GRI inventory number of SERCA inhibitors measureable potencies. Inhibitors **1** – **13** had IC₅₀ values below 50 μ M (“potent inhibitors”) whereas inhibitors **14** – **17** had IC₅₀ values between 50 and 100 μ M (“weak inhibitors”). Compounds **11** – **13**, **15** and **17** were identified by Protocol 1 and all others by Protocol 2.

Figure 5.

Docking-predicted poses and interactions between compounds **7** (left panels) and **5** (right panels). In the upper panels, the pose of BHQ according to the X-ray crystal structure is displayed in yellow as a reference. The upper panels emphasize the complementarity in shape and in hydrophobic profile of the inhibitors and the binding site. The MolCad surface (Connolly) is colored according to hydrophobicity (brown: more hydrophobic; green: less hydrophobic). The lower panels show residues predicted to be engaged in hydrogen bonds with the inhibitors.

Figure 6.
Classification tree employed by Model 1. Terminal leaves are colored according to their classification as “active” (green) or “inactive” (purple).

Figure 7.
Example for the occurrence of the substructure of BHQ that is captured by the descriptor n4ga12.

Table 1

Descriptive statistics for octanol/water partition coefficients (P) and water solubility (WS) of the training set compounds.

parameter	mean	standard deviation	minimum	maximum
log P, actives	4.5	1.2	4.4	6.6
log P, inactives	3.8	1.6	−0.12	6.3
log WS, actives	−4.7	1.2	−6.9	−1.8
log WS, inactives	−3.8	1.5	−7.0	−0.3

Table 2

Definition of bin assignments used in range binning for the development of RP models 2 and 3. Entries indicate the logarithm of IC₅₀ values expressed in μM .

lower limit	upper limit	class (bin) assigned
> 0		1
-0.5	0	2
-1.0	-0.5	3
-1.5	-1.0	4
-2.0	-1.5	5
-2.5	-2.0	6
inactive		7

Table 3

Experimentally determined and computationally predicted potencies of all compounds tested. Compounds declared as “inactive” had no measureable inhibitory potency below 100 μM .

GRI Compound Code	Selection Protocol	Observed IC_{50} / μM	Observed $\log 1/\text{IC}_{50}$	Model-1	Model-2	Model-3	ChemScore kJ mol^{-1}
382864	1	12 ± 8	-1.1	1	2	3	N/A
402297	1	60 ± 15	-1.8	1	2	3	N/A
911098	1	13 ± 5	-1.1	1	2	2	N/A
945830	1	44 ± 13	-1.6	1	2	3	N/A
954355	1	64 ± 8	-1.8	1	2	2	N/A
99248	2	28 ± 9	-1.4	1	6	4	31.9
225920	2	19 ± 12	-1.3	1	4	4	30.9
283992	2	14 ± 3	-1.2	1	6	6	28.2
290199	2	5.2 ± 2.5	-0.71	1	7	4	27.2
305506	2	69 ± 37	-1.8	1	6	5	28.8
376381	2	32 ± 5	-1.5	1	7	4	28.6
387297	2	3.8 ± 0.5	-0.58	1	4	6	33.0
406402	2	52 ± 39	-1.7	0	3	4	30.1
481715	2	2.2 ± 0.5	-0.35	1	7	3	27.9
519791	2	25 ± 16	-1.4	0	7	6	27.3
765292	2	3.2 ± 1.0	-0.51	0	4	5	30.6
857336	2	30 ± 9	-1.5	0	7	3	32.3
116648	1	Inactive	Inactive	1	2	3	N/A
121165	1	Inactive	Inactive	1	2	3	N/A
122573	1	Inactive	Inactive	1	2	3	N/A
155757	1	Inactive	Inactive	1	2	3	N/A
171714	1	Inactive	Inactive	1	2	2	N/A
193280	1	Inactive	Inactive	1	2	3	N/A
251080	1	Inactive	Inactive	1	2	3	N/A
251843	1	Inactive	Inactive	1	2	3	N/A

GRI Compound Code	Selection Protocol	Observed IC ₅₀ / μM	Observed log I/IC ₅₀	Model-1	Model-2	Model-3	ChemScore kJ mol ⁻¹
263581	1	Inactive	Inactive	1	2	3	N/A
281668	1	Inactive	Inactive	1	2	3	N/A
284382	1	Inactive	Inactive	1	2	2	N/A
378898	1	Inactive	Inactive	1	2	3	N/A
379040	1	Inactive	Inactive	1	2	3	N/A
390755	1	Inactive	Inactive	1	2	3	N/A
396074	1	Inactive	Inactive	1	2	3	N/A
405901	1	Inactive	Inactive	1	2	3	N/A
496399	1	Inactive	Inactive	1	2	3	N/A
499492	1	Inactive	Inactive	1	2	2	N/A
507319	1	Inactive	Inactive	1	2	3	N/A
514940	1	Inactive	Inactive	1	2	3	N/A
519324	1	Inactive	Inactive	1	2	3	N/A
526497	1	Inactive	Inactive	1	2	3	N/A
705836	1	Inactive	Inactive	1	2	3	N/A
767792	1	Inactive	Inactive	1	2	3	N/A
768153	1	Inactive	Inactive	1	2	2	N/A
772348	1	Inactive	Inactive	1	2	3	N/A
858475	1	Inactive	Inactive	1	2	3	N/A
859176	1	Inactive	Inactive	1	2	2	N/A
883820	1	Inactive	Inactive	1	2	3	N/A
912715	1	Inactive	Inactive	1	2	3	N/A
913188	1	Inactive	Inactive	1	2	3	N/A
917150	1	Inactive	Inactive	1	2	2	N/A
932556	1	Inactive	Inactive	1	2	3	N/A
933527	1	Inactive	Inactive	1	2	3	N/A
249548	2	Inactive	Inactive	0	7	5	36.4
290792	2	Inactive	Inactive	1	3	6	32.0

GRI Compound Code	Selection Protocol	Observed IC ₅₀ / μ M	Observed log 1/IC ₅₀	Model-1	Model-2	Model-3	ChemScore kJ mol ⁻¹
291028	2	Inactive	Inactive	1	7	4	27.2
374051	2	Inactive	Inactive	1	7	4	28.2
374288	2	Inactive	Inactive	1	7	4	27.9
374408	2	Inactive	Inactive	1	7	5	38.1
377112	2	Inactive	Inactive	0	7	4	27.0
413652	2	Inactive	Inactive	0	2	3	27.3
415530	2	Inactive	Inactive	1	4	4	28.7
455028	2	Inactive	Inactive	1	4	5	30.1
491311	2	Inactive	Inactive	0	7	6	28.7
497671	2	Inactive	Inactive	1	5	4	27.5
511231	2	Inactive	Inactive	0	3	6	30.6
511254	2	Inactive	Inactive	0	6	6	35.1
511257	2	Inactive	Inactive	0	2	3	27.6
513898	2	Inactive	Inactive	1	3	4	27.1
539261	2	Inactive	Inactive	1	7	4	30.4
857863	2	Inactive	Inactive	0	7	4	28.5
946572	2	Inactive	Inactive	0	6	4	30.6
947022	2	Inactive	Inactive	1	7	4	28.0
954235	2	Inactive	Inactive	0	7	4	27.1