

Review

Reviewing Ligand-Based Rational Drug Design: The Search for an ATP Synthase Inhibitor

Chia-Hsien Lee ¹, Hsuan-Cheng Huang ^{2,*} and Hsueh-Fen Juan ^{1,3,*}

¹ Graduate Institute of Biomedical Electronics and Bioinformatics, National Taiwan University, Taipei 106, Taiwan; E-Mail: r99945042@ntu.edu.tw

² Institute of Biomedical Informatics and Center for Systems and Synthetic Biology, National Yang-Ming University, Taipei 112, Taiwan

³ Institute of Molecular and Cellular Biology and Department of Life Science, National Taiwan University, Taipei 106, Taiwan

* Authors to whom correspondence should be addressed; E-Mails: hsuancheng@ym.edu.tw (H.-C.H.); yukijuan@ntu.edu.tw (H.-F.J.); Tel.: +886-2-2826-7357 (H.-C.H.); +886-2-3366-4536 (H.-F.J.); Fax: +886-2-2367-3374 (H.-F.J.).

Received: 23 February 2011; in revised form: 4 August 2011 / Accepted: 11 August 2011 /

Published: 17 August 2011

Abstract: Following major advances in the field of medicinal chemistry, novel drugs can now be designed systematically, instead of relying on old trial and error approaches. Current drug design strategies can be classified as being either ligand- or structure-based depending on the design process. In this paper, by describing the search for an ATP synthase inhibitor, we review two frequently used approaches in ligand-based drug design: The pharmacophore model and the quantitative structure-activity relationship (QSAR) method. Moreover, since ATP synthase ligands are potentially useful drugs in cancer therapy, pharmacophore models were constructed to pave the way for novel inhibitor designs.

Keywords: pharmacophore; quantitative structure-activity relationships; ATP synthase ligands

1. Introduction

Previously, new drugs were found only by a trial and error approach. Potentially, anything deemed to be of medicinal value in a specific disease could be tested on patients to determine its efficacy [1]. As the science of chemistry advanced, researchers started to purify the active compounds in herbal preparations known to have medicinal properties, and deduce the structures of these active compounds [1]. The science of drug design progressed further with advances in molecular biology and biochemistry, which elaborated the concepts of genes and ligand-receptor relationships [2]. The most important breakthrough came with the emergence of genomics, proteomics and the development of bioinformatics and chemoinformatics. The concept and techniques of large-scale screening, which can be applied to identify disease-related genes and proteins, were first introduced in genomics and proteomics [3], as bioinformatics [4] and chemoinformatics [5] further sought to uncover methods for large-scale data processing and *in silico* experimentation.

The goal of drug design is to select a target in the etiology of a disease and find one or more compounds which interacts with that target [6], and in doing so activates or blocks the given target [2]. Ideally, the resultant changes in target protein activity will go on to influence a series of reactions and lead to an improvement in the clinical outcome [7].

After a suitable target has been identified, the drug design process can begin. If reliable information on the 3-D structure and active sites of the target protein can be obtained from X-ray crystallography, nuclear magnetic resonance, or 3-D structure databases, and incorporated into a computer model, compounds binding to the target can be designed [8]. This approach is known as “structure-based drug design”. Frequently used techniques in this approach are docking and molecular dynamics simulation [9]. Potent ligands can be found by screening a molecule database with docking software [10]. Molecular dynamics simulation can be useful to ascertain not only how a molecule interacts with the target protein, but also to determine some other properties of the molecule itself, such as membrane permeability [11,12].

In some cases, usually in which data pertaining to the 3-D structure of a target protein are not available, drug design can instead be based on processes using the known ligands of a target protein as the starting point. This approach is known as “ligand-based drug design”. Molecular similarity approaches, quantitative structure-activity relationships (QSAR) and pharmacophore models are frequently used methods in the ligand-based drug design process [13]. By using the molecular fingerprints of known ligands, databases can be screened to find molecules with similar fingerprints [14]. Common structural features of ligands can be found using pharmacophore modeling, which can then be used to screen for molecules with these features [15]. To predict the activity of a novel molecule, models can be built with QSAR [16]. While a pharmacophore model may only indicate the activity-conferring features of an active ligand, the relationship between chemical or physical properties of ligand and biological activity can be more fully explored using the QSAR model.

This review will focus on QSAR and pharmacophore modeling and elaborate on their fundamental concepts, workflows for building models, and their applications. Pharmacophore models of the ATP synthase beta subunit-binding ligands selected from existing literature are also discussed here as an illustration.

2. Pharmacophore

The term “pharmacophore” was first defined by Ehrlich as: “a molecular framework that carries the essential features responsible for a drug’s biological activity” [17]. It follows from this definition that a pharmacophore defines the necessary features that an active ligand should possess. Generally, feature type, position, and direction of an active ligand would be encoded into a pharmacophore model, along with possible steric constraints of the active compound [18].

A 3-D pharmacophore would reflect how key amino acids are positioned in the active site of a target protein [19]. For example, a key amino acid residue which acts as a hydrogen-bond acceptor should be in the vicinity of a hydrogen-bond donor feature in the pharmacophore model, accounting, in part, for the protein-ligand interaction. Once a ligand binds to the target protein in the correct conformation and interacts with key amino acid residues, the conformation of the protein may be changed or become locked, depending on the mechanism of ligand-protein interaction [2].

A pharmacophore model can be generated from a set of known ligands. However, data pertaining to 3-D protein structure or protein-ligand complexes combined with information on active sites can also be used to model a pharmacophore [20]. By studying the binding site, possible interactions between the active compound and the protein can be inferred, and pharmacophore models can be built from data on target protein structure.

Pharmacophore models are widely used to elicit specific inhibitors of disease-related proteins, including G-protein coupled receptors, enzymes, and ion channels [21]. It is also used with other drug discovery methods, as has been described in the “Applications” subsection.

2.1. Construction of a Pharmacophore Model

The detailed workflow of pharmacophore model construction will depend on the software package one uses. However, the general procedure can be summarized as follows [18]: Ligands are first selected and their usability confirmed. If a 3-D pharmacophore model is being constructed, sets of conformers, or conformations of a ligand, are generated. According to IUPAC, conformations are defined as the spatial arrangement of the atoms affording distinction between stereoisomers [22]. The conformers, or the 2-D structures if a 2-D pharmacophore model, which define only the required atom types and their connectivity [23] being built, are then used to determine features common to the selected ligands required to build the pharmacophore model. In general, more than one pharmacophore model is generated from the selected ligand set. Ranking and validating the model is the final step in pharmacophore model construction. Most pharmacophore modelers conduct the whole workflow with the help of a chemoinformatics software package, since all the required tools are found in this type of package.

The first step in pharmacophore modeling is to assemble an active compound set, usually from literature searches and molecular database querying. The researchers can use all databases available, including commercial, in-house, or public databases. A consistent threshold to define a compound active should be applied while searching active compounds from multiple sources. Activity levels of the compounds are not necessary if it is not necessary to build a pharmacophore-based QSAR model subsequently. After a set of molecules is selected, ensuring that all selected molecules exert their biological effects via the same mechanism is another important issue which should be considered

before continuing pharmacophore model construction [18]. The structures of the molecules should be confirmed after checking their biological mechanism. In some cases, the chemical structures are manually sketched in a software package, which is an error-prone process. Even if the chemical information comes from a well-known database, a modeler should be cautious since such data is not always accurate. The pharmacophore model describes the interaction between the active site of a target protein and a ligand [19]; therefore the use of data pertaining to different active sites on the target would produce an entirely spurious pharmacophore model. Even if the pharmacophore elucidation software could generate results from such input, the resulting model would be meaningless.

If a representative 3-D pharmacophore model is to be built, a reasonable conformation set should be generated prior to alignment. The quality of the output set of the conformation generator, including such considerations as coverage and energy of the conformers, impacts on the quality of generated model [24].

The next step is to build the pharmacophore model from the common features found among the ligands. To find common features, the ligands are aligned, or superimposed with the assistance either of a conformer database of the relevant ligands or of an on-the-fly conformation generator [20]. After the alignment step is completed, the pharmacophore models are generated by a pharmacophore elucidation algorithm. More than one output model can often be obtained from this procedure, and selecting the best ones from the results is an important task [25]. This is usually assisted by the scoring function contained within the pharmacophore-building software. Generally, the model with the highest score is the most representative member of the output model set.

The best models are subjected to the validation process. If the receptor-ligand binding mechanism is clearly understood, this information can be utilized to validate the pharmacophore model, since the pharmacophore model can sometimes reflect the 3-D structure of the binding site, as previously stated [19]. If a pharmacophore model conflicts with previously established binding mechanisms, the modeler should consider rejecting it, or try to account for the discrepancies between documented binding data and the newly built pharmacophore model. A pharmacophore model should always be validated by ligands that are not included in the training set despite whether previously known binding mechanisms exist or not. If a pharmacophore model cannot pass this validation, it should be rejected and the modeling process reviewed.

2.2. Applications of Pharmacophore Models

Pharmacophore models are often employed in virtual screening processes. For example, Mustata *et al.* discovered novel Myc-Max heterodimer disruptors [26] by using a pharmacophore model generated using known disruptors. In this study, the authors performed the virtual screening with the pharmacophore model after it was validated by using a non-training molecule set. Bioassays were then conducted to check the potency of the screened molecules.

Another example of the use of pharmacophore models in drug discovery is the identification of novel PPAR γ partial agonists by Petersen *et al.* [27]. The authors found that full agonists at the PPAR γ had unwanted side effects, as previous research had indicated [28], so it was a priority to find new compounds which could act as partial agonists at PPAR γ . A pharmacophore model was built from a collection of known partial agonists, and it was validated with a newly discovered partial agonist.

After inspecting the hit list generated by virtual screening with the pharmacophore model, methyl oleanonate, a derivative of oleanonic acid, was selected because oleanonic acid can be easily obtained. Purification and bioassay were conducted after its potential to act as a partial agonist was confirmed by docking simulation.

In addition to the uses described above, pharmacophore models can be used as filters to reduce false positives, either during preprocessing of a molecular database before submission to another virtual screening procedure [29], or to filter the results of other virtual screening processes [30].

An example of the use of a pharmacophore model as a preprocessor can be found in the Chk-1 kinase study by Lyne *et al.* [29]. In this research, the authors preprocessed the molecule database using molecular weight and rotational bond limits and a user-defined 3-D pharmacophore model. The remaining ligands were then docked into the binding site of the kinase and scored by a scheme defined by the authors. The potencies of the resulting molecules were tested in a Chk-1 kinase assay, an *in vitro* bioassay protocol.

Pharmacophore models can also be used to process the output of virtual screening protocols, as shown in the research of Peach *et al.* [30]. Although the pharmacophore model in this study was generated manually from receptor binding site data obtained from literature and 3-D structure database searches, it aptly demonstrated the way in which a pharmacophore model can be used as a filter. In this research, the authors supported the validity of their method with three examples. First, the pharmacophore model was constructed from 3-D receptor-ligand complexes and from literature data. A docking simulation was performed, with all output tested by the pharmacophore models. During this process, a molecule is only recognized as a hit if it matches with all generated pharmacophore models. This approach performed better than scoring functions provided by the docking algorithms used here. The value of this approach, in contrast to the pharmacophore preprocessor method, is that the time-consuming docking simulation does not need to be repeated if the pharmacophore model is subsequently changed [30].

3. QSAR

QSAR, which stands for “quantitative structure-activity relationships”, is a method that relates chemical structure to biological or chemical activity using mathematical models [31]. If the activity of a set of ligands can be determined, a model can be constructed to describe this relationship. Unlike a pharmacophore model, which encodes only the essential features of an active ligand, the QSAR model allows one to determine the effect of a certain property on the activity of a molecule. For example, the QSAR model may reveal a property to have a highly negative, or alternatively a weak positive effect on ligand activity. Such information is not available using a pharmacophore model.

Quantifying the structure and activity of a ligand is important in the modeling process. Structure quantification is not a trivial problem, since a structure cannot be represented by a mere value. Instead, a set of properties, usually known as the “descriptors”, is computed from the structure and used to quantify it. By using structural descriptors as independent variables and activity as a dependent variable, a model can be built to describe the relationship between the two [32].

After a QSAR model is built and validated, it can predict the biological activity of novel molecules from their structural properties. A QSAR model can also screen potentially active molecules from a

database, as described in the section on applications of the technique. Because the QSAR model can incorporate a wide range of different variables, be it physical, chemical or biological, it can also be utilized in industries apart from drug design [33], such as toxicology [34], food chemistry [35], and other fields.

3.1. Building a QSAR Model

The process of constructing a QSAR model can be summarized as follows [36]: First, ligands and their activities are collected. Descriptors are calculated and selected before a mathematical modeling method is chosen and the ligand data are then used to construct the QSAR models. After the models are completed, they are tested by internal and external validation procedures. Only then can a QSAR model be used in any practical applications, such as predicting the activity of a novel compound.

As is the case when building a pharmacophore model, the active ligand set must be gathered from molecular databases or from literature searches before QSAR modeling begins. The process requires not only the collection of ligand structures but also of their activities. Generally, IC_{50} s (half maximal inhibitory concentration) [37], EC_{50} s (half maximal effective concentration) [38], and K_i values (inhibition constant) [39] are commonly used to quantify drug activity. However, the quantification of ligand activity as used in QSAR is not limited to pharmacokinetic parameters. Other activity indexes can also be incorporated into model depending on the phenomena one wishes to predict. For example, Chen *et al.* used corneal permeability coefficient as the activity index to construct a model that predicted the efficiency of drug delivery in various eye drop medications [40]. In addition to structure verification as described in the section on pharmacophore model construction, ligand activity data should also be checked. All activity data should come from the same experimental procedure or assay, and it is preferable if the data comes from the same laboratory, and even the same researcher [32].

Before a QSAR model can be built, ligand structure descriptors should be ascertained or calculated. Some descriptors obtained directly from data sources or calculated using simple arithmetic operations take into account the specific number of atoms, molecular chain length, molecular mass, *etc.* However, other descriptors may require complex computation, for example pharmacophore-based descriptors [41], molecular field descriptors, which are derived from the interaction of probes and molecules and used in CoMFA and CoMSIA [42,43], and spectral descriptors derived from the IR spectrum of the ligand [44]. It is important that the descriptors are related to the biological or chemical activity which the model will be used to predict [45]. In other words, if a descriptor is not related to activity, one should avoid incorporating the descriptor into the modeling process.

After the activity index (the dependent variable) and descriptors (the independent variables) are prepared for each ligand, a variable selection method and a modeling method can be selected, and a model is built. The selection process should avoid redundant descriptors, and then a model is built using the selected descriptors [46]. If two descriptors represent a similar biological or chemical parameter, one of them should be disregarded. In order to select descriptors, genetic algorithms [47,48], principle component analysis [49], artificial neural networks [50] and k -nearest neighbor [51] approaches can all be used. If a linear model is assumed, some conventional statistical methods, such as the partial least squares method and multiple linear regression [16] can be used. If a nonlinear

model is preferred on the other hand, machine learning methods like artificial neural networks [52] or support vector machines [53] can be applied.

The main differences among the frequently used QSAR algorithms reside in their means of descriptor generation. For example, most QSAR algorithms, like CoMFA [42], CoMSIA [43], CoMMA [54], and HypoGen [55], use similar linear statistical models to explore the relationship between activity and descriptors, which are calculated by different processes. In CoMFA and CoMSIA, pre-aligned molecules are put onto a grid, or lattice. The descriptors are calculated by the interaction of the molecule and a probe is placed at each intersection of the lattice. The differences between CoMFA and CoMSIA are in the use of different probes and interaction-calculating functions. In CoMFA, only probes representing steric and electrostatic interactions can be used. In CoMSIA, probes representing hydrophobic and hydrogen bond interactions, in addition to CoMFA probes may be selected. In addition, CoMSIA uses a Gaussian-type function for calculating prober-molecule interaction. By using such a smooth function, the result value is more reasonable than the function used in CoMFA, and defining a cut-off limit to remove invalid values is no longer required. The descriptors used in CoMMA are generated by computing the spatial moments of the molecules. On the other hand, the fit value is the only descriptor source employed in the HypoGen model generation process. The fit value describes the goodness of alignment between a ligand and a pharmacophore model and is obtained from a pharmacophore model generated and optimized using known structure and activity data.

The model must then be validated before it can be used to predict activity. There are some popular methods used to validate a QSAR model [56], including internal validation approaches (such as the “leave-one-out” or “leave-n-out” cross validation methods [57]), and external validation approaches. In cross validation, one (leave-one-out) or more (leave-n-out) ligand of the training set is excluded. The excluded data is predicted by the model constructed with reduced training set data. These steps are repeated until all data has been excluded and predicted, and the power of a model is determined by the accuracy of prediction [57]. External validation is a widely used method, and is considered important in the QSAR building pipeline [58]. In external validation, the capability of the model is tested using data which is not included in the training set, in contrast to internal validation, which utilizes data taken from the training set to validate the model [59]. In most of the studies, both internal and external validations are performed to ensure the reliability of the model (see following subsection “Applications of QSAR”). After the model has passed these strict validation tests, it can be used to predict the activity of novel molecules.

3.2. Applications of QSAR

As previously discussed, QSAR is widely used when a predictive model is required. In the work of De Melo, a PC (principle component) -SAR model was developed to predict if a molecule is active, and a PLS (partial least squares) -QSAR model was developed to predict the degree of activity of an active compound [37]. The author validated the QSAR model employing both internal (the “leave-n-out” method) and external validation approaches. From the two models, the author found several important chemical properties that contributed to the activity of the molecule.

Another example of this type of method is the QSAR analysis of CCR2 inhibitors, by Saghale *et al.* [60]. CCR2 is a chemokine G-protein coupled receptor important in certain inflammatory

disorders. In this study, the authors used a genetic algorithm to select key descriptors, and then built the QSAR model using a stepwise multiple linear regression approach. The model was validated in accordance with the literature [58] both internally (using the “leave-one-out” and “leave-five-out” methods) and externally. However, no bioassay data were given in this study.

In a study conducted by Obiol-Pardo *et al.* [61], two QSAR models were incorporated into a multiscale simulation system which was designed to predict drug-induced cardiotoxicity by the interactions of drug molecules with two potassium channels, hERG and KCNQ1. First, a docking simulation was executed to determine the active conformation of the ligands. QSAR models, which used field-based descriptors, were constructed for each potassium channel by the partial least squares method. The QSAR models were validated internally (with the “leave-one-out” method) and externally after model construction. The IC_{50} values for drugs acting on hERG and KCNQ1 predicted by the QSAR models were used as the input for two previously studied models, which could predict action potential duration and pseudo-ECG (for extraction of the QT interval) respectively. Drug-induced arrhythmia was reported to affect the QT interval and action potential duration. By employing this multiscale model, the cardiotoxicity of a novel drug was assessed without using any *in vitro* or *in vivo* assays.

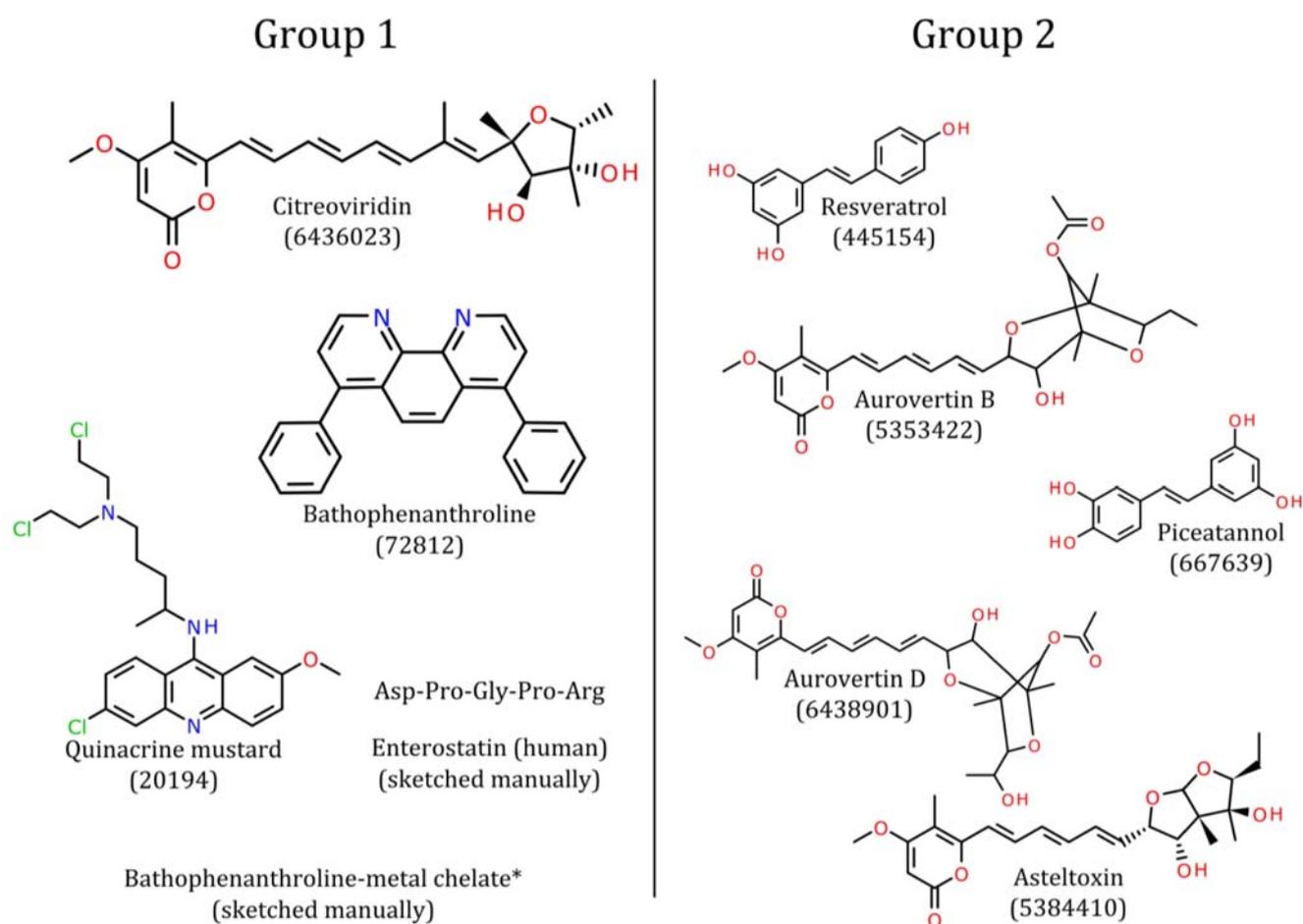
4. Pharmacophore Models of ATP Synthase Beta Subunit-Binding Ligands

In our previous research, ATP synthase was discovered to be over-expressed on the cell membranes of a breast cancer cell line [62]. Aurovertin B, which binds to the beta subunits of ATP synthase, was also found to be a drug of potential value in the targeted therapy of breast cancer [62]. However, because of the structural complexity of aurovertin B, which makes it a difficult compound to synthesize, it has become important to design a new drug to target ATP synthase for use in cancer research. To achieve this goal, the common backbone, a substructure bearing required functional groups of the ligands, should be deduced. Although a shared backbone might be found by aligning all the molecules, important groups might not be inferred from such an alignment. The backbone and important functional groups are both vital to the design of a new inhibitor. For this reason, a pharmacophore model was built to determine common functional groups. After relevant functional groups are identified, shared backbone structure can be deduced from original structure by removing branches which contain no necessary functional groups if the molecules can be soundly aligned to the pharmacophore and each other.

Data pertaining to known ligands of the ATP synthase beta subunit were retrieved from the PubChem [63] database by querying with their names found on the literature [64], with the exception of the peptide enterostatin and bathophenanthroline-metal chelates (see Figure 1). These two molecules were sketched manually and checked carefully using the Discovery Studio package. The ligands were separated into two groups, Group 1 (G1) and Group 2 (G2), according to their binding site location as described in the literature [64]. The compounds of G2 are known to bind to the same site on the ATP synthase beta subunit, whereas the G1 molecules are known to bind to the ATP synthase beta subunit, although the precise location of their binding site(s) was not clear: molecules in this group may bind to different location of ATP synthase beta subunit. Because reliable activity data

for some molecules could not be obtained, we decided to generate a common feature pharmacophore model instead of a 3-D QSAR pharmacophore model.

Figure 1. ATP synthase beta subunit ligands used in the pharmacophore modeling process. Structures, names, and PubChem Compound database IDs are listed. * Bathophenanthroline-metal chelate was constructed from bathophenanthroline and ferrous iron in the ratios 1:1 and 3:1, but a conformer could not be generated (as stated in the following paragraph). Since the conformer generation of the chelate failed, it was not included in the training set when building the pharmacophore model.

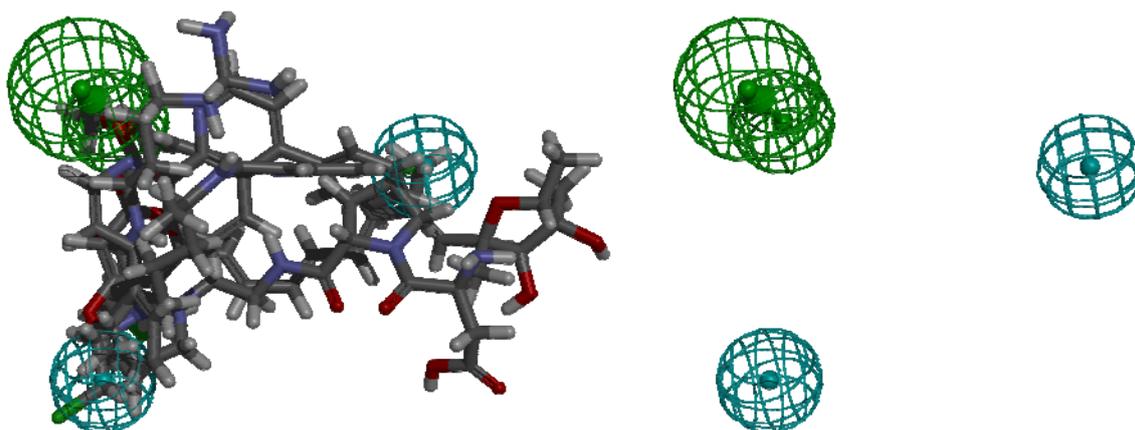


Common feature pharmacophore models for each group were constructed using the HipHop algorithm [65] after the conformers were generated using the “BEST” algorithm [66]. All the algorithms are included in the Discovery Studio package (version 2.1) from Accelrys.

The best pharmacophore model of the G1 set, as determined by the ranking score generated by the HipHop algorithm, is shown in Figure 2. As described in previous discussions, since the molecules in G1 did not share a common binding site on ATP synthase beta subunit, they should not be used as training sets when building pharmacophore models. However, we would like to demonstrate that pharmacophore models can be built by pharmacophore elucidating software even if the training set cannot meet the criteria discussed above. Bathophenanthroline-metal chelate is not present in the alignment of the final result. The conformation generator could not generate conformers for the chelate mixture, thus the chelate was automatically excluded from the input training set. Despite the fact that

G1 molecules are not known to have a common binding site, a pharmacophore model could be generated by HipHop, which was an unexpected finding. The ligands might share the same binding site on the ATP synthase beta subunit because they share some chemical features, as indicated by the generated pharmacophore model; further structural studies of ATP synthase are required to reveal the binding site(s) of these ligands. Nevertheless, researchers need to shoulder the responsibility of checking the binding characteristics of ligands, as described in the previous section; otherwise a pharmacophore model which stands a high chance of being meaningless may be generated, as demonstrated by the pharmacophore model presented here.

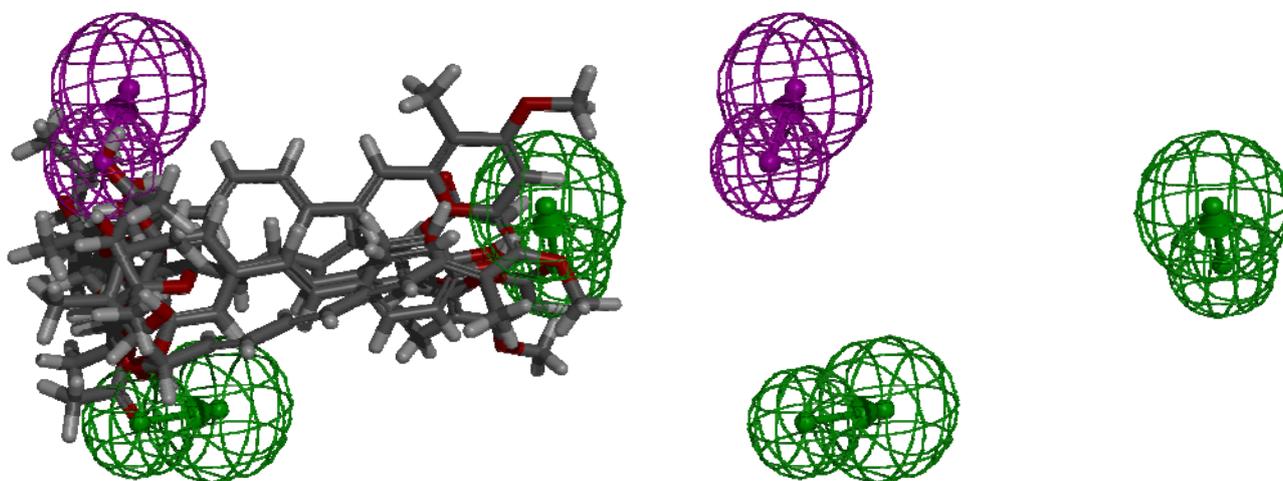
Figure 2. The pharmacophore model generated from the G1 molecule set. Left panel: pharmacophore model and aligned ligands; Right panel: pharmacophore model only; Green indicates a hydrogen bond acceptor feature; Cyan, a hydrophobic feature. The direction of the hydrogen bond acceptor (shown with arrows) indicated an optimal alignment of the features.



The best pharmacophore model of the G2 set, as determined by the ranking score generated by the HipHop algorithm, is shown in Figure 3. As predicted, the G2 molecules were aligned better than the G1 molecules. Three similar molecules, aurovertin B, aurovertin D, and asteltoxin, were in this group. In addition, resveratrol and piceatannol are two very similar molecules. However, it can be clearly seen that the G2 molecules can be separated into two sets by putting resveratrol and piceatannol together into one group, and putting the remaining molecules into the other. Although this action might affect the generated pharmacophore model, the results bear some resemblance to the results of the docking simulation performed in our previous study [62]: The hydrogen bond donor feature can be mapped to the Lue342 residue, and the hydrogen bond acceptor features can be mapped to the Lys382, Arg412 residues of the ATP synthase beta subunit. These observations were not beyond what we expected to achieve, since a 3-D pharmacophore model can reflect the arrangement of the key residues of the binding site [19]. However, after careful inspection, the pharmacophore model feature vectors indicating the direction of the hydrogen bonds were different from those of the previous docking simulation, and the interaction poses were different as well. In addition, the fit value (not shown) of aurovertin B was not as high as expected. As in our previous research, the activity of aurovertin B was significantly higher than resveratrol and piceatannol [62]. Furthermore, accurate alignment of the two

aurovertins and asteltoxin was expected because of their structural similarity but was not observed in our results. Although this pharmacophore model can reflect the binding mode discovered in our previous research, this model was not validated experimentally. We present the model here for demonstration of the pharmacophore modeling process. One should conduct validation before using this pharmacophore model in their research.

Figure 3. The pharmacophore model generated from the G2 molecules. Left panel: the pharmacophore model and the aligned ligands; Right panel: pharmacophore model only; Green indicates a hydrogen bond acceptor feature; Purple, a hydrogen bond donor feature. The direction of the hydrogen bond acceptor and hydrogen bond donor (shown with arrows) indicates optimal alignment of the features.



5. Conclusions

QSAR and pharmacophore models are two widely used techniques in the rational drug design process. In this review, the basics of the pharmacophore and QSAR techniques were introduced, and the basic modeling processes for both methods, along with some possible pitfalls, were discussed. After the fundamental concepts of the two methods were covered, some studies using each technique were explored. Previously unpublished pharmacophore models generated from a range of ATP synthase beta subunit ligands were also used to demonstrate the workflow of a pharmacophore modeling process. It is important to know the basics and pitfalls of these computer-aided drug design techniques. Knowledge of such techniques can aid researchers in selecting the best technique for their given objectives. However, results may still be prone to error, especially when there is disagreement between input and assumptions of the selected techniques. It is also important to bear in mind that better results may be yielded by combining QSAR and pharmacophore models with other drug design methods, as shown in the example sections of this review.

Acknowledgements

This work was supported by the National Research Program for Genomic Medicine, National Science Council of Taiwan (NSC 100-3112-B-002-002, NSC 99-2621-B-002-005-MY3), and the National Taiwan University Frontier and Innovative Research Program.

References

1. Sneader, W. *Drug Discovery: A History*; John Wiley & Sons, Ltd.: Chichester, UK, 2005.
2. Drews, J. Drug discovery: a historical perspective. *Science* **2000**, *287*, 1960–1964.
3. Butcher, E.C.; Berg, E.L.; Kunkel, E.J. Systems biology in drug discovery. *Nat. Biotechnol.* **2004**, *22*, 1253–1259.
4. Chen, Y.P.; Chen, F. Identifying targets for drug discovery using bioinformatics. *Expert Opin. Ther. Targets* **2008**, *12*, 383–389.
5. Engel, T. Basic overview of chemoinformatics. *J. Chem. Inf. Model.* **2006**, *46*, 2267–2277.
6. Moon, J.B.; Howe, W.J. Computer design of bioactive molecules: a method for receptor-based *de novo* ligand design. *Proteins* **1991**, *11*, 314–328.
7. Sleno, L.; Emili, A. Proteomic methods for drug target discovery. *Curr. Opin. Chem. Biol.* **2008**, *12*, 46–54.
8. Cavasotto, C.N.; Phatak, S.S. Homology modeling in drug discovery: current trends and applications. *Drug Discov. Today* **2009**, *14*, 676–683.
9. Ivanov, A.S.; Veselovsky, A.V.; Dubanov, A.V.; Skvortsov, V.S. Bioinformatics platform development: from gene to lead compound. *Methods Mol. Biol.* **2006**, *316*, 389–431.
10. Chen, Z.; Li, H.L.; Zhang, Q.J.; Bao, X.G.; Yu, K.Q.; Luo, X.M.; Zhu, W.L.; Jiang, H.L. Pharmacophore-based virtual screening *versus* docking-based virtual screening: a benchmark comparison against eight targets. *Acta. Pharmacol. Sin.* **2009**, *30*, 1694–1708.
11. Wang, Y.; Shaikh, S.A.; Tajkhorshid, E. Exploring transmembrane diffusion pathways with molecular dynamics. *Physiology* **2010**, *25*, 142–154.
12. Bemporad, D.; Luttmann, C.; Essex, J.W. Behaviour of small solutes and large drugs in a lipid bilayer from computer simulations. *Biochim. Biophys. Acta* **2005**, *1718*, 1–21.
13. Tintori, C.; Manetti, F.; Botta, M. Pharmacophoric models and 3D QSAR studies of the adenosine receptor ligands. *Curr. Top. Med. Chem.* **2010**, *10*, 1019–1035.
14. Vogt, M.; Bajorath, J. Predicting the performance of fingerprint similarity searching. *Methods Mol. Biol.* **2011**, *672*, 159–173.
15. Langer, T.; Hoffmann, R.D. Virtual screening: an effective tool for lead structure discovery? *Curr. Pharm. Des.* **2001**, *7*, 509–527.
16. Scior, T.; Medina-Franco, J.L.; Do, Q.T.; Martínez-Mayorga, K.; Yunes Rojas, J.A.; Bernard, P. How to recognize and workaroud pitfalls in QSAR studies: a critical review. *Curr. Med. Chem.* **2009**, *16*, 4297–4313.
17. Ehrlich, P. Ueber den jetzigen Stand der Chemotherapie. *Ber. Dtsch. Chem. Ges.* **1909**, *42*, 17–47.
18. van Drie, J.H. Pharmacophore discovery-lessons learned. *Curr. Pharm. Des.* **2003**, *9*, 1649–1664.
19. Wolber, G.; Seidel, T.; Bendix, F.; Langer, T. Molecule-pharmacophore superpositioning and pattern matching in computational drug design. *Drug Discov. Today* **2008**, *13*, 23–19.

20. Yang, S.Y. Pharmacophore modeling and applications in drug discovery: challenges and recent advances. *Drug Discov. Today* **2010**, *15*, 444–450.
21. Kubinyi, H. Success Stories of Computer-Aided Design. In *Computer Applications in Pharmaceutical Research and Development*; Ekins, S., Wang, B., Eds.; Wiley-Interscience: Hoboken, NJ, USA, 2006; pp. 377–424.
22. McNaught, A.D.; Wilkinson, A. *Compendium of Chemical Terminology-IUPAC Recommendations*, 2nd ed.; Blackwell Science: Hoboken, NJ, USA, 1997.
23. Rankovic, Z.; Morphy, R. *Lead Generation Approaches in Drug Discovery*; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2011; pp. 86–90.
24. Chen, I.J.; Foloppe, N. Is conformational sampling of drug-like molecules a solved problem? *Drug Dev. Res.* **2011**, *72*, 85–94.
25. Güner, O.F. History and evolution of the pharmacophore concept in computer-aided drug design. *Curr. Top. Med. Chem.* **2002**, *2*, 1321–1332.
26. Mustata, G.; Follis, A.V.; Hammoudeh, D.I.; Metallo, S.J.; Wang, H.; Prochownik, E.V.; Lazo, J.S.; Bahar, I. Discovery of novel Myc-Max heterodimer disruptors with a three-dimensional pharmacophore model. *J. Med. Chem.* **2009**, *52*, 1247–1250.
27. Petersen, R.K.; Christensen, K.B.; Assimopoulou, A.N.; Fretté, X.; Papageorgiou, V.P.; Kristiansen, K.; Kouskoumvekaki, I. Pharmacophore-driven identification of PPAR γ agonists from natural sources. *J. Comput. Aided Mol. Des.* **2011**, *25*, 107–116.
28. Barroso, I.; Gurnell, M.; Crowley, V.E.; Agostini, M.; Schwabe, J.W.; Soos, M.A.; Maslen, G.L.; Williams, T.D.; Lewis, H.; Schafer, A.J. Dominant negative mutations in human PPAR γ associated with severe insulin resistance, diabetes mellitus and hypertension. *Nature* **1999**, *402*, 880–883.
29. Lyne, P.D.; Kenny, P.W.; Cosgrove, D.A.; Deng, C.; Zabudoff, S.; Wendoloski, J.J.; Ashwell, S. Identification of compounds with nanomolar binding affinity for checkpoint kinase-1 using knowledge-based virtual screening. *J. Med. Chem.* **2004**, *47*, 1962–1968.
30. Peach, M.L.; Nicklaus, M.C. Combining docking with pharmacophore filtering for improved virtual screening. *J. Cheminform.* **2009**, *1*, 6:1–6:15.
31. Esposito, E.X.; Hopfinger, A.J.; Madura, J.D. Methods for applying the quantitative structure-activity relationship paradigm. *Methods. Mol. Biol.* **2004**, *275*, 131–214.
32. Perkins, R.; Fang, H.; Tong, W.; Welsh, W.J. Quantitative structure-activity relationship methods: perspectives on drug discovery and toxicology. *Environ. Toxicol. Chem.* **2003**, *22*, 1666–1679.
33. Du, Q.S.; Huang, R.B.; Chou, K.C. Recent advances in QSAR and their applications in predicting the activities of chemical molecules, peptides and proteins for drug design. *Curr. Protein Pept. Sci.* **2008**, *9*, 248–260.
34. Bradbury, S.P. Quantitative structure-activity relationships and ecological risk assessment: an overview of predictive aquatic toxicology research. *Toxicol. Lett.* **1995**, *79*, 229–237.
35. Martinez-Mayorga, K.; Medina-Franco, J.L. Chemoinformatics-applications in food chemistry. *Adv. Food. Nutr. Res.* **2009**, *58*, 33–56.
36. Myint, K.Z.; Xie, X.Q. Recent advances in fragment-based QSAR and multi-dimensional QSAR methods. *Int. J. Mol. Sci.* **2010**, *11*, 3846–3866.

37. De Melo, E.B. Multivariate SAR/QSAR of 3-aryl-4-hydroxyquinolin-2(1H)-one derivatives as type I fatty acid synthase (FAS) inhibitors. *Eur. J. Med. Chem.* **2010**, *45*, 5817–5826.
38. Zhou, T.; Shi, Q.; Chen, C.H.; Zhu, H.; Huang, L.; Ho, P.; Lee, K.H. Anti-AIDS agents 79. Design, synthesis, molecular modeling and structure-activity relationships of novel dicamphanoyl-2',2'-dimethyldihydropyranochromone (DCP) analogs as potent anti-HIV agents. *Bioorg. Med. Chem.* **2010**, *18*, 6678–6689.
39. Karolidis, D.A.; Agatonovic-Kustrin, S.; Morton, D.W. Artificial neural network (ANN) based modelling for D1 like and D2 like dopamine receptor affinity and selectivity. *Med. Chem.* **2010**, *6*, 259–270.
40. Chen, C.; Yang, J. MI-QSAR models for prediction of corneal permeability of organic compounds. *Acta Pharmacol. Sin.* **2006**, *27*, 193–204.
41. Tropsha, A.; Zheng, W. Identification of the descriptor pharmacophores using variable selection QSAR: applications to database mining. *Curr. Pharm. Des.* **2001**, *7*, 599–612.
42. Cramer, R.D., III; Patterson, D.E.; D. Bunce, J.D. Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. *J. Am. Chem. Soc.* **1988**, *110*, 5959–5967.
43. Klebe, G.; Abraham, U.; Mietzner, T. Molecular similarity indices in a comparative analysis (CoMSIA) of drug molecules to correlate and predict their biological activity. *J. Med. Chem.* **1994**, *37*, 4130–4146.
44. Turner, D.B.; Willett, P. The EVA spectral descriptor. *Eur. J. Med. Chem.* **2000**, *35*, 367–375.
45. Katritzky, A.R.; Gordeeva, E.V. Traditional topological indices vs. electronic, geometrical, and combined molecular descriptors in QSAR/QSPR research. *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 835–857.
46. Whitley, D.C.; Ford, M.G.; Livingstone, D.J. Unsupervised forward selection: a method for eliminating redundant variables. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 1160–1168.
47. Chen, H.; Zhou, J.; Xie, G. PARM: a genetic evolved algorithm to predict bioactivity. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 243–250.
48. Rogers, D.; Hopfinger, A.J. Application of genetic function approximation to quantitative structure-activity relationships and quantitative structure-property relationships. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 854–866.
49. Xue, L.; Godden, J.W.; Bajorath, J. Evaluation of descriptors and mini-fingerprints for the identification of molecules with similar activity. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 1227–1234.
50. Wikel, J.H.; Dow, E.R. The use of neural networks for variable selection in QSAR. *Bioorg. Med. Chem. Lett.* **1993**, *3*, 645–651.
51. Itskowitz, P.; Tropsha, A. k nearest neighbors QSAR modeling as a variational problem: theory and applications. *J. Chem. Inf. Model.* **2005**, *45*, 777–785.
52. Jain, A.N.; Koile, K.; Chapman, D. Compass: Predicting biological activities from molecular surface properties. performance comparisons on a steroid benchmark. *J. Med. Chem.* **1994**, *37*, 2315–2327.
53. Shahlaei, M.; Fassihi, A.; Saghale, L. Application of PC-ANN and PC-LS-SVM in QSAR of CCR1 antagonist compounds: a comparative study. *Eur. J. Med. Chem.* **2010**, *45*, 1572–1582.

54. Silverman, B.D.; Platt, D.E. Comparative molecular moment analysis (CoMMA): 3D-QSAR without molecular superposition. *J. Med. Chem.* **1996**, *39*, 2129–2140.
55. Kurogi, Y.; Güner, O.F. Pharmacophore modeling and three-dimensional database searching for drug design using catalyst. *Curr. Med. Chem.* **2001**, *8*, 1035–1055.
56. Verma, J.; Khedkar, V.M.; Coutinho, E.C. 3D-QSAR in drug design—a review. *Curr. Top. Med. Chem.* **2010**, *10*, 95–115.
57. Cramer, R.D., III; Bunce, J.D.; Patterson, D.E.; Frank, I.E. Crossvalidation, bootstrapping, and partial least squares compared with multiple regression in conventional QSAR studies. *Quant. Struct.-Act. Relat.* **1988**, *7*, 18–25.
58. Tropsha, A.; Gramatica, P.; Gombar, V.K. The importance of being earnest: validation is the absolute essential for successful application and interpretation of QSPR models. *QSAR Comb. Sci.* **2003**, *22*, 69–77.
59. Consonni, V.; Ballabio, D.; Todeschini, R. Comments on the definition of the Q2 parameter for QSAR validation. *J. Chem. Inf. Model.* **2009**, *49*, 1669–1678.
60. Saghale, L.; Shahlaei, M.; Fassihi, A.; Madadkar-Sobhani, A.; Gholivand, M.B.; Pourhossein, A. QSAR analysis for some diaryl-substituted pyrazoles as CCR2 inhibitors by GA-stepwise MLR. *Chem. Biol. Drug. Des.* **2011**, *77*, 75–85.
61. Obiol-Pardo, C.; Gomis-Tena, J.; Sanz, F.; Saiz, J.; Pastor, M. A multiscale simulation system for the prediction of drug-induced cardiotoxicity. *J. Chem. Inf. Model.* **2011**, *51*, 483–492.
62. Huang, T.C.; Chang, H.Y.; Hsu, C.H.; Kuo, W.H.; Chang, K.J.; Juan, H.F. Targeting therapy for breast carcinoma by ATP synthase inhibitor aurovertin B. *J. Proteome Res.* **2008**, *7*, 1433–1444.
63. Sayers, E.W.; Barrett, T.; Benson, D.A.; Bolton, E.; Bryant, S.H.; Canese, K.; Chetvernin, V.; Church, D.M.; Dicuccio, M.; Federhen, S.; *et al.* Database resources of the national center for biotechnology information. *Nucleic Acids Res.* **2010**, *38*, D5–D16.
64. Hong, S.; Pedersen, P.L. ATP synthase and the actions of inhibitors utilized to study its roles in human health, disease, and other scientific areas. *Microbiol. Mol. Biol. Rev.* **2008**, *72*, 590–641.
65. Barnum, D.; Greene, J.; Smellie, A.; Sprague, P. Identification of common functional configurations among molecules. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 563–571.
66. Smellie, A.; Teig, S.L.; Towbin, P. Poling: promoting conformational variation. *J. Comp. Chem.* **1995**, *16*, 171–187.