

# Teaching and Learning Computational Drug Design: Student Investigations of 3D Quantitative Structure–Activity Relationships through Web Applications

Rino Ragno,<sup>\*</sup> Valeria Esposito,<sup>†</sup> Martina Di Mario,<sup>†</sup> Stefano Masiello,<sup>†</sup> Marco Viscovo,<sup>†</sup> and Richard D. Cramer



Cite This: *J. Chem. Educ.* 2020, 97, 1922–1930



Read Online

ACCESS |



Metrics & More



Article Recommendations



Supporting Information

**ABSTRACT:** The increasing use of information technology in the discovery of new molecular entities encourages the use of modern molecular-modeling tools to help teach important concepts of drug design to chemistry and pharmacy undergraduate students. In particular, statistical models such as quantitative structure–activity relationships (QSAR)—often as its 3D QSAR variant—are commonly used in the development and optimization of a leading compound. We describe how these drug discovery methods can be taught and learned by means of free and open-source web applications, specifically the online platform [www.3d-qsar.com](http://www.3d-qsar.com). This new suite of web applications has been integrated into a drug design teaching course, one that provides both theoretical and practical perspectives. We include the teaching protocol by which pharmaceutical biotechnology master students at Pharmacy Faculty of Sapienza Rome University are introduced to drug design. Starting with a choice among recent articles describing the potencies of a series of molecules tested against a biological target, each student is expected to build a 3D QSAR ligand-based model from their chosen publication, proceeding as follows: creating the initial data set (Py-MolEdit); generating the global minimum conformations (Py-ConfSearch); proposing a promising mutual alignment (Py-Align); and finally, building, and optimizing a robust 3D QSAR models (Py-CoMFA). These student activities also help validate these new molecular modeling tools, especially for their usability by inexperienced hands. To more fully demonstrate the effectiveness of this protocol and its tools, we include the work performed by four of these students (four of the coauthors), detailing the satisfactory 3D QSAR models they obtained. Such scientifically complete experiences by undergraduates, made possible by the efficiency of the 3D QSAR methodology, provide exposure to computational tools in the same spirit as traditional laboratory exercises. With the obsolescence of the classic Comparative Molecular Field Analysis Sybyl host, the 3dqsar web portal offers one of the few available means of performing this well-established 3D QSAR method.

**KEYWORDS:** *Upper-Division Undergraduate, Graduate Education/Research, Continuing Education, Chemoinformatics, Interdisciplinary/Multidisciplinary, Computer-Based Learning, Molecular Modeling, Drugs/Pharmaceuticals, Medicinal Chemistry, 3D QSAR*

## INTRODUCTION

A basic knowledge of pharmaceutical chemistry is one fundamental goal for students in master's degree (MD) courses such as Pharmaceutical Biotechnology (PB, PBMD) or Medicinal Chemistry (MC, MCMD) and Pharmaceutical Technology (Industrial Pharmacy Degree, PT, PTMD). To undertake these courses, students are required to have knowledge about biology, biochemistry, chemistry, pharmacology, and general pathology, normally acquired from introductory courses for medicinal chemistry during a bachelor's degree program. Master's degree courses usually emphasize frontal lectures delivered by teachers with the students' evaluations being written and/or oral student exams. Only a few of them include practical training in the application of theoretical rules and learned knowledge. At Pharmacy and Medicine Faculty of Sapienza University of Rome PBMD (Sapienza PBMD, SPBMD), this traditional approach to teaching and learning is being enriched by increasing the number of practical lessons and by augmenting the final

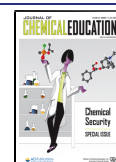
evaluation exam with the student's multimedia presentation, given in a classroom in the presence of the teacher and other students. Although perhaps coincidentally, the number of SPBMD enrolled students has increased during the last five years (Figure 1).

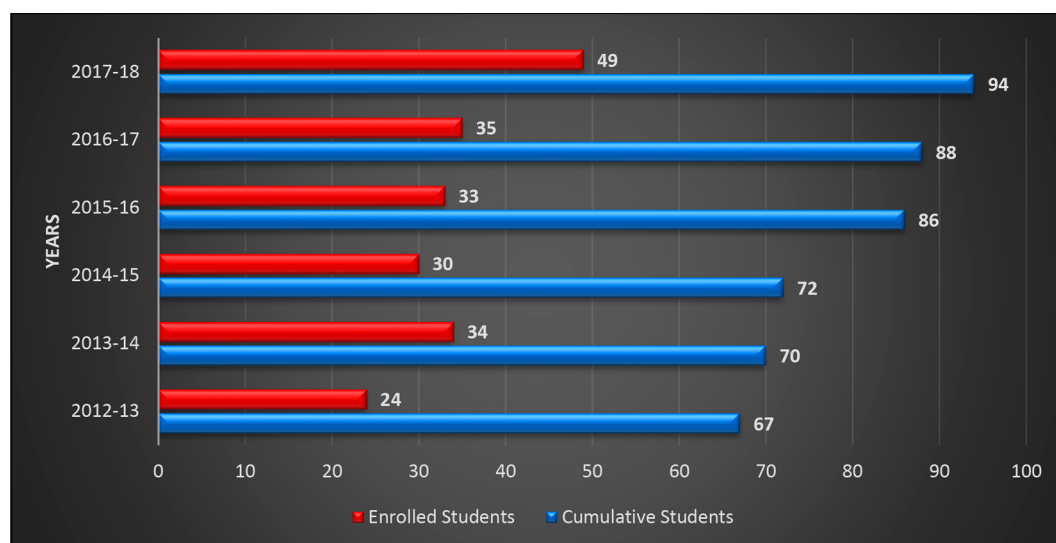
More specifically, as other Italian universities<sup>1–3</sup> and non-Italian universities<sup>4–6</sup> have also been doing, SPBMD has been offering a drug design (DD) course, previously named medicinal chemistry or computational medicinal chemistry. At SPBMD, this DD course has gradually evolved from pure frontal to interactive lectures, increasing the students'

**Received:** February 29, 2020

**Revised:** June 6, 2020

**Published:** June 23, 2020





**Figure 1.** Number of SPBMD enrolled and cumulative students per year. Cumulative students are the total student in the full course (first + second years).

proficiency by exposing them to such computational resources as online databases:

- Protein Data Bank (PDB)<sup>7</sup>
- ChEMBL<sup>8</sup>
- PubChem<sup>9</sup>

Students also gain exposure to specialized computational chemistry software and techniques:

- Multiple linear regression (MLR)<sup>10</sup>
- Principal component analysis (PCA)<sup>11</sup>
- Partial least-squares (PLS)<sup>12,13</sup>
- Quantitative structure–activity relationships (QSAR)<sup>14</sup>
- Three-dimensional QSAR (3D QSAR)<sup>15</sup>
- Pharmacophore modeling (PM)<sup>16</sup>
- Molecular docking (MDock)<sup>17</sup>

### Pedagogic Aspects

This article describes how, in MOOC-like fashion (MOOC: Massive Online Open Courses<sup>18</sup>) with interactive lectures and online platforms, students are learning these DD methods, in particular 3D QSAR, by means of the [www.3d-qsar.com](http://www.3d-qsar.com) portal.<sup>19</sup> While doing so, students also improve their computer skills, in general and within this applied field, which can of course benefit any future career.

From a somewhat different perspective, that is, the latest thinking on the most effective approaches to education, this combination of technology usage with a socio-constructivist approach<sup>20–22</sup> gives students the opportunity to benefit from real learning by doing.<sup>23</sup> The curriculum combines interactive lectures and technical interaction tools that involve them in the construction of a scientific product. We are thus instantiating the so-called triological learning approach<sup>24</sup> (TLA) and the achievement of the knowledge considered as coconstruction mediated by cultural and social artifacts and implemented at the interpersonal level, through communication and interaction with peers and experts.<sup>25</sup> As in a recent definition,<sup>26</sup> in the our TLA approach the students are actively collaborating in developing and creating new science work, as similarly described by Sins and Andreissen.<sup>27</sup>

However, the TLA approach can represent a novel challenge to a student who has been exposed only to traditional

instruction and who may have not had much direct exposure to executables and files. At first the student's feelings are uncertain. On the one hand, the volume of new concepts to be acquired, may be intimidating; yet on the other side there can be a great curiosity that drives them to explore a new way of learning. Among the 50 students that attended the course for the academic year 2017–2018, four of them were particularly interested in its educational program and were therefore invited to participate in preparing and compiling this report, by reporting their fulfillments of these assignments in this work and becoming coauthors of this publication.

Students were also asked about the difficulties encountered in performing the assignment, and their doubts and questions helped to improve the way to present some topics in the frontal lessons. Among the most frequently difficulties was the recognition of the 3-D aspects of molecules and the fact that to develop a 3D QSAR model these 3-D aspects must somehow be aligned. To help overcome some of the difficulties a blog session was created for students to pose written questions to be answered by the teacher, assistants, and also other more skilled colleagues. At the time of writing this article, this blog lists more than 300 answered questions (Italian language) and a blog session has been added to the public [www.3d-qsar.com](http://www.3d-qsar.com)<sup>19</sup> (English language, the blog is accessible only after registering at the site).

### Similar Approaches

Providing students with a fuller understanding of computer-aided drug design (CADD) by the hands-on application of its methods is the goal of several educational programs that have recently been described.<sup>28</sup> Of course, these programs do vary somewhat in their emphases among the large variety of CADD methods that are currently practiced. All start with the nature of the problem: the primality of 3-dimensional shape and flexibility; the multiple properties that an effective drug molecule must have the many repetitive make-and-test cycles. Hands-on instruction begins with the basic and often clunky tools of the trade, today freely available from many sources, to perform: the usage of public databases; the computer representations of chemical structure and shape, and their

means of modification; the generation of canonical 3D shapes from 2D structures.

At this point, these published CADD educational programs diverge. Usually the emphasis is on docking, today's representation, and the attempted quantification and optimization of Ehrlich's foundational "lock-and-key" intuition, specifically of one or two ligands into a single receptor, often with a further "pharmacophoric" goal of attributing an affinity to specific interactions between ligand and receptor atoms. Our 3D QSAR approach thus rather complements docking, as it involves comparisons among many molecular structures, themselves considered as wholes.

Other noteworthy features among these educational programs include the following. Tantillo et al. agree with us in the motivational value of utilizing the exposures to these concepts and tools to complete a DD project.<sup>28</sup> The Computational Structural Biology group at Molecular Modeling Group of the SIB in Lausanne describes an elegant web-based GUI's intended to introduce DD and CADD to a very different audience, the general public, including high school teachers and students.<sup>29</sup> Johnson et al.<sup>30</sup> incorporated pharmacokinetic considerations in their use of lighter computational tools to strengthen a conventional medicinal chemistry course. A Brazilian group<sup>31</sup> included ligand-based approaches and free computational resources in prescribing a receptor-based study of a specific target, COX-1. Sutch et al.<sup>32</sup> gave special attention to pharmacokinetic approaches in the application of various web-available receptor-based tools to the caspase-3 target.

#### DD Course Overview at SPBMD

The SPBMD DD course includes frontal and practical lessons. During the frontal lessons, students are taught about the history of classical and modern medicinal chemistry concepts mainly focused on structure–activity relationship (SAR), from Lipinski rules<sup>33</sup> to current definitions of QSAR, pharmacophores, and molecular docking, and is also given an introduction to pharmacology, emphasizing ADMET (Adsorption, Distribution, Metabolism and Excretion–Toxicity) and thereby pharmacokinetics, with some details about the different macro-groups of drugs. Also included, in preparation for the second part of the coursework, are some theory and usage of the following molecular modeling computer programs:

- ChemAxon MarvinSketch:<sup>34</sup> used to edit and draw molecules
- UCSF Chimera:<sup>35</sup> used to visualize the structure target and in preparation for a docking study
- AutoDock Vina:<sup>36</sup> a Chimera tool, to perform a docking study
- OpenBabel:<sup>37</sup> used to interconvert molecular formats
- [www.3d-qsar.com](http://www.3d-qsar.com):<sup>38</sup> a comprehensive online platform to build 3D QSAR models from scratch

Participating in practical sessions, in a multimedia room, helped the students in the use of personal computers to apply these computational chemistry tools. A series of tutorials were prepared and shared by means of the Moodle platform used in Sapienza University.<sup>39</sup> Basic training in MS Windows, MAC OS, and Linux operating systems and office suites (MS Office and Libreoffice) was also available as needed.

As final test, each student was assigned a project in which all the skills acquired during both the lessons and practical sessions were to be actively applied to the design of new

analogues with improved biological potency (the goal of any actual drug design project). Each student selected a scientific article starting from the current year issue of the *Journal of Medicinal Chemistry* or its ASAP articles.<sup>40</sup> The selected article should include, at minimum, a list of 40 newly synthesized molecules including their affinities for a macromolecular target whose experimental structure was listed in the protein data bank.<sup>7</sup> On their selected article the student should perform the following tasks:

1. Review its context in depth, considering:
  - a. The reason for carrying out the reported research
  - b. The chemical routes used to synthesize the presented compounds
  - c. The biology and biochemistry of the macromolecular target
2. Create models of each listed molecule and tabulate their structures, in SMILES and MOL2 formats
3. Calculate and include a first series of descriptors, in order to
  - a. Evaluate the molecules' "drug-likeness" by means of the Lipinski<sup>41</sup> and Veber<sup>42</sup> rules and attach the calculated data
  - b. Examine the overall physicochemical profile of these molecules, based on the parameters that underly these "drug-like" rules, by means of PCA
  - c. Make QSAR models using only these same parameters, by means of MLR and PLS statistical approaches
4. Use the modeled molecules as a starting point to perform a ligand-based (LB) 3D QSAR study with the [www.3d-qsar.com](http://www.3d-qsar.com) web applications.
5. Use the modeled molecules in a structure-based (SB) application to
  - a. Investigate the binding modes of the molecules by molecular docking
  - b. Obtain a SB molecular alignment for subsequent SB 3D QSAR study
6. Prepare a final report, which includes:
  - a. An electronic document describing the work and its results
  - b. An electronic presentation by means of MS Powerpoint or Libreoffice Impress
7. Make an oral presentation of a maximum 30–40 min

While all of these theoretical background matters are also described in the didactic material of the course, shared and freely available from the Moodle platform at Sapienza University,<sup>39</sup> here the discussion will be focused only on their 3D QSAR aspects, since it seems that elsewhere 3D QSAR and analogous techniques are simply mentioned, without any practical components. Please note that our 3D QSAR facilities and experiences can provide anyone with a fresh template for teaching and learning 3D QSAR, again with no costs whatsoever and without any need to install any specialized software.

The four selected students (coauthors in this publication; V.E., M.D.M., S.M., and M.V.) were among the best students of the 2017–2018 academic year. Among their course (2017–2018), they were among the first to take the master's degree and all of them earned the maximum degree rank cum laude. Also, their reports were among the best then reported, although with the maturation of the course and with higher



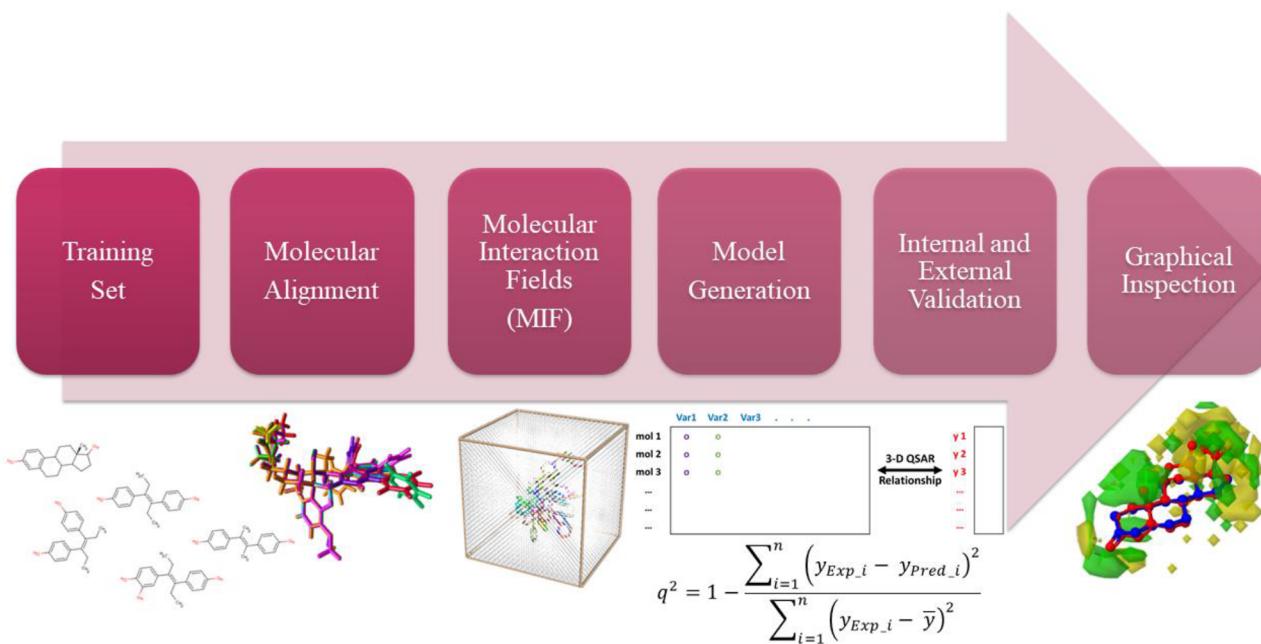


Figure 2. Workflow of a FB 3D QSAR procedure.

reference many other students of the same academic year and many others among the following one (2018–2019) performed just as well. The high quality of the 3D QSAR models themselves was not considered a requisite to pass the exam, as not all data sets can yield satisfactory models. Instead the quality of a final presentation allowed the examining committee to evaluate the student's understanding and performance. Only a few percent of students actually failed to pass the exam (4%) while about 16% passed with the maximum rank (30/30 cum laude), about 40% were very good (28–30/30), about 30 gained an average rank (24–27/30) and the last 10% had a low rank ((18–23/30).

### ■ A BRIEF INTRODUCTION TO 3D QSAR

Many reviews and a few books have been published on 3D QSAR.<sup>15,43–47</sup> While other 3D QSAR methods are known, with some discussed below, the field-based (FB) 3D QSARs (FB 3D QSAR) were the first introduced and are by far the most practiced, starting with the first Cramer et al. comparative molecular field analysis (CoMFA) published article,<sup>47</sup> and with little subsequent enhancement of CoMFA beyond Cramer's developments of "topomers"<sup>48,49</sup> and "template CoMFA"<sup>50–52</sup> as simple yet robust options for producing ligand-based (LB) alignments (in general drug design nomenclature, "Field-based" (FB) approaches are a major subset of "ligand-based" (LB) approaches, with LB in turn to be contrasted with the set of "receptor-based" or "structure-based" approaches).

To summarize the CoMFA methodology, the set of 3-D aligned molecular conformations is virtually embedded into a 3-D lattice spanning at least 5 Å, the minimum volume typically required to enclose all the molecules. At each grid point a hypothetical probe atom, typically an sp<sup>3</sup> carbon atom bearing a positive charge, is placed, and then its Lennard-Jones and Coulomb law potentials with each of the atoms in every molecule are accumulated, collectively comprising the respective steric and electrostatic "interaction fields (MIF)" for each molecule. This result can be visualized as a structure/activity table or matrix with far more columns (lattice point

count × 2), than rows (structures)—a severely under-constrained base for conventional model derivation! Happily, by treating this matrix as a whole rather than as a collection of independently manipulable descriptor columns, the PLS algorithm coupled with cross-validation can yield a robust mathematical model from these MIFs, although one described by a long unreadable regression equation containing thousands of terms

$$y = \sum_{i=1}^n (x_i \times \text{PLScoeff}_i)$$

where  $y$  is the calculated or predicted response;  $x_i$  are the field values; and  $\text{PLScoeff}_i$  are the PLS correlation coefficients.

Yet the graphical representation of  $\text{PLScoeff}_i$  (after multiplication by each node point's average value) yields very informative polyhedron whose interpretation can aid in the design of new molecules.

### Steps for the FB 3D QSAR Procedure

Carrying out a FB 3D QSAR model involves the following steps, as shown in Figure 2.

**Data Set Compilation.** Selection of a molecular data set with associated biological activities or properties (responses) which do not need to share a structural scaffold. It is recommended that biological activities, when as typically expressed in concentrations ( $\text{IC}_{50}$ ,  $\text{EC}_{50}$ ,  $K_i$ ), are expressed by the " $p = -\log()$  (logistic) function", with the negative sign associating positive model coefficients with improved activities. To ensure that the differences among the activities exceed random experimental variabilities, the range of these logistic values should span at least 2 log values. Many authors are actually indicating a range of 4–5 logs to get good QSAR models, but it is not always possible to have such a range of activity.

To permit an estimate of the predictive ability of an ensuing model, the intended conclusion with CoMFA, the data set can also be split into a training (model building) and a test (model

evaluation) set. To develop such a reliable model a minimum number of 15–20 molecules for the training set is desirable.

**Generation of 3-D Conformations.** If the data set molecules are available only in SMILES format, their “flat” representations must first be converted into 3-D conformations. Appropriate tools include graphical sketch-based (UCSF Chimera<sup>55</sup>) or command line methods (OpenBabel<sup>53</sup>) or free web services (ochem.eu).<sup>54</sup>

**Definition of Alignment Rules.** This is the most critical step. Its goal is to superimpose (align) each of the molecules so that the differences among their atoms’ identities and positions represent the resulting field differences in a way that then yields a satisfactory CoMFA model. Different alignment strategies can be adopted.<sup>55</sup> Many articles have reported several approaches to obtain the aligned training set through automatic alignment programs<sup>56</sup> or by atom-by-atom superposition of the maximum common substructure<sup>57</sup> or by using pharmacophore modeling.<sup>58</sup> One of us (Cramer) has found that alignment rules, such as “topomers” and “template CoMFA”, which prefer the field differences intentionally introduced by chemical synthesis to those imposed by physicochemical realities, often easily generate remarkably powerful and versatile 3D-QSAR models. However, with the demise of classical CoMFA’s Sybyl host but the continuance of relevant patents, these alignment methods are unfortunately available only as crude, unsupported, and almost unvalidated open source codes that also require the acquisition and installation of underlying commercial software.

**Calculation of MIF.** The aligned molecules are virtually placed in a grid of proper dimension and the MIF are calculated.

**Model Generation.** At this point, all the data needed to run the PLS and generate the models for a given number of extracted principal components have been assembled. The goodness of the models is evaluated through the squared correlation coefficient  $r^2$  calculated by

$$r^2 = 1 - \frac{\sum_{i=1}^n (y_{\text{exp},i} - y_{\text{calc},i})^2}{\sum_{i=1}^n (y_{\text{exp},i} - \bar{y})^2}$$

**Model Validation.** The best model is selected by means of

- **Robustness:** Cross-validation that indicates the most robust model on the basis of an internal predictive coefficient called  $q^2$  evaluated by the following equation:

$$q^2 = 1 - \frac{\sum_{i=1}^n (y_{\text{exp},i} - y_{\text{pred},i})^2}{\sum_{i=1}^n (y_{\text{exp},i} - \bar{y})^2}$$

- **Lack of Chance Correlation:** Even a robust model could have been obtained by chance. A further test is the “Y-scrambling” (Y-S) method, by randomly reassigning the experimental potencies to the compounds. Upon rederivation of this now clearly invalid model, the obtained  $r_{Y-S}^2$  and  $q_{Y-S}^2$  values should be always lower than those of the original model.
- **Predictive Ability:** In the case of availability of additional molecules external to this model-building set (perhaps by subsequent synthesis), it is the usual practice in CoMFA to evaluate the predictive ability of its final model, by calculating the standard deviation error of prediction (SDEP), and possibly computing an  $r_{\text{pred}}^2$  ( $r^2$  calculated using the predicted and experimental

responses of the test set). The calculation of SDEP can be easily achieved with the standard deviation equation:

$$\text{SDEP} = \frac{\sqrt{\sum_{i=1}^n (y_{\text{exp},i} - y_{\text{pred},i})^2}}{n - 1}$$

**Graphical Interpretation.** A series of plots are generated to inspect and describe the model and design new derivatives on the basis of the observation of the plots superimposed on the training set molecules.

### Field-Based Instructional Resources

As mentioned above, 3D QSAR was introduced by Cramer et al. in 1988,<sup>47</sup> with the name of CoMFA as an acronym for “Comparative Molecular Field Analysis”. Nevertheless, despite the application of 3D QSAR as CoMFA in tens of thousands of articles and many books, we can find no current evidence of any practical presentation within a DD academic MD course. CoMFA had been available only in the Tripos Suites (aka Sybyl), in a very friendly form so easy to use that with a few “mouse clicks” almost any user was able to develop their own 3D QSAR model, and was well documented in the CoMFA manual. Nevertheless some IT knowledge and costs were required to obtain, install and run the software. Although other 3D QSAR software appeared on the scene, the original CoMFA patent prevented its commercialization by any other software house until its expiration in 2011. Possibly the further development of 3D QSAR was inhibited; 4D QSAR,<sup>59</sup> 5D QSAR,<sup>60</sup> and 6D QSAR<sup>61</sup> appeared but few scientific applications followed. However, the Cresset Group now commercially offers a field-based nonlattice 3D QSAR package.<sup>62</sup> In 1994, Cruciani et al. introduced GOLPE,<sup>63</sup> a chemometric tool that in combination with Goodford’s GRID provided highly reliable 3D QSAR models. In 2011, after the CoMFA patent expired, Open3DQSAR,<sup>64</sup> was announced, the first free and open source program that with the companion softwares Open3DALIGN<sup>65</sup> and Open3DGRID ([open3dgrid.sourceforge.net](http://open3dgrid.sourceforge.net)) provided all the steps in 3D QSAR. However, a web-based implementation of Open3DQSAR in a Web site<sup>66</sup> lacked model building and graphical analysis. In 2012, 3D-QSAutogrid/R,<sup>67</sup> a second open source software based 3D QSAR procedure, was reported in which the MIFs were calculate by means of the AutoDock suite AutoGrid program<sup>68</sup> and the statistical calculations (PLS, cross-validation and Y-scrambling) were performed by means of the R-CRAN pls package.<sup>69</sup> The only other CoMFA source, an open source Open Eye-based code, was mentioned above.

Thus, although these other tools for building other types of FB 3D QSAR models exist, it is evident that, today, well-validated classic CoMFA 3D QSAR models can easily be created at no cost with Open3DQSAR and via the readily accessible and [www.3d-qsar.com](http://www.3d-qsar.com) portal.<sup>19</sup>

## DEVELOPING 3D QSAR MODELS BY MEANS OF THE QSAR WEB PORTAL

### Overview of the Web Portal Procedure

Here is an overview of this portal’s workflow, serving also as background for the following summaries of the four student projects. The scientific and technical rationales underlying each of its six successive stages have already been summarized.

**Data Set Compilation through the Web Portal.** At first, an empty data set container, a sort of work folder or

Table 1. Comparison of 3D QSAR Models Metrics<sup>a</sup>

target	N	fields	ONPC <sup>b</sup>	r <sup>2</sup>	SDEC <sup>c</sup>	q <sup>2</sup>		SDEP <sup>d</sup>		r <sup>2</sup> YS	q <sup>2</sup> YS		settings				
						LOO <sup>e</sup>	LSO <sup>f</sup>	LOO <sup>e</sup>	LSO <sup>f</sup>		LOO <sup>e</sup>	LSO <sup>f</sup>	PA <sup>g</sup>	GS <sup>h</sup>	GE <sup>i</sup>	MS <sup>j</sup>	C <sup>k</sup>
IDO1	51	Ste <sup>l</sup>	2	0.54	0.77	-0.06	-0.06	1.17	1.17	0.52	-0.27	-0.28	C.2	1.312	5	0.05	15
		Ele <sup>m</sup>	8	0.98	0.13	0.47	0.40	0.84	0.87	0.88	-1.45	-0.94					
		Both <sup>n</sup>	8	0.98	0.12	0.15	0.19	1.04	1.02	0.94	-0.36	-0.42					
SHP-2	40	Ste <sup>l</sup>	4	0.84	0.31	0.41	0.41	0.61	0.61	0.76	-0.67	-0.46	H	2.200	10	1.50	25
		Ele <sup>m</sup>	1	0.18	0.71	0.12	0.12	0.74	0.74	0.05	-0.12	-0.10					
		Both <sup>n</sup>	4	0.78	0.36	0.27	0.29	0.67	0.66	0.61	-0.61	-0.51					
IRAK4	58	Ste <sup>l</sup>	6	0.98	0.13	0.29	0.31	0.94	0.93	0.96	-0.30	-0.29	O.3	1.000	5	2.00	25
		Ele <sup>m</sup>	4	0.83	0.46	0.21	0.26	0.99	0.96	0.62	-0.36	-0.47					
		Both <sup>n</sup>	6	0.97	0.16	0.44	0.45	0.83	0.83	0.93	-0.30	-0.27					
BRD4	45	Ste <sup>l</sup>	2	0.69	0.40	0.28	0.31	0.61	0.60	0.55	-0.64	-0.39	H	2.200	5	0.05	25
		Ele <sup>m</sup>	2	0.31	0.60	0.04	0.08	0.71	0.69	0.22	-0.39	-0.23					
		Both <sup>n</sup>	4	0.88	0.24	0.54	0.54	0.49	0.48	0.74	-0.67	-0.36					

<sup>a</sup>Note: For further details, see the [Supporting Information](#). <sup>b</sup>ONPC: Optimal number of principal components. <sup>c</sup>SDEP: Cross-validated standard deviation error prediction. <sup>d</sup>SDEC: Standard deviation error calculation. <sup>e</sup>LOO: Leave-one-out. <sup>f</sup>LSO: Leave-some-out. <sup>g</sup>PA: Probe atom. <sup>h</sup>GS: Grid step. <sup>i</sup>GE: Grid extension. <sup>j</sup>MS: Minimum sigma. <sup>k</sup>C: Max/min energy of cutoff value. <sup>l</sup>Ste: Steric MIF. <sup>m</sup>Ele: Electrostatic MIF. <sup>n</sup>Both: Steric and electrostatic fields.

virtual spreadsheet, is prepared by means of the Py-MolEdit web application. Molecules may be added to this data set in several different ways. For these projects 3D structures were created with MarvinSketch and uploaded to the data set container using the “Add Multiple Molecules” command. A separately prepared file added biological activities to these data set molecules. (See the Py-MolEdit tutorial link in the [Supporting Information](#).)

**Generation of 3-D Conformations through the Web Portal.** A ligand-based (LB) conformation analysis was performed with the Py-ConfSearch module on all of the molecules in this data set. Its goal is to explore the conformational space of each molecule and provide an idea of their flexibility, with a family of conformations being saved for every flexible molecule. (See the Py-ConfSearch tutorial link in the [Supporting Information](#).)

**Definition of Alignment Rules through the Web Portal.** Single conformations, selected from the conformation family for each of the molecules are aligned by means of the Py-Align application, forming one training set. The Py-Align application provides 16 alignment approaches, with the expectation that many different training sets, conformational analysis and alignment complex, will be considered from each data set. (See the Py-Align tutorial link in the [Supporting Information](#).)

**Calculation of MIF Fields and Model Generation through the Web Portal.** For each such aligned training set, a 3D QSAR model can be generated using the Py-CoMFA module<sup>70</sup> with the default settings, which can then be further explored by varying the probe type, the grid spacing, the min-max cutoff energy and the minimum sigma. To evaluate model robustness, cross-validation is done as the model is built. Other tests for chance correlation, as discussed above, can be performed on the completed model. (See the Py-CoMFA tutorial link in the [Supporting Information](#).)

**Model Validation and Prediction through the Web Portal.** Since the main goal of any 3D QSAR model is to predict unknown molecules activities by means of their structures and the alignment rules, an external test set is usually prepared along with the training set. The aim of external validation is to verify the predictive capacity of the model. Including this capability can be made part of the

derivation of the 3D QSAR model, by random selection from the test set molecules in the data set container.

**Graphical Interpretation through the Web Portal.** Still within the Py-CoMFA application, once the model has been built and validated the standard CoMFA 3-D map can be created and inspected directly through the web browser without the use of any supplementary software. The polyhedron images can generated for an individual molecule's MIF as well as for the entire 3D QSAR model.

#### Protocol Most Commonly Followed by the Students

**Target Assignment.** As described above (see also DD course overview at SPBMD) the students were each assigned a *J. Med. Chem.*<sup>71</sup> article describing more than 40 molecules for a given biological target. The four targets chosen were: indoleamine 2,3-dioxygenase 1 (IDO1),<sup>72,73</sup> tyrosine-protein phosphatase nonreceptor type II (SHP-2),<sup>74,75</sup> interleukin-1 receptor associated kinase-4 (IRAK4),<sup>76,77</sup> and bromodomain-containing protein 4 (BRD4).<sup>78</sup>

**General Procedures for Initial Molecular Modeling of Ligands.** All of the molecules listed in the target articles were drawn and edited through ChemAxon, MarvinSketch, or the Py-MolEdit web application available on the Web site [www.3d-qsar.com](http://www.3d-qsar.com).<sup>19</sup> (For use, see the [Supporting Information](#).) With ChemAxon, the molecules must be downloaded from another Web site, and the direct editing on Py-MolEdit saves the molecule in the database. Conformation analysis was performed on each of the four completed data sets (Py-ConfSearch), specifically using the Balloon<sup>79</sup> method, and an arbitrary number of conformations for each molecule, specifically 70, were aligned by Py-Align. Py-Align provides 16 possibilities from a training set as templates for alignment, such as the most active, the less active, or the most flexible. Once the template is selected, Py-Align provides several options for automatic alignment; for most projects the option chosen was ShaEP<sup>56</sup> with the default settings. From the result of each training set alignment, a 3D QSAR model was generated.

**General Procedure for the 3D QSAR Models' Development.** The 3D QSAR Ligand-Based approach was performed with the same default settings for all listed targets. Therefore, for each one a conformational analysis was carried out, various alignment rules were applied and finally, several



models were generated. Then, these models were optimized, modifying the settings as following:

- *Probes atom*: C.3, C.2, C.cat, O.3, N.3, H
- *Grid spacing*: range 1–3 with 0.1 unit of difference
- *Grid extension*: range 5–10 with 1 unit of difference
- *Min/max cutoff energy*: range 20–40 with 5 units of difference
- *Minimum Sigma*: range 0.05–2 with 0.05 units of difference

A total of about 80 3D QSAR models were built for each target, and for the one yielding the highest  $r^2$  and  $q^2$  values, both steric and electrostatic contours maps were generated and analyzed through the graphical plots implemented within [www.3d-qsar.com](http://www.3d-qsar.com).<sup>19</sup>

### STUDENT WORK EXAMPLES

The students applied the above protocol to their selected *J. Med. Chem.* articles to generate ligand-based aligned FB 3D QSAR models by means of the [www.3d-qsar.com](http://www.3d-qsar.com) portal<sup>19</sup> (Table 1). Considering all four of the projects, a total of 194 inhibitors were modeled and more than 320 3D QSAR models were developed and internally validated for fitting, robustness, and lack of chance correlation. External validations were also performed in three of the four projects, yielding low errors of predictions (see the Supporting Information). Each of the four final models were further analyzed by means of classical CoMFA contour maps revealing the possibility for future drug design.

### CONCLUSIONS

In this publication, we believe we have demonstrated the effectiveness and the convenience of the free Web site platform [www.3d-qsar.com](http://www.3d-qsar.com) as a tool to teach 3D QSAR in DD courses, including the difficult underlying chemical concepts. Our experiences suggest that the 3D QSAR approach is quite appropriate for undergraduate students of pharmaceutical biotechnology, and more generally as a part of bachelor, master, and Ph.D. degree programs, and perhaps even within high school courses. Particularly relevant evidence for this view could be the work performed by the four students from the SPBMD course during the academic year 2017–2018, who constructively criticized and then employed four modules of the [www.3d-qsar.com](http://www.3d-qsar.com) website<sup>19</sup> (Py-MolEdit, Py-ConfSearch, Py-Align, Py-CoMFA) to complete a ligand-based 3D QSAR study. This experience helped these students to obtain a high score in the drug design exam, increasing their biological computational skills, and encouraging them to continue these kind of studies.

To also provide students with some experience in receptor-based methodologies, since the targets of the four publications used for this study are available in crystallized form in the online database PDB (protein data bank),<sup>7</sup> another study will describe the use of these targets to generate alternative molecular alignments for 3D QSAR. Also, video tutorial are planned for release soon to provide even better guides to develop 3D QSAR models.

In conclusion, beyond the value of the students' models themselves, their work clearly demonstrates that by using the protocol we describe they learned many skills and perspectives in computational medicinal chemistry, as was the main objective of the course.

### ASSOCIATED CONTENT

#### Supporting Information

The Supporting Information is available at <https://pubs.acs.org/doi/10.1021/acs.jchemed.0c00117>.

Details of student work examples with tutorial links for interacting at the 3D QSAR Web site (PDF)

### AUTHOR INFORMATION

#### Corresponding Author

**Rino Ragno** – Rome Center for Molecular Design, Department of Drug Chemistry and Technology, Sapienza Rome University, 00185 Rome, Italy; [orcid.org/0000-0001-5399-975X](https://orcid.org/0000-0001-5399-975X); Email: [rino.ragno@uniroma1.it](mailto:rino.ragno@uniroma1.it)

#### Authors

**Valeria Esposito** – Pharmacy and Medicine Faculty, Pharmaceutical Biotechnology Master Degree Course, Sapienza Rome University, 00185 Rome, Italy

**Martina Di Mario** – Pharmacy and Medicine Faculty, Pharmaceutical Biotechnology Master Degree Course, Sapienza Rome University, 00185 Rome, Italy

**Stefano Masiello** – Pharmacy and Medicine Faculty, Pharmaceutical Biotechnology Master Degree Course, Sapienza Rome University, 00185 Rome, Italy

**Marco Viscovo** – Pharmacy and Medicine Faculty, Pharmaceutical Biotechnology Master Degree Course, Sapienza Rome University, 00185 Rome, Italy

**Richard D. Cramer** – Retired, Santa Fe, New Mexico 87501, United States

Complete contact information is available at: <https://pubs.acs.org/doi/10.1021/acs.jchemed.0c00117>

#### Author Contributions

<sup>†</sup>V.E., M.D.M., S.M., and M.V. contributed equally to this work.

#### Notes

The authors declare no competing financial interest.

### ACKNOWLEDGMENTS

Many thanks to Professor F. Altieri, chair of Pharmaceutical Biotechnology MD, for providing us with the data about university and students entries from 2011 to 2018. We also thank the Department of Pharmaceutical Chemistry and Technology, Faculty of Pharmacy and Medicine, Sapienza University of Rome. Many thanks are also owed to the Ph.D. students Manuela Sabatino, Alexandros Patsilinos, and Federica Micale for their teaching assistance. In addition, we thank Alessio Ragno and Dylan Savoia, two software engineering and robotics master degree students who supported the development of [www.3d-qsar.com](http://www.3d-qsar.com). This work was supported by PRIN 2017 (prot. 2017JL8SRX) (to RR), Ateneo 2019 (prot. RM11916B8876093E) (to RR) and Ateneo 2018 (prot. RM118164361B425B) (to RR).

### REFERENCES

- (1) University of Siena, Italy. <http://www.unisi.it> (accessed June 2019).
- (2) University of Milan, Italy. <http://www.unimi.it> (accessed June 2019).
- (3) University of Padua, Italy. <http://www.unipd.it> (accessed June 2019).

- (4) Ulm University, Germany. <https://www.uni-ulm.de> (accessed June 2019).
- (5) Department of Biological Sciences, Columbia University, New York. <https://biology.columbia.edu> (accessed June 2019).
- (6) University of California Los Angeles, California. <https://biology.columbia.edu> (accessed June 2019).
- (7) Burley, S. K.; Berman, H. M.; Kleywegt, G. J.; Markley, J. L.; Nakamura, H.; Velankar, S. Protein Data Bank (PDB): The Single Global Macromolecular Structure Archive. *Methods Mol. Biol.* **2017**, *1607*, 627–641.
- (8) Nowotka, M. M.; Gaulton, A.; Mendez, D.; Bento, A. P.; Hersey, A.; Leach, A. Using ChEMBL web services for building applications and data processing workflows relevant to drug discovery. *Expert Opin. Drug Discov.* **2017**, *12* (8), 757–767.
- (9) Li, Q.; Cheng, T.; Wang, Y.; Bryant, S. H. PubChem as a public resource for drug discovery. *Drug Discovery Today* **2010**, *15* (23–24), 1052–7.
- (10) Šoškić, M.; Plavšić, D.; Trinajstić, N. Link between Orthogonal and Standard Multiple Linear Regression Models. *J. Chem. Inf. Model.* **1996**, *36* (4), 829–832.
- (11) Balsera, M. A.; Wriggers, W.; Oono, Y.; Schulten, K. Principal component analysis and long time protein dynamics. *J. Phys. Chem.* **1996**, *100* (7), 2567–2572.
- (12) Wold, S.; Ruhe, A.; Wold, H.; Dunn, W. J., III The Collinearity Problem in Linear Regression. The Partial Least Squares (PLS) Approach to Generalized Inverses. *SIAM J. Sci. and Stat. Comput.* **1984**, *5* (3), 735–743.
- (13) Martin, Y. C.; Lin, C. T.; Hetti, C.; DeLazzer, J. PLS analysis of distance matrices to detect nonlinear relationships between biological potency and molecular properties. *J. Med. Chem.* **1995**, *38* (16), 3009–15.
- (14) Gao, H.; Williams, C.; Labute, P.; Bajorath, J. Binary quantitative structure-activity relationship (QSAR) analysis of estrogen receptor ligands. *J. Chem. Inf. Comput. Sci.* **1999**, *39* (1), 164–8.
- (15) Melo-Filho, C. C.; Braga, R. C.; Andrade, C. H. 3D-QSAR approaches in drug design: perspectives to generate reliable CoMFA models. *Curr. Comput.-Aided Drug Des.* **2014**, *10* (2), 148–59.
- (16) Yang, S. Y. Pharmacophore modeling and applications in drug discovery: challenges and recent advances. *Drug Discovery Today* **2010**, *15* (11–12), 444–50.
- (17) de Ruyck, J.; Brysbaert, G.; Blossey, R.; Lensink, M. F. Molecular docking as a popular tool in drug design, an in silico travel. *Adv. Appl. Bioinform Chem.* **2016**, *9*, 1–11.
- (18) Leontyev, A.; Baranov, D. Massive Open Online Courses in Chemistry: A Comparative Overview of Platforms and Features. *J. Chem. Educ.* **2013**, *90* (11), 1533–1539.
- (19) 3D QSAR Portal. <https://www.3d-qsar.com> (accessed March 2019).
- (20) Engeström, Y., Activity theory and individual and social transformation. In *Perspectives on activity theory*; Cambridge University Press: New York, 1999; pp 19–38.
- (21) Scardamalia, M.; Bereiter, C. *Knowledge building: Theory, pedagogy, and technology* **2005**, 97.
- (22) Porath, M. The Culture of Education. *Canadian Journal of Education/Revue canadienne de l'éducation* **2000**, *25* (3), 236–238.
- (23) Anzai, Y.; Simon, H. A. The theory of learning by doing. *Psychological Review* **1979**, *86* (2), 124–140.
- (24) Paavola, S.; Hakkarainen, K. The knowledge creation metaphor - An emergent epistemological approach to learning. *Science and Education* **2005**, *14*, 535.
- (25) Topçiu, M.; Myftiu, J. Vygotsky Theory on Social Interaction and its Influence on the Development of Pre-School Children. *European Journal of Social Sciences Education and Research* **2015**, *4*, 172.
- (26) Paavola, S.; Engeström, R.; Hakkarainen, K., The Triological Approach as a New form of Mediation. In *Collaborative Knowledge Creation: Practices, Tools, Concepts*; Moen, A., Mørch, A. I., Paavola, S., Eds.; SensePublishers: Rotterdam, 2012; pp 1–14.
- (27) Sins, P.; Andriessen, J., Working within Knowledge Communities as a Context for Developing Knowledge Practices. In *Collaborative Knowledge Creation: Practices, Tools, Concepts*; Moen, A., Mørch, A. I., Paavola, S., Eds.; SensePublishers: Rotterdam, 2012; pp 233–248.
- (28) Tantillo, D. J.; Siegel, J. B.; Saunders, C. M.; Palazzo, T. A.; Painter, P. P.; O'Brien, T. E.; Nuñez, N. N.; Nouri, D. H.; Lodewyk, M. W.; Hudson, B. M.; Hare, S. R.; Davis, R. L. Computer-Aided Drug Design for Undergraduates. *J. Chem. Educ.* **2019**, *96* (5), 920–925.
- (29) Daina, A.; Blatter, M.-C.; Baillie Gerritsen, V.; Palagi, P. M.; Marek, D.; Xenarios, I.; Schwede, T.; Michielin, O.; Zoete, V. Drug Design Workshop: A Web-Based Educational Tool To Introduce Computer-Aided Drug Design to the General Public. *J. Chem. Educ.* **2017**, *94* (3), 335–344.
- (30) Johnson, A. T.; Prospective Method, A. To Guide Small Molecule Drug Design. *J. Chem. Educ.* **2015**, *92* (5), 836–842.
- (31) Rodrigues, R. P.; Andrade, S. F.; Mantoani, S. P.; Eifler-Lima, V. L.; Silva, V. B.; Kawano, D. F. Using Free Computational Resources To Illustrate the Drug Design Process in an Undergraduate Medicinal Chemistry Course. *J. Chem. Educ.* **2015**, *92* (5), 827–835.
- (32) Sutch, B. T.; Romero, R. M.; Neamati, N.; Haworth, I. S. Integrated Teaching of Structure-Based Drug Design and Biopharmaceutics: A Computer-Based Approach. *J. Chem. Educ.* **2012**, *89* (1), 45–51.
- (33) Lipinski, C. A. Lead- and drug-like compounds: the rule-of-five revolution. *Drug Discovery Today: Technol.* **2004**, *1* (4), 337–41.
- (34) Chemaxon. <https://chemaxon.com> (accessed March 2019).
- (35) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **2004**, *25* (13), 1605–12.
- (36) Trott, O.; Olson, A. J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **2009**, *31* (2), 455–61.
- (37) O'Boyle, N. M.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.; Hutchison, G. R. Open Babel: An open chemical toolbox. *J. Cheminf.* **2011**, *3*, 33.
- (38) Ragno, R. [www.3d-qsar.com](http://www.3d-qsar.com): A Portal to Build 3D QSAR Models. *Proceedings* **2019**, *22*, 76.
- (39) E-Learning Sapienza. <https://elearning.uniroma1.it> (accessed March 2019).
- (40) Journal of Medicinal Chemistry Ahead of Print. <https://pubs.acs.org/toc/jmcmr/0/0> (accessed March 2019).
- (41) Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Delivery Rev.* **1997**, *23* (1–3), 3–25.
- (42) Veber, D. F.; Johnson, S. R.; Cheng, H. Y.; Smith, B. R.; Ward, K. W.; Kopple, K. D. Molecular properties that influence the oral bioavailability of drug candidates. *J. Med. Chem.* **2002**, *45* (12), 2615–2623.
- (43) Kellogg, G. E.; Semus, S. F. 3D QSAR in modern drug design. *EXS* **2003**, *93*, 223–41.
- (44) Kubinyi, H. QSAR and 3D QSAR in drug design 0.2. Applications and problems. *Drug Discovery Today* **1997**, *2* (12), 538–546.
- (45) Kubinyi, H. QSAR and 3D QSAR in drug design 0.1. methodology. *Drug Discovery Today* **1997**, *2* (11), 457–467.
- (46) Verma, J.; Khedkar, V. M.; Coutinho, E. C. 3D-QSAR in drug design—a review. *Curr. Top. Med. Chem.* **2010**, *10* (1), 95–115.
- (47) Cramer, R. D.; Patterson, D. E.; 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* (18), 5959–67.
- (48) Jilek, R. J.; Cramer, R. D. Topomers: A Validated Protocol for Their Self-Consistent Generation. *J. Chem. Inf. Comput. Sci.* **2004**, *44* (4), 1221–1227.



- (49) Cramer, R. D. Topomer CoMFA: a design methodology for rapid lead optimization. *J. Med. Chem.* **2003**, *46* (3), 374–88.
- (50) Cramer, R. D.; Wendt, B. Template CoMFA: the 3D-QSAR Grail? *J. Chem. Inf. Model.* **2014**, *54* (2), 660–71.
- (51) Cramer, R. D. Template CoMFA applied to 116 biological targets. *J. Chem. Inf. Model.* **2014**, *54* (7), 2147–56.
- (52) Cramer, R. D. Template CoMFA Generates Single 3D-QSAR Models that, for Twelve of Twelve Biological Targets, Predict All ChEMBL-Tabulated Affinities. *PLoS One* **2015**, *10* (6), e0129307.
- (53) O'Boyle, N. M.; Morley, C.; Hutchison, G. R. Pybel: a Python wrapper for the OpenBabel cheminformatics toolkit. *Chem. Cent. J.* **2008**, *2*, 5.
- (54) Online Chemical Database. <https://ochem.eu> (accessed March 2019).
- (55) Lemmen, C.; Lengauer, T. Computational methods for the structural alignment of molecules. *J. Comput.-Aided Mol. Des.* **2000**, *14* (3), 215–32.
- (56) Vainio, M. J.; Puranen, J. S.; Johnson, M. S. ShaEP: molecular overlay based on shape and electrostatic potential. *J. Chem. Inf. Model.* **2009**, *49* (2), 492–502.
- (57) Lesk, A. M. Extraction of well-fitting substructures: root-mean-square deviation and the difference distance matrix. *Folding Des.* **1997**, *2* (3), S12–4.
- (58) Seidel, T.; Schuetz, D. A.; Garon, A.; Langer, T. The Pharmacophore Concept and Its Applications in Computer-Aided Drug Design. *Prog. Chem. Org. Nat. Prod* **2019**, *110*, 99–141.
- (59) Hopfinger, A. J.; Wang, S.; Tokarski, J. S.; Jin, B.; Albuquerque, M.; Madhav, P. J.; Duraiswami, C. Construction of 3D-QSAR Models Using the 4D-QSAR Analysis Formalism. *J. Am. Chem. Soc.* **1997**, *119* (43), 10509–10524.
- (60) Vedani, A.; Dobler, M. 5D-QSAR: the key for simulating induced fit? *J. Med. Chem.* **2002**, *45* (11), 2139–49.
- (61) Vedani, A.; Dobler, M.; Lill, M. A. Combining Protein Modeling and 6D-QSAR. Simulating the Binding of Structurally Diverse Ligands to the Estrogen Receptor. *J. Med. Chem.* **2005**, *48* (11), 3700–3703.
- (62) Cresset. <https://www.cresset-group.com> (accessed March 2019).
- (63) Cruciani, G.; Watson, K. A. Comparative molecular field analysis using GRID force-field and GOLPE variable selection methods in a study of inhibitors of glycogen phosphorylase b. *J. Med. Chem.* **1994**, *37* (16), 2589–601.
- (64) Tosco, P.; Balle, T. Open3DQAR: a new open-source software aimed at high-throughput chemometric analysis of molecular interaction fields. *J. Mol. Model.* **2011**, *17* (1), 201–8.
- (65) Tosco, P.; Balle, T.; Shiri, F. Open3DALIGN: an open-source software aimed at unsupervised ligand alignment. *J. Comput.-Aided Mol. Des.* **2011**, *25* (8), 777–83.
- (66) 3D Qserver. <http://bifuohyd.ac.in/qserver/> (accessed March 2019).
- (67) Ballante, F.; Ragno, R. 3-D QSAutogrid/R: an alternative procedure to build 3D QSAR models. Methodologies and applications. *J. Chem. Inf. Model.* **2012**, *52* (6), 1674–85.
- (68) Goodsell, D. S.; Morris, G. M.; Olson, A. J. Automated docking of flexible ligands: applications of AutoDock. *J. Mol. Recognit.* **1996**, *9* (1), 1–5.
- (69) Mevik, B.-H.; Wehrens, R. The pls Package: Principal Component and Partial Least Squares Regression in R. *Journal of Statistical Software* **2007**, *18*, 2007.
- (70) Ragno, R. [www.3d-qsar.com](http://www.3d-qsar.com): a web portal that brings 3D QSAR to all electronic devices—the Py-CoMFA web application as tool to build models from pre-aligned datasets. *J. Comput.-Aided Mol. Des.* **2019**, *33*, 855.
- (71) The Journal of Medicinal Chemistry. <https://pubs.acs.org/journal/jmcmr> (accessed March 2019).
- (72) Weng, T.; Qiu, X.; Wang, J.; Li, Z.; Bian, J. Recent discovery of indoleamine-2,3-dioxygenase 1 inhibitors targeting cancer immunotherapy. *Eur. J. Med. Chem.* **2018**, *143*, 656–669.
- (73) Rohrig, U. F.; Majjigapu, S. R.; Vogel, P.; Zoete, V.; Michielin, O. Challenges in the Discovery of Indoleamine 2,3-Dioxygenase 1 (IDO1) Inhibitors. *J. Med. Chem.* **2015**, *58* (24), 9421–37.
- (74) Chen, Y. N.; LaMarche, M. J.; Chan, H. M.; Fekkes, P.; Garcia-Fortanet, J.; Acker, M. G.; Antonakos, B.; Chen, C. H.; Chen, Z.; Cooke, V. G.; Dobson, J. R.; Deng, Z.; Fei, F.; Firestone, B.; Fodor, M.; Fridrich, C.; Gao, H.; Grunenfelder, D.; Hao, H. X.; Jacob, J.; Ho, S.; Hsiao, K.; Kang, Z. B.; Karki, R.; Kato, M.; Larrow, J.; La Bonte, L. R.; Lenoir, F.; Liu, G.; Liu, S.; Majumdar, D.; Meyer, M. J.; Palermo, M.; Perez, L.; Pu, M.; Price, E.; Quinn, C.; Shakya, S.; Shultz, M. D.; Slisz, J.; Venkatesan, K.; Wang, P.; Warmuth, M.; Williams, S.; Yang, G.; Yuan, J.; Zhang, J. H.; Zhu, P.; Ramsey, T.; Keen, N. J.; Sellers, W. R.; Stams, T.; Fortin, P. D. Allosteric inhibition of SHP2 phosphatase inhibits cancers driven by receptor tyrosine kinases. *Nature* **2016**, *535* (7610), 148–52.
- (75) Xie, J.; Si, X.; Gu, S.; Wang, M.; Shen, J.; Li, H.; Shen, J.; Li, D.; Fang, Y.; Liu, C.; Zhu, J. Allosteric Inhibitors of SHP2 with Therapeutic Potential for Cancer Treatment. *J. Med. Chem.* **2017**, *60* (24), 10205–10219.
- (76) Scott, J. S.; Degorce, S. L.; Anjum, R.; Culshaw, J.; Davies, R. D. M.; Davies, N. L.; Dillman, K. S.; Dowling, J. E.; Drew, L.; Ferguson, A. D.; Groombridge, S. D.; Halsall, C. T.; Hudson, J. A.; Lamont, S.; Lindsay, N. A.; Marden, S. K.; Mayo, M. F.; Pease, J. E.; Perkins, D. R.; Pink, J. H.; Robb, G. R.; Rosen, A.; Shen, M.; McWhirter, C.; Wu, D. Discovery and Optimization of Pyrrolopyrimidine Inhibitors of Interleukin-1 Receptor Associated Kinase 4 (IRAK4) for the Treatment of Mutant MYD88(L265P) Diffuse Large B-Cell Lymphoma. *J. Med. Chem.* **2017**, *60* (24), 10071–10091.
- (77) Lee, K. L.; Ambler, C. M.; Anderson, D. R.; Boscoe, B. P.; Bree, A. G.; Brodfuehrer, J. I.; Chang, J. S.; Choi, C.; Chung, S.; Curran, K. J.; Day, J. E.; Dehnhardt, C. M.; Dower, K.; Drozda, S. E.; Frisbie, R. K.; Gavrin, L. K.; Goldberg, J. A.; Han, S.; Hegen, M.; Hepworth, D.; Hope, H. R.; Kamtekar, S.; Kilty, I. C.; Lee, A.; Lin, L. L.; Lovering, F. E.; Lowe, M. D.; Mathias, J. P.; Morgan, H. M.; Murphy, E. A.; Papaioannou, N.; Patny, A.; Pierce, B. S.; Rao, V. R.; Saiah, E.; Samardjiev, I. J.; Samas, B. M.; Shen, M. W. H.; Shin, J. H.; Soutter, H. H.; Strohbach, J. W.; Symanowicz, P. T.; Thomason, J. R.; Trzuppek, J. D.; Vargas, R.; Vincent, F.; Yan, J.; Zapf, C. W.; Wright, S. W. Discovery of Clinical Candidate 1-[[{(2S,3S,4S)-3-Ethyl-4-fluoro-5-oxopyrrolidin-2-yl]methoxy}-7-methoxyisoquinolin-6-carboxamide (PF-06650833), a Potent, Selective Inhibitor of Interleukin-1 Receptor Associated Kinase 4 (IRAK4), by Fragment-Based Drug Design. *J. Med. Chem.* **2017**, *60* (13), 5521–5542.
- (78) Collin, A. BRD4 (bromodomain containing 4). *Atlas of Genetics and Cytogenetics in Oncology and Haematology* **2011**, *11* (3), 180–181.
- (79) Vainio, M. J.; Johnson, M. S. Generating Conformer Ensembles Using a Multiobjective Genetic Algorithm. *J. Chem. Inf. Model.* **2007**, *47* (6), 2462–2474.