

# Contemporary Computational Applications and Tools in Drug Discovery

Philip B. Cox\* and Rishi Gupta\*

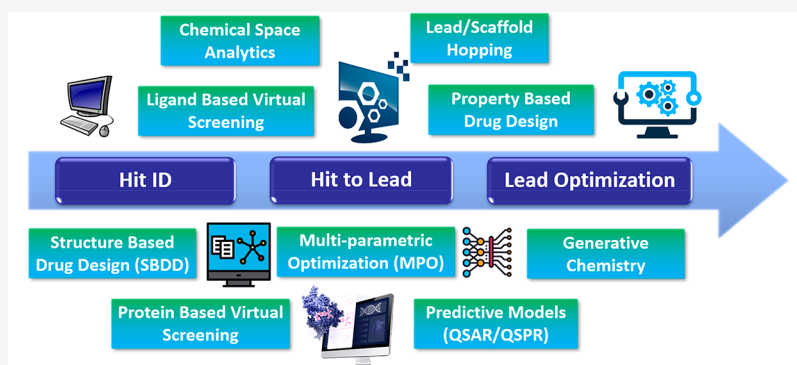
Cite This: *ACS Med. Chem. Lett.* 2022, 13, 1016–1029

Read Online

ACCESS |

Metrics &amp; More

Article Recommendations



**ABSTRACT:** In the past decade or so there has been a dramatic increase in the number of computational applications and tools that have been developed to enable medicinal chemists to prosecute modern drug discovery programs more efficiently. The upsurge of user-friendly, well-designed computational tools that enable structure-based drug design (SBDD) and cheminformatics (CI)-based drug design has equipped the medicinal chemist with an arsenal of tools and applications that significantly augments the entire design process, thereby enhancing the speed and efficiency of the design–make–test–analyze cycle. Modern computational applications and tools transcend all areas of drug discovery, and most savvy medicinal chemists can employ them effectively in a myriad of drug discovery applications. Indeed, the sheer scope and breadth of tools available to the medicinal chemist is vast and, to our knowledge, has not been comprehensively reviewed. In this article we have catalogued many computational tools, platforms, and applications that are currently available, with four main areas highlighted: commercially available tools/platforms, open-source applications, internally developed platforms (software tools developed within a pharma or biotech organization), and artificial intelligence/machine learning-based platforms. For ease of interpretation, for these categories we provide tables organized by vendor or organization name, the name of the application, whether the tool/application is employed predominantly for SBDD or CI-based design, and a summary of the main function of the tools, with associated hyperlinks to vendor Web sites. We have tried to be as comprehensive and as inclusive as possible; however, the pace of development of new and existing tools is so rapid that there may be omissions with respect to newly developed tools and current versions of the software.

**KEYWORDS:** Structure-based drug design, Cheminformatics, Artificial intelligence, Machine learning, Retrosynthetic analysis, DMTA, CADD, Drug discovery paradigm, Design applications and tools, Software, Open source

Computational applications and tools have been adopted to enable and accelerate the discovery of drugs since the early 1980s, regarded as the birth of computer-aided drug design (CADD). Since then, advances in the field have blossomed due to innovative applications of theoretical chemistry and physics, computer science, and statistics and the vast concomitant increase in computational power. On the heels of vast computer power has come the data science evolution that has enabled the derivatization of vast amounts of knowledge, thus enabling the burgeoning fields of bio- and cheminformatics. In the past decade, with the development of user-friendly software, the use of computational tools for drug discovery has shifted from a predominantly CADD specialist

role to a reality now where the medicinal chemist routinely employs a spectrum of desktop tools for modern drug design purposes. Indeed, some companies have encouraged, empowered, and trained specialized medicinal chemists to utilize computational tools in a specific “designer” role to potentially

Received: November 23, 2021

Accepted: May 16, 2022

Published: June 1, 2022



streamline the drug discovery process. Whether the designer model is adopted is largely governed by the philosophy and culture of individual organizations, and the value of this approach remains an area of debate among the medicinal chemistry community. Ultimately, though, the fact remains that savvy medicinal chemists embrace state-of-the-art computational tools to effectively augment drug design. In the past decade alone, the number of computational tools that have been developed by vendors has increased significantly, with the spectrum of applications broadening appreciably in addition. Whereas 10 or 15 years ago pharma companies were more apt to develop their own bespoke tools, today the availability of commercial tools has shifted the focus for many companies toward drug design using these user-friendly, robust, well-built, and well-conceived design platforms in an attempt to improve the entire design–make–test–analyze (DMTA) cycle. The optimization of hits to leads to candidates and ultimately approved drugs is very much an iterative process where many parameters must be optimized in parallel. Balancing and optimizing multiple parameters (multiple-parameter optimization (MPO), or scoring) to produce chemical matter that is potent, selective, soluble, permeable, and metabolically stable, and has good oral bioavailability, with an adequate safety profile, is a vastly complex and challenging process which can take many years. Parallel processing is not something that humans are particularly adept at; computers, on the other hand, were built for such purposes. In addition, computers are quite useful at repetitive work where large amounts of data are handled and the likelihood of human error is high. With the advent of machine learning and, more recently, deep learning, efficient computer-based MPO is a reality and, combined with generative chemistry methods, can significantly reduce the number of DMTA cycles, the number of compounds to be synthesized, and more importantly the time and resources it takes for a drug to reach the clinic. Currently no artificial intelligence/machine learning (AI/ML)-derived compounds have made it all the way to the market; however, it is just a matter of time before this eventually transpires. This illustrates how the burgeoning field of computational drug discovery has led to a concomitant growth in innovative computational tools and applications that are now available to drug hunters.

Computational tools have applicability in many areas of drug discovery and development. Many of the applications of these tools toward specific areas of drug discovery have been reviewed several times before<sup>1</sup> and therefore will not be the focus of this article. Instead, for this review we have catalogued many of the available computational applications and tools that have recently emerged and evolved specifically to enable drug discovery, particularly as it relates to the DMTA cycle. Additionally, the focus of the article is centered on tools (Tables 1–5) that enable structure-based drug design (SBDD) and/or cheminformatics (CI)-based design. SBDD facilitates the design of potential drugs utilizing crystallographically derived protein structures or related homology models with and without bound ligands. Examples of such applications are virtual docking/screening and de novo design. CI-based design tools are those that facilitate the design of potential drugs through the curation, manipulation, and analysis of chemical and biological data. Examples of CI applications are chemical similarity calculations and searching, clustering, R-group analysis, and matched molecular pair analysis. Five main areas are highlighted in the tables: Table 1 highlights

commercially available tools/platforms; Table 2, internally developed platforms by individual pharma companies; Table 3, open-source tools and software; Table 4, companies whose focus is predominantly on AI/ML applications to drug discovery; and Table 5, tools employed specifically for synthesis predictions using ML methods. Note also that, for AI/ML-based companies in Table 4, we have made the distinction between companies that are predominantly software and service providers and companies that use AI/ML as the prime drug discovery engine and may not necessarily offer access to their platforms/software. For ease of interpretation, these tables are organized by vendor/organization name, the name of the application, whether the tool/application is employed predominantly for SBDD or CI-based design, and a summary of the main function of the tools, with the associated hyperlinks to vendor Web sites—note that many of the descriptions are taken directly from the vendor's Web site. The assignment of whether the computational tools are predominantly used for SBDD or CI design in some cases is subjective, as the software may have multiple functions that cross both design paradigms. For clear examples of this we have checked both boxes, but there may be instances where the bias between the two paradigms is not as clear and only one box (paradigm) was selected. We have tried to be as comprehensive and as inclusive as possible; however, the pace of development of new and existing tools is so rapid that there may be some omissions with respect to newly developed tools and the current versions of the software.

As we consider the drug discovery continuum from Hit ID → Hit to Lead → Lead Optimization (see Scheme 1), computational tools are essential for all these stages of drug discovery. For Hit ID, for example, there are many approaches to virtual screening that use both protein-based and ligand-based computational approaches. There are pros and cons for both of these approaches; however, there have been significant innovations and enhancements in recent years, with many vendors offering tailored searching/docking algorithms that often identify both complementary (similar) and orthogonal (dissimilar) hits.

At the early stages of a project, where screening is required for Hit ID, it is essential to understand the scope, breadth, and extent of the chemical space that one needs to explore to identify tractable chemical matter. Corporate screening collections are finite and have numbers that do not adequately sample the vastness of drug-like chemical space. Even with the introduction of DNA encoded library (DEL) technologies, which has dramatically expanded synthetically accessible chemical space into the billions, both curating and exploring ultra-vast chemical space can only be achieved using innovative computational methods. There are many ways to approach building virtual libraries, either using fully enumerated structures with large but limited numbers ( $10^6$ ) or with topological/pharmacophoric feature-encoded fragmented molecular representations, such as Feature Trees (FTrees<sup>51</sup>), that enable the virtual expansion of chemical space to unprecedented levels ( $10^{26}$ ).<sup>2</sup> There are undoubtedly pros and cons associated with these approaches, but they have been effective at Hit ID, employed either individually or in concert. The exploration of chemical space is enabled using computational methods, with vendors offering a variety of tools to enable searching, clustering, similarity assessments, and mapping (dimensionality reduction), to mention a few. Note that virtual screening can be employed at any stage of the drug

Table 1. Various Commercial Software Tools Available for Cheminformatics (CI)-Based and Structure-Based Drug Design (SBDD)

vendor	application	CI	SBDD	main function	Web site
Schrödinger	Maestro <sup>7</sup>	✓	✓	workhorse platform for all-purpose molecular modeling	<a href="https://www.schrodinger.com/maestro">https://www.schrodinger.com/maestro</a>
	AutoQSAR <sup>8</sup>	✓	✓	automated creation and application of predictive QSAR models	<a href="https://www.schrodinger.com/autosar">https://www.schrodinger.com/autosar</a>
	ConfGen <sup>9</sup>	✓	✓	accurate and efficient bioactive conformational searching	<a href="https://www.schrodinger.com/confgen">https://www.schrodinger.com/confgen</a>
	Desmond <sup>10</sup>	✓	✓	high-performance MD simulations	<a href="https://www.schrodinger.com/desmond">https://www.schrodinger.com/desmond</a>
	FEP+ <sup>11</sup>	✓	✓	high-performance free energy calculations for drug discovery	<a href="https://www.schrodinger.com/fep">https://www.schrodinger.com/fep</a>
	Glide <sup>12</sup>	✓	✓	solution for ligand–receptor docking	<a href="https://www.schrodinger.com/glide">https://www.schrodinger.com/glide</a>
	Induced Fit <sup>13</sup>	✓	✓	fast and accurate prediction of ligand-induced changes in receptor active sites	<a href="https://www.schrodinger.com/induced-fit">https://www.schrodinger.com/induced-fit</a>
	LigPrep <sup>14</sup>	✓	✓	versatile generation of accurate 3D molecular models	<a href="https://www.schrodinger.com/ligprep">https://www.schrodinger.com/ligprep</a>
	MacroModel <sup>15</sup>	✓	✓	molecular modeling platform	<a href="https://www.schrodinger.com/macromodel">https://www.schrodinger.com/macromodel</a>
	Phase <sup>16</sup>	✓	✓	easy-to-use pharmacophore modeling solution for ligand- and structure-based drug design	<a href="https://www.schrodinger.com/phase">https://www.schrodinger.com/phase</a>
	PyMOL <sup>17</sup>	✓	✓	high-performance molecular graphics platform for 3D visualization	<a href="https://www.schrodinger.com/pymol">https://www.schrodinger.com/pymol</a>
	QSite <sup>18</sup>	✓	✓	high-performance QM/MM program	<a href="https://www.schrodinger.com/qsite">https://www.schrodinger.com/qsite</a>
	Shape Screening <sup>19</sup>	✓	✓	VLS using 3D shape-based similarity	<a href="https://www.schrodinger.com/shape-screening">https://www.schrodinger.com/shape-screening</a>
	Core Hopping <sup>20</sup>	✓	✓	comprehensive ligand- and receptor-based scaffold exploration tool for lead optimization	<a href="https://www.schrodinger.com/core-hopping">https://www.schrodinger.com/core-hopping</a>
	e-Pharmacophores <sup>21</sup>	✓	✓	energetically optimized structure-based pharmacophores	<a href="https://www.schrodinger.com/e-pharmacophores">https://www.schrodinger.com/e-pharmacophores</a>
	Field-Based QSAR <sup>22</sup>	✓	✓	discover and optimize new lead compounds using quantitative predictions of binding-site chemistry	<a href="https://www.schrodinger.com/field-based-qsar">https://www.schrodinger.com/field-based-qsar</a>
	Jaguar <sup>23</sup>	✓	✓	rapid ab initio QM electronic structure package	<a href="https://www.schrodinger.com/jaguar">https://www.schrodinger.com/jaguar</a>
LiveDesign <sup>24</sup>	✓	✓	next-generation platform for collaborative drug design	<a href="https://www.schrodinger.com/live-design/drug-discovery">https://www.schrodinger.com/live-design/drug-discovery</a>	
QikProp <sup>25</sup>	✓	✓	rapid ADMET prediction of drug candidates	<a href="https://www.schrodinger.com/qikprop">https://www.schrodinger.com/qikprop</a>	
WaterMap <sup>26</sup>	✓	✓	solution for desolvation thermodynamics	<a href="https://www.schrodinger.com/watermap">https://www.schrodinger.com/watermap</a>	
SiteMap <sup>27</sup>	✓	✓	fast, accurate, and practical binding site predictor	<a href="https://www.schrodinger.com/sitemap">https://www.schrodinger.com/sitemap</a>	
Canvas <sup>28</sup>	✓	✓	comprehensive cheminformatics computing environment	<a href="https://www.schrodinger.com/canvas">https://www.schrodinger.com/canvas</a>	
EPIK <sup>29</sup>	✓	✓	fast and robust pK <sub>a</sub> predictions	<a href="https://www.schrodinger.com/epik">https://www.schrodinger.com/epik</a>	
CovDock <sup>30</sup>	✓	✓	work-flow for pose prediction for covalently bound ligands	<a href="https://www.schrodinger.com/covdock">https://www.schrodinger.com/covdock</a>	
Cresset	Forge <sup>31</sup>	✓	✓	ligand-based workbench for molecular design and SAR analysis	<a href="https://www.cresset-group.com/software/forge/">https://www.cresset-group.com/software/forge/</a>
	Activity Atlas <sup>32</sup>	✓	✓	component of Forge for SAR interpretation and visualization	<a href="https://www.cresset-group.com/software/forge-activity-atlas/">https://www.cresset-group.com/software/forge-activity-atlas/</a>
	Activity Miner <sup>32</sup>	✓	✓	component of Forge to find and understand activity cliffs	<a href="https://www.cresset-group.com/software/forge-activity-miner/">https://www.cresset-group.com/software/forge-activity-miner/</a>
	FieldTemplater <sup>32</sup>	✓	✓	component of Forge to generate the most accurate field pharmacophores available	<a href="https://www.cresset-group.com/software/field-templater/">https://www.cresset-group.com/software/field-templater/</a>
Open-Eye	Flare <sup>33</sup>	✓	✓	protein–ligand analysis platform for advanced SBDD	<a href="https://www.cresset-group.com/software/flare/">https://www.cresset-group.com/software/flare/</a>
	Spark <sup>32</sup>	✓	✓	scaffold hopping and R-group replacement for innovative molecular design	<a href="https://www.cresset-group.com/software/spark/">https://www.cresset-group.com/software/spark/</a>
	Blaze <sup>34</sup>	✓	✓	ligand-based virtual screening	<a href="https://www.cresset-group.com/software/blaze/">https://www.cresset-group.com/software/blaze/</a>
	PickR <sup>32</sup>	✓	✓	advanced electrostatic diversity monomer selection tool for library design	<a href="https://www.cresset-group.com/software/pickr/">https://www.cresset-group.com/software/pickr/</a>
	Lead Finder <sup>35</sup>	✓	✓	high-throughput docking and scoring for VLS	<a href="https://www.cresset-group.com/software/lead-finder/">https://www.cresset-group.com/software/lead-finder/</a>
Torx <sup>32</sup>	✓	✓	next-generation platform for smallmolecules team-working and collaboration	<a href="https://www.cresset-group.com/software/torx/">https://www.cresset-group.com/software/torx/</a>	
cheminformatics tool kits	cheminformatics tool kits	✓	✓	real-time shape similarity for VLS, lead hopping, and shape clustering	<a href="https://www.eyesopen.com/molecular-modeling-fastrocs">https://www.eyesopen.com/molecular-modeling-fastrocs</a>
	FastROCS TK <sup>36</sup>	✓	✓	core chemistry handling and representation	<a href="https://www.eyesopen.com/oechem-tk">https://www.eyesopen.com/oechem-tk</a>
	OEChem TK <sup>37</sup>	✓	✓	2D molecule rendering and depiction	<a href="https://www.eyesopen.com/oe depict-tk">https://www.eyesopen.com/oe depict-tk</a>
OEdepict TK	✓	✓			

Table 1. continued

vendor	application	CI	SBDD	main function	Web site
	Grapheme TK	✓		Advanced molecule rendering and report generation	<a href="https://www.eyesopen.com/grapheme-tk">https://www.eyesopen.com/grapheme-tk</a>
	GraphSim TK <sup>38</sup>	✓		2D molecular similarity (e.g., fingerprints)	<a href="https://www.eyesopen.com/graphsim-tk">https://www.eyesopen.com/graphsim-tk</a>
	Lexichem TK <sup>39</sup>	✓		name-to-structure and structure-to-name translator	<a href="https://www.eyesopen.com/lexichem-tk">https://www.eyesopen.com/lexichem-tk</a>
	MolProp TK	✓		molecular property calculation and filtering	<a href="https://www.eyesopen.com/molprop-tk">https://www.eyesopen.com/molprop-tk</a>
	Quacpac TK <sup>40</sup>	✓		tautomer enumeration and charge assignment	<a href="https://www.eyesopen.com/quacpac-tk">https://www.eyesopen.com/quacpac-tk</a>
	MedChemTK	✓		matched molecular pair analysis, fragmentation, and molecular complexity metrics	<a href="https://www.eyesopen.com/oemedchem-tk">https://www.eyesopen.com/oemedchem-tk</a>
	modeling tool kits				
	OEDocking TK <sup>41</sup>	✓		molecular docking and scoring	<a href="https://www.eyesopen.com/oedocking-tk">https://www.eyesopen.com/oedocking-tk</a>
	Omega TK <sup>42</sup>	✓		conformer generation	<a href="https://www.eyesopen.com/omega-tk">https://www.eyesopen.com/omega-tk</a>
	Shape TK <sup>43</sup>	✓		3D shape description, manipulators, and interrogation	<a href="https://www.eyesopen.com/shape-tk">https://www.eyesopen.com/shape-tk</a>
	Spicoli TK	✓		surface generation, manipulation, and interrogation	<a href="https://www.eyesopen.com/spicoli-tk">https://www.eyesopen.com/spicoli-tk</a>
	Szmap TK	✓		understanding water interactions in binding sites	<a href="https://www.eyesopen.com/szmap-tk">https://www.eyesopen.com/szmap-tk</a>
	Zap TK <sup>44</sup>	✓		calculate Poisson–Boltzmann electrostatic potentials	<a href="https://www.eyesopen.com/zap-tk">https://www.eyesopen.com/zap-tk</a>
	Szybki TK <sup>45</sup>	✓		general purpose optimization with MMFF94	<a href="https://www.eyesopen.com/szybki-tk">https://www.eyesopen.com/szybki-tk</a>
	Orion	✓		cloud-based platform to develop custom computational drug discovery workflows and visualizations	<a href="https://www.eyesopen.com/orion">https://www.eyesopen.com/orion</a>
BioSolve-IT	SeeSAR <sup>46</sup>	✓		visual compound prioritization and evolution; structure-based design work supporting MPO; visualize binding affinity at interfaces	<a href="https://www.biosolveit.de/SeeSAR/">https://www.biosolveit.de/SeeSAR/</a>
	infiniSee <sup>47</sup>	✓		finds molecules chemically similar to given molecule or template; navigates through “almost infinite space” and employs FTrees technology	<a href="https://www.biosolveit.de/infiniSee/">https://www.biosolveit.de/infiniSee/</a>
	FlexX <sup>48</sup>	✓		structure-based docking; binding mode prediction	<a href="https://www.biosolveit.de/FlexX/">https://www.biosolveit.de/FlexX/</a>
	HYDE <sup>49</sup>	✓		structure-based scoring; compound classification	<a href="https://www.biosolveit.de/HYDE/">https://www.biosolveit.de/HYDE/</a>
	FlexS <sup>50</sup>	✓		ligand-based alignment in 3D; prep for 3D QSAR; find scaffold/compound mimics; virtual screening	<a href="https://www.biosolveit.de/FlexS/">https://www.biosolveit.de/FlexS/</a>
	FTrees <sup>51</sup>	✓		ligand-based similarity; fuzzy matching to detect novel molecular scaffolds	<a href="https://www.biosolveit.de/FTrees/">https://www.biosolveit.de/FTrees/</a>
	ReCore <sup>51</sup>	✓		fragment-based; replace central elements in known bioactive molecules to create new custom 3D scaffold	<a href="https://www.biosolveit.de/ReCore/">https://www.biosolveit.de/ReCore/</a>
	FTrees-FS <sup>51</sup>	✓		combinatorial fragment space extension module for FTrees; builds up compounds from fragments while still comparing and searching for similarity	<a href="https://www.biosolveit.de/products/#FTrees">https://www.biosolveit.de/products/#FTrees</a>
	CoLibri <sup>51</sup>	✓		toolkit for chemical space exploration; generate chemical space; focus on synthetically accessible entities	<a href="https://www.biosolveit.de/CoLibri/">https://www.biosolveit.de/CoLibri/</a>
	REAL Space Navigator	✓		collaboration with Enamine, “world’s largest, ultrafast” searchable chemical space; 13 billion compounds	<a href="https://www.biosolveit.de/REALSpaceNavigator/">https://www.biosolveit.de/REALSpaceNavigator/</a>
	GalaXi	✓		explore chemical space commercially available on demand from Wuxi	<a href="https://www.biosolveit.de/2019/10/10/launch-of-galaxi/">https://www.biosolveit.de/2019/10/10/launch-of-galaxi/</a>
	KnowledgeSpace	✓		virtual chemical space containing 100+ literature reactions to produce likely synthetically accessible compounds	<a href="https://www.biosolveit.de/CoLibri/spaces.html">https://www.biosolveit.de/CoLibri/spaces.html</a>
Dotmatics	Browser	✓		searching and browsing across all corporate scientific databases	<a href="https://www.dotmatics.com/products/browser">https://www.dotmatics.com/products/browser</a>
	D40	✓		“Dotmatics for Microsoft Office” connector to integrate with Microsoft Office suite	<a href="https://www.dotmatics.com/products">https://www.dotmatics.com/products</a>
	Blueprint	✓		Web-based platform for visualization and scientific data analysis for small-molecule design	<a href="https://www.dotmatics.com/products/blueprint">https://www.dotmatics.com/products/blueprint</a>
	Vortex	✓		chemically intuitive data visualization and analysis’ expansion on the “spreadsheet”; can do statistical analysis and create visualizations	<a href="https://www.dotmatics.com/products/vortex/">https://www.dotmatics.com/products/vortex/</a>
Discngine	Assay	✓		design and manage cellular, molecular, and high-content screening campaigns	<a href="https://www.discngine.com/assay">https://www.discngine.com/assay</a>
	3Decision	✓		store, analyze, and share protein–ligand structures, sequences, and associated data	<a href="https://3decision.discngine.com/">https://3decision.discngine.com/</a>
	Chemistry Collection	✓		generate and visualize scaffold networks, perform molecular fragmentation, or use Pharmacophore Graph to design matched molecule pairs analysis	<a href="https://www.discngine.com/chemistry-collection">https://www.discngine.com/chemistry-collection</a>
	Network Collection	✓		represent, manage, and analyze complex network structures using advanced graph theory	<a href="https://www.discngine.com/network-collection">https://www.discngine.com/network-collection</a>

Table 1. continued

vendor	application	CI	SBDD	main function	Web site
Optibrium	Stardrop	✓		small-molecule design, optimization, and data analysis; can do predictive modeling and QSAR	<a href="https://www.optibrium.com/stardrop/stardrop-features.php">https://www.optibrium.com/stardrop/stardrop-features.php</a>
	Asteris	✓		iPad app for drug discovery	<a href="https://www.asteris-app.com/">https://www.asteris-app.com/</a>
	Ocura	✓		load and browse molecular structures; data viewer	<a href="https://www.optibrium.com/ocura/">https://www.optibrium.com/ocura/</a>
TIBCO	Spotfire TIBCO Data Science	✓ ✓		AI-powered search-driven experience; built-in data wrangling and advanced analytics share and deploy analytics models across organizations for increased collaboration	<a href="https://www.tibco.com/products/tibco-spotfire">https://www.tibco.com/products/tibco-spotfire</a> <a href="https://www.tibco.com/products/data-science">https://www.tibco.com/products/data-science</a>
Dassault Systèmes	Biovia Pipeline Pilot Discovery Studio	✓ ✓	✓	data pipelining tool for creating simple as well as advanced workflows for various data science initiatives such as cheminformatics, bioinformatics, ML, etc. data pipelining tool for advanced 3D design such as molecular mechanics, free energy calculations, biotherapeutics developability, and more in a common environment	<a href="https://www.3ds.com/products-services/biovia/products/data-science/pipeline-pilot/">https://www.3ds.com/products-services/biovia/products/data-science/pipeline-pilot/</a> <a href="https://www.3ds.com/products-services/biovia/products/molecular-modeling-simulation/biovia-discovery-studio/">https://www.3ds.com/products-services/biovia/products/molecular-modeling-simulation/biovia-discovery-studio/</a>
CCDC	mercury GOLD	✓ ✓	✓ ✓	knowledge-based geometry and interaction preferences protein–ligand docking	<a href="https://www.ccdc.cam.ac.uk/mercury/">https://www.ccdc.cam.ac.uk/mercury/</a> <a href="https://www.ccdc.cam.ac.uk/solutions/csd-discovery/components/gold/">https://www.ccdc.cam.ac.uk/solutions/csd-discovery/components/gold/</a>
	CSD-CrossMiner	✓		discovery-oriented data mining of protein and small-molecule crystal structures	<a href="https://www.ccdc.cam.ac.uk/solutions/csd-discovery/components/csd-crossminer">https://www.ccdc.cam.ac.uk/solutions/csd-discovery/components/csd-crossminer</a>
	SuperStar	✓		knowledge-based pharmacophore prediction	<a href="https://www.ccdc.cam.ac.uk/solutions/csd-discovery/components/superstar/">https://www.ccdc.cam.ac.uk/solutions/csd-discovery/components/superstar/</a>
	CSD Python API	✓		Python-based programmatic access to the tools	<a href="https://www.ccdc.cam.ac.uk/solutions/csd-materials/components/csd-python-api/">https://www.ccdc.cam.ac.uk/solutions/csd-materials/components/csd-python-api/</a>
	Mogul	✓	✓	knowledge-based library of molecular geometry derived from the Cambridge Structural Database (CSD)	<a href="https://www.ccdc.cam.ac.uk/solutions/csd-system/components/mogul/">https://www.ccdc.cam.ac.uk/solutions/csd-system/components/mogul/</a>
CCG	Molecular Operating Environment (MOE) <sup>52</sup> MOESaic <sup>52</sup>	✓ ✓	✓ ✓	integrated computer-aided molecular design platform; includes 3D visualization, structure-based design, SAR explorer, data modeling, virtual screening, simulations, QSAR, etc. for small molecules as well as antibody–drug conjugates and biologics allows small-molecule SAR analysis such as MMPs, R-group, etc.	<a href="https://www.chemcomp.com/Products.htm">https://www.chemcomp.com/Products.htm</a> <a href="https://www.chemcomp.com/Products.htm#MOESaic-SAR_Explorer">https://www.chemcomp.com/Products.htm#MOESaic-SAR_Explorer</a>
	PSILO <sup>52</sup>	✓	✓	data repository and visualization application to store curated RCSB protein structures, pocket similarity, and other features for data analysis	<a href="https://www.chemcomp.com/Products.htm#PSILO-Structure_Database">https://www.chemcomp.com/Products.htm#PSILO-Structure_Database</a>
Medchemica	MCPairs	✓	✓	AI/ML-based SAR analysis application to generate match pairs and new ideas	<a href="https://www.medchemica.com/products/">https://www.medchemica.com/products/</a>
ChemAlive	ConstruQt	✓	✓	high-throughput quantum chemistry for molecular design; allows automated library-scale deployment of quantum chemical calculations	<a href="https://www.chemalive.com/construqt/">https://www.chemalive.com/construqt/</a>
CDD Vault	CDD Vault	✓	✓	informatics platform; molecular/data viewing and browsing	<a href="https://www.collaborative-drug.com/benefits/">https://www.collaborative-drug.com/benefits/</a>
BioSymetrics	Augusta	✓	✓	biomedical-specific ML framework; drug discovery through small-molecule activity prediction	<a href="https://www.biosymetrics.com/products/mona-prediction/">https://www.biosymetrics.com/products/mona-prediction/</a>
Eidogen Ser-tanty	Target Informatics Platform (TIP) TIP Calculation Engine (STRUCFAST, SiteSeeker, SiteSorter, SLiC)	✓	✓	interrogate the druggable genome from a structural perspective; bridge the gap between bio- and cheminformatics structure determination; binding site annotation; similarity assessment	<a href="https://www.eidogen.com/tip.php">https://www.eidogen.com/tip.php</a> <a href="https://www.eidogen.com/tae.php">https://www.eidogen.com/tae.php</a>



Table 1. continued

vendor	application	CI	SBDD	main function	Web site
	Eidogen Visualization Environment (EVE)	✓	✓	visualize and compare small-molecule binding sites of targets of interest	<a href="https://www.eidogen.com/eve.php">https://www.eidogen.com/eve.php</a>
	Oncology Knowledgebase (OKB)	✓		database of structure–activity data across targets of oncological interest	<a href="https://www.eidogen.com/oncologykb.php">https://www.eidogen.com/oncologykb.php</a>
	Kinase Knowledgebase (KKB)	✓		database of kinase structure–activity and chemical synthesis data	<a href="https://www.eidogen.com/kinasekb.php">https://www.eidogen.com/kinasekb.php</a>
	ChIP	✓		algorithm-driven enumeration engine for automated generation of medically relevant, novel molecules with proven synthetic access	<a href="https://www.eidogen.com/chip.php">https://www.eidogen.com/chip.php</a>
VeraChem	VM2	✓	✓	software package for molecular-binding free energy calculations	<a href="https://www.verachem.com/products/vm2">https://www.verachem.com/products/vm2</a>
	VConf	✓	✓	2D-to-3D small-molecule conversion	<a href="https://www.verachem.com/products/vconf">https://www.verachem.com/products/vconf</a>
	VCharge	✓	✓	accurate partial atomic charges of drug-like molecules	<a href="http://www.verachem.com/products/vcharge">http://www.verachem.com/products/vcharge</a>
	VFilter	✓	✓	analyze ensembles of molecular conformations and remove repeats	<a href="http://www.verachem.com/products/vfilter">http://www.verachem.com/products/vfilter</a>
	Vrms	✓	✓	provides symmetry-corrected root-mean-square deviation between molecular conformers	<a href="https://www.verachem.com/products/vrms/">https://www.verachem.com/products/vrms/</a>
	VDisplay	✓	✓	3D molecular viewer	<a href="https://www.verachem.com/products/vdisplay/">https://www.verachem.com/products/vdisplay/</a>
NextMove	Arthor <sup>53</sup>	✓		fast, state-of-the-art substructure and chemical similarity search capabilities for ultra-large databases of 100s of millions of compounds, using SMARTS optimization, Just-In-Time compilation, and/or GPUs	<a href="https://www.nextmovesoftware.com/arthor.html">https://www.nextmovesoftware.com/arthor.html</a>
	Matsy <sup>54</sup>	✓	✓	set of tools for creating and analyzing matched molecular series (the general form of MMPS); In particular, can be used to suggest what compound to make next in a med-chem program	<a href="https://www.nextmovesoftware.com/matsy.html">https://www.nextmovesoftware.com/matsy.html</a>
	MPSearch <sup>54</sup>	✓	✓	rapidly searches a database to find matched pairs related to a query molecule. This type of search is used to explore previous medicinal chemistry strategies	<a href="https://www.nextmovesoftware.com/mpsearch.html">https://www.nextmovesoftware.com/mpsearch.html</a>
	Patsy <sup>55</sup>	✓	✓	speed up SMARTS pattern matching by creating optimized SMARTS patterns of source code; speed gains are particularly large when multiple SMARTS patterns are matched against a single structure	<a href="https://www.nextmovesoftware.com/patsy.html">https://www.nextmovesoftware.com/patsy.html</a>
	SmallWorld <sup>56</sup>	✓	✓	index of chemical space based on more than 230 billion substructures; can be used to measure similarity based on graph-edit distance, find the maximum common subgraph of two or more molecules, analyze HTS results, and more	<a href="https://www.nextmovesoftware.com/smallworld.html">https://www.nextmovesoftware.com/smallworld.html</a>
	CaffeineFix <sup>57</sup>	✓	✓	rapidly match chemical names or terms against a dictionary of grammar (e.g., a grammar for IUPAC names); use in text mining; can be used to provide autocorrected functionality and spell-correction	<a href="https://www.nextmovesoftware.com/caffeinefix.html">https://www.nextmovesoftware.com/caffeinefix.html</a>
	LeadMine <sup>58</sup>	✓	✓	extracts chemical names and terms from text; incorporates CaffeineFix technology to find terms that match appropriate dictionaries or grammars; has enhanced functionality to handle patent literature	<a href="https://www.nextmovesoftware.com/leadmine.html">https://www.nextmovesoftware.com/leadmine.html</a>
	Casandra <sup>59</sup>	✓	✓	server for delivering real-time safety warning of experimental hazards straight to pharmaceutical ELNs	<a href="https://www.nextmovesoftware.com/casandra.html">https://www.nextmovesoftware.com/casandra.html</a>
	HazELNut <sup>60</sup>	✓	✓	suite of tools used to extract, normalize, and analyze information in ELNs; can be used to implement a search interface, find/eliminate duplicates, find similar reactions, etc.	<a href="https://www.nextmovesoftware.com/hazelnut.html">https://www.nextmovesoftware.com/hazelnut.html</a>
	NameRxn <sup>61</sup>	✓	✓	used to classify and name reactions; particularly useful in the context of ELN analysis but also as a plug-in to chemical drawing software; builds on NextMove Patsy technology	<a href="https://www.nextmovesoftware.com/namerxn.html">https://www.nextmovesoftware.com/namerxn.html</a>
	Pistachio <sup>62</sup>	✓	✓	reaction dataset browser providing loading, querying, and analytics of chemical reactions; with over 9 million chemical reactions extracted from U.S. and EPO patents, it demonstrates an AI interface to faceted (structure) search	<a href="https://www.nextmovesoftware.com/pistachio.html">https://www.nextmovesoftware.com/pistachio.html</a>
Alvascience	alvaMolecule	✓		visual analytics platform allowing users to standardize and organize chemistry project data	<a href="https://www.alvascience.com/alvamolecule/">https://www.alvascience.com/alvamolecule/</a>
	alvaModel	✓		build and deploy QSAR/QSPR regression models; consists of two pieces of software: alvaModel and alvaRunner	<a href="https://www.alvascience.com/alvamodel/">https://www.alvascience.com/alvamodel/</a>
	alvaDesc	✓		allows calculation of over 5000 0D/1D/2D/3D molecular descriptors	<a href="https://www.alvascience.com/alvadesc/">https://www.alvascience.com/alvadesc/</a>
	alvaBuilder	✓		molecule design platform utilizing user-selected property criteria	<a href="https://www.alvascience.com/alvabuilder/">https://www.alvascience.com/alvabuilder/</a>
Datagrok	Datagrok for Cheminformatics	✓		chemistry data visualization and analytics; provides a structure/substructure search as well as mechanism to build ML models	<a href="https://datagrok.ai/cheminformatics">https://datagrok.ai/cheminformatics</a>

Table 1. continued

vendor	application	CI	SBDD	main function	Web site
Molsoft	ICM Chemist Pro	✓		allows scientists to draw and edit chemicals, create and view chemical spreadsheets, and perform chemical searching, chemical clustering, and many other routine cheminformatics functionalities	<a href="https://www.molsoft.com/icm-chemist-pro.html">https://www.molsoft.com/icm-chemist-pro.html</a>
	ICM Pro		✓	desktop software environment for high-quality protein structure analysis, modeling, and docking	<a href="https://www.molsoft.com/icm_pro.html">https://www.molsoft.com/icm_pro.html</a>

discovery continuum, often to augment existing hit series and to identify new orthogonal tractable chemical series with the goal of increasing the probability of success. Indeed, if a project is enabled with a robust protein crystal structure, virtual docking is arguably the main vehicle for the optimization of ligand binding. Without structure, but equipped with a known binder, ligand-based virtual screening or lead/scaffold hopping can be used to identify new series to pursue. Fortunately, there are a growing number of computational tools from vendors that enable both ligand-based virtual screening and lead/scaffold hopping employing very innovative approaches that have been used successfully, particularly in Hit ID, but also in the Hit to Lead phase of the drug discovery continuum.

The Hit to Lead and Lead Optimization stages of the discovery continuum are complex and challenging paradigms, where many different properties of lead series must be optimized concurrently and iteratively to achieve the desired efficacious *in vivo* (pharmacokinetics/pharmacodynamics, PK/PD) profile in a relevant disease model. This may be achieved efficiently by adopting a MPO approach, a process that is becoming mainstream in contemporary medicinal chemistry optimization campaigns.<sup>3</sup> Therefore, in concert with the optimization of binding and functional activity, concurrent physicochemical property-based optimization is essential for the success of any medicinal chemistry campaign. To achieve this, a mix of both *in vitro* and *in silico* parameters must be employed to derive multi-parametric scoring functions tailored to an individual project. A growing number of computational design platforms are enabled to perform MPO using a range of tailored methods; in addition, vendors now offer this capability employing a variety of mathematical approaches.

Oftentimes, particularly in Lead Optimization, there is sufficient data to build robust quantitative structure–activity relationship (QSAR) models which may significantly enable a project team by augmenting the optimization of a lead to the desired outcome of candidate selection. Additionally, as most pharmaceutical companies have vast repositories of absorption, distribution, metabolism, excretion, and toxicity (ADMET) data, quantitative structure–property relationship (QSPR) models using a variety of deep-learning (DL)/machine-learning (ML) methods are now par for the course, with the predicted ADMET properties often calculated after registration of new chemical entities prior to capturing initial primary *in vitro* ADMET assay data. Models can then be augmented and improved continuously through machine-based learning methods employing the newly added data, with accuracy often improved for a given similar series. Many pharmaceutical and biotech companies are enabled with IT infrastructure that facilitates many of the computational methods needed to sufficiently augment and improve the drug discovery process. The extent and utility of these methods depends on the individual companies, with a mix of both internally and externally developed applications and platforms adopted. In the past couple of years, the power of computational drug discovery has been demonstrated, and the concomitant interest in the medicinal chemistry community has been heightened by the development of generative *de novo* design methods (Table 2). Using DL models, these methods can generate novel chemical entities which are optimized, *in silico*, to possess targeted affinity, PK, and even safety profiles. Moreover, AI-based platforms can learn which classes of chemical entities bind selective protein targets and generate associated novel chemical matter capable of binding and functionally modulat-

Table 2. Various Pharma Proprietary Software Tools Available for Cheminformatics (CI)- and Structure-Based Drug Design (SBDD)

vendor	application	CI	SBDD	main function	Web site
AbbVie	AIDEAS	✓	✓	AIDEAS (AbbVie's Integrated Design Explorer and Analytics Solution) is a one-stop shop for all cheminformatics tools such as phys-chem property calculation, ADME ML calculations, library enumeration, docking, R-group/MMP analysis, patent data extraction, clustering, etc.	<a href="https://www.bio-itworldexpo.com/software">https://www.bio-itworldexpo.com/software</a> (Oral presentation)
Vertex	ASAP	✓	✓	ASAP (Affinity, Selectivity, Activity, Properties and PK) is built using OpenEye toolkits and ChemAxon, for searching and exploring chemical and biological data	<a href="https://chemaxon.com/presentation/asap-emphasizing-multidimensional-drug-discovery">https://chemaxon.com/presentation/asap-emphasizing-multidimensional-drug-discovery</a>
	MedChem2	✓	✓	a modeling and design tool that integrates a number of cheminformatics tools such as molecular superposition, protein–ligand docking, property calculation, library design etc.	<a href="https://link.springer.com/article/10.1007/s10822-016-9994-0">https://link.springer.com/article/10.1007/s10822-016-9994-0</a> (ref 1a)
Pfizer	PCAT	✓	✓	a tool for clustering, organizing, and visualizing molecules with their associated properties and biological activities	<a href="https://pubs.acs.org/doi/pdf/10.1021/ci9002443?src=recsys">https://pubs.acs.org/doi/pdf/10.1021/ci9002443?src=recsys</a> (ref 63)
	PGVL			virtual library design-based platform	<a href="https://pubmed.ncbi.nlm.nih.gov/23020747/">https://pubmed.ncbi.nlm.nih.gov/23020747/</a> (ref 64)
	CCT			?	<a href="https://link.springer.com/article/10.1007/s10822-016-9997-x">https://link.springer.com/article/10.1007/s10822-016-9997-x</a> (ref 65)
GlaxoSmithKline	BRADSHAW	✓	✓	BRADSHAW (Biological Response Analysis and Design System using an Heterogenous, Automated Workflow) is GSK's experimental automated design environment	<a href="https://link.springer.com/article/10.1007/s10822-019-00234-8">https://link.springer.com/article/10.1007/s10822-019-00234-8</a>
Novartis	FOCUS	✓	✓	platform to produce, visualize, and share information on various aspects of a drug discovery project, such as cheminformatics, data analysis, structural information, and design; built on MolSoft's ICM	<a href="https://pubs.acs.org/doi/abs/10.1021/ci500598e">https://pubs.acs.org/doi/abs/10.1021/ci500598e</a> (ref 66)



Table 3. Various Open-Source Software Tools Available for Cheminformatics (CI) - and Structure-Based Drug Design (SBDD)

vendor	application	CI	SBDD	main function	Web site
OpenMolecules	DataWarrior	✓	✓	open-source cheminformatics application; includes several data analysis and visualization methods such as clustering, phys-chem property calculations, large-scale similarity calculations, dimensionality reduction, etc.	<a href="http://www.openmolecules.org/datawarrior/">http://www.openmolecules.org/datawarrior/</a> (ref 67)
MayaChem	MayaChem Tools	✓	✓	open-source package for property computation, library enumeration, R-group decomposition, Python scripts for PyMOL plug-ins, etc.	<a href="http://www.mayachemtools.org/">http://www.mayachemtools.org/</a> (ref 68)
KNIME	KNIME	✓	✓	open-source software for data science innovation; create workflows in GUI-like pipeline pilot	<a href="https://www.knime.com/">https://www.knime.com/</a>
AutoDock	AutoDock Vina	✓	✓	open-source suite of automated docking tools; designed to predict how small molecules, such as substrates or drug candidates, bind to a receptor of known 3D structure	<a href="http://vina.scripps.edu/">http://vina.scripps.edu/</a> (ref 69)
Open Drug Discovery Toolkit	ODDT	✓	✓	open-source modular toolkit written in Python for cheminformatics and chemical modeling	<a href="https://cheminf.biomedcentral.com/articles/10.1186/s13321-015-0078-2">https://cheminf.biomedcentral.com/articles/10.1186/s13321-015-0078-2</a> (ref 70)
Source Forge	Python Prescription (PyRx)	✓	✓	virtual screening software for computational drug discovery; screen libraries against potential targets; includes docking wizard (uses AutoDock Vina) and built on Python	<a href="https://pyrx.sourceforge.io/">https://pyrx.sourceforge.io/</a>

ing the protein target of interest, ultimately leading to the optimization of leads to clinical candidates very rapidly.<sup>4</sup> Of course, the success of generative platforms is directly proportional to the extent and accuracy of the data used in the training sets for both the generation and optimization phases. Typically, therefore, this approach works most effectively with well-trodden targets such as kinases, where a wealth of robust data is at hand to enable more accurate predictions and the most promising in silico de novo designs.<sup>5</sup> Drugging less tractable therapeutic targets, however, by targeting more challenging targets such as PROTACS and protein–protein interactions (PPIs) using this approach is significantly more challenging, as is the case for traditional medicinal chemistry approaches. However, the rate of change in the development of these generative platforms is so rapid that it will be very interesting to see how far these evolve in the coming years and whether they are able to generate novel chemical matter that can drug these challenging, high-value targets.

In addition to discussing generative models, as well as a plethora of software tools developed to generate and test a variety of hypotheses, it is important to consider methods related to prediction of chemical synthesis.<sup>6</sup> Over the past few years, several new tools have emerged (Table 3) that allow medicinal chemistry teams to quickly assess the synthesis of difficult-to-synthesize or novel chemotypes. These approaches range from ML-based methods trained on large reaction datasets (e.g., USPTO dataset) to hand-coded rules. Some of these methods are commercial, and a small handful of them are open-source (e.g., AiZynthFinder) or part of a consortium (e.g., ASKCOS). The overall outlook is favorable for these tools, and this will continue due to the continuous addition of new reaction datasets. With these tools now emerging online, it is important that each pharma company ensure that clean and reliable electronic lab notebook (ELN) data, utilizing both positive and negative data (failed reactions), is employed to train models. In addition, combining internal together with publicly available data may improve the overall quality of the models. A variety of commercial software providers are already moving in this direction, i.e., allowing the retraining of models with both internal and external datasets.

One important area of debate currently is whether to implement Web-based or desktop-based software, a deployment issue that leads to multiple concerns such as optimal machine configuration, security, etc. Routine software upgrades and updates can be a significant challenge for IT organizations to implement for each scientist, while understanding the implications of these upgrades on other software. In the past few years, most cheminformatics and molecular modeling applications have been moving to Web-based platforms. In addition, hosting applications on the “cloud” not only makes them easy for the user to access but also simplifies application deployment for the IT organizations. Over the past decade, with computer resources becoming relatively inexpensive, newer methods are being continuously developed and adopted that realize the potential of computationally intensive calculations. This raises an important issue about running some of the intensive calculations on high-performance computing (HPC) facilities available on-premises versus hosted on external infrastructures (e.g., rented within academic institutions or cloud providers). The decision to build an in-house HPC versus other options may provide a huge barrier to software acquisition from vendors and its deployment. Based

Table 4. Leading Cutting-Edge AI/ML-Based Organizations Providing Innovation in Drug Discovery

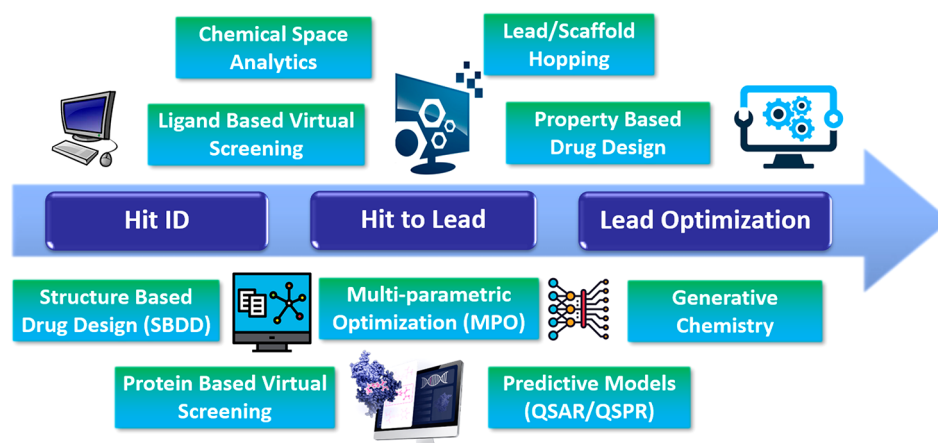
vendor	application	CI	SBDD	main function	Web site
Exscientia <sup>a</sup>	Centaur Chemist	✓	✓	uses AI to learn best-practices from drug discovery data and experienced drug hunters	<a href="https://www.exscientia.ai/">https://www.exscientia.ai/</a>
Insilico Medicine	Chemistry42	✓	✓	AI/DL generative platform for de novo drug design	<a href="https://insilico.com/chemistry42">https://insilico.com/chemistry42</a>
Atomwise <sup>a</sup>	Services Options	✓	✓	utilizes deep-learning AI technology for structure-based small-molecule drug discovery.	<a href="https://www.atomwise.com/">https://www.atomwise.com/</a> (ref 71)
Cyclica	Ligand Design, Ligand Express	✓	✓	apply a combination of computational biophysics and ML to design novel drugs	<a href="https://www.cyclicacx.com/">https://www.cyclicacx.com/</a> (ref 72)
XtalPi	RENOVA	✓	✓	generate novel drug-like molecules by combining QM and AI; predict crystal formation of small molecules	<a href="https://www.xtalpi.com/en/">https://www.xtalpi.com/en/</a>
Recursion	ReChem, RePredict, ReAnalyze	✓	✓	develop novel chemical compounds by utilizing AI to conduct experimental biology, at scale, by testing thousands of compounds on hundreds of cellular disease models in parallel	<a href="https://www.recursion.com/">https://www.recursion.com/</a>
IBM <sup>a</sup>	Generative Models	✓	✓	develop new AI methods for de novo compound generation as well as methods for retrosynthetic analysis to accelerate drug discovery	<a href="https://research.ibm.com/science/generative-models/">https://research.ibm.com/science/generative-models/</a>
1910 Genetics <sup>a</sup>	ELVIS, ROSALYND	✓	✓	discover new drugs using computation, physics, and ML; develop a robust in silico platform utilizing AI/ML methods	<a href="https://1910genetics.com/">https://1910genetics.com/</a>
Insitro <sup>a</sup>		✓	✓	Insitro is a drug discovery and development startup that utilizes ML and biology to transform drug discovery	<a href="https://insitro.com/">https://insitro.com/</a>
OneThree Biotech <sup>a</sup>		✓	✓	combining systems biology with AI to uncover new insights and build next-generation drug discovery	<a href="https://onethree.bio/">https://onethree.bio/</a> (ref 73)
Iktos	Makya, Spaya	✓	✓	AI/DL platforms for generative de novo drug design and retrosynthetic analysis	<a href="https://iktos.ai/">https://iktos.ai/</a>
DeepMind	AlphaFold	✓	✓	uses AI to accelerate drug discovery; discovered one of the most successful protein structure prediction algorithms, AlphaFold 2.0	<a href="https://deepmind.com/">https://deepmind.com/</a>
ACELLERA	PlayMolecule	✓	✓	one-click molecular discovery platform consisting of innovative MD- and DL-based tools that transcend the drug discovery continuum design paradigm	<a href="https://www.acellera.com/products/playmolecule/">https://www.acellera.com/products/playmolecule/</a>
Glamorous AI (now X-Chem)	ROSALIND <sup>AI</sup>	✓	✓	SaaS platform for AI/ML-based drug discovery; largest repository of state-of-the-art AI/ML-based models employed for applications including de novo generative design	<a href="https://www.glamorous.ai/rosalindai">https://www.glamorous.ai/rosalindai</a>
BenevolentAI <sup>a</sup>	The Benevolent Platform	✓	✓	knowledge graph-based platform for drug discovery	<a href="https://www.benevolent.com/">https://www.benevolent.com/</a>

<sup>a</sup>Drug discovery companies which may not offer access to AI/ML tools and applications.

Table 5. Leading Cutting-Edge AI/ML-Based Computer-Aided Synthetic Prediction (CASP) Organizations Providing Innovation in Drug Discovery

vendor	application	main function	Web site
ChemPass	ChemPass	rule-based AI for forward reaction-based design	<a href="https://chempassltd.com/">https://chempassltd.com/</a>
DeepMatter	ICSynth	ML-based method to generate chemistry rules	<a href="https://www.deepmatter.io/products/icsynth/">https://www.deepmatter.io/products/icsynth/</a>
Sigma-Aldrich/ Merck KGaA	SYNTHIA (previously known as Chematica)	algorithms using a database of manually generated rules	<a href="https://www.sigmaaldrich.com/US/en/services/software-and-digital-platforms/synthia-retrosynthesis-software?gclid=EAlaQobChMI94G7_Zqn8g1VE73IC1yTQjEAAAYAAAgZxfD_BwE">https://www.sigmaaldrich.com/US/en/services/software-and-digital-platforms/synthia-retrosynthesis-software?gclid=EAlaQobChMI94G7_Zqn8g1VE73IC1yTQjEAAAYAAAgZxfD_BwE</a>
IBM	IBM RXN	natural language-based model on reaction data sets	<a href="https://www.research.ibm.com/blog/rxn-cleaning-chemical-datasets">https://www.research.ibm.com/blog/rxn-cleaning-chemical-datasets</a>
CAS	ChemPlanner (previously known as ARChem)	method-based or rule-based as well as ML algorithms	<a href="https://www.cas.org/solutions/cas-scifinder-discovery-platform/cas-scifinder/retrosynthesis-planning">https://www.cas.org/solutions/cas-scifinder-discovery-platform/cas-scifinder/retrosynthesis-planning</a>
Elsevier	Reaxys	algorithms using a database of manually generated rules	<a href="https://www.elsevier.com/solutions/reaxys">https://www.elsevier.com/solutions/reaxys</a>
AstraZeneca	AIzynthFinder	open-source MonteCarlo tree-based search using ANN	<a href="https://github.com/MolecularAI/aizynthfinder">https://github.com/MolecularAI/aizynthfinder</a>
Eli Lilly Co.	LillyMol	ML-based method using model trained on reaction transformations	<a href="https://github.com/EliLillyCo/LillyMol">https://github.com/EliLillyCo/LillyMol</a>
MIT Consortia	ASKCOS	a multifaceted AI/ML-based retrosynthetic analysis tool developed at the Massachusetts Institute of Technology	<a href="https://askcos.mit.edu/help/modules">https://askcos.mit.edu/help/modules</a>
Iktos	Spaya.ai	AI-fueled tool to discover and prioritize synthetic routes	<a href="https://iktos.ai/spaya/">https://iktos.ai/spaya/</a>
University of Notre Dame	C-CAS	provides synthetic chemistry prediction using quantitative, data-driven approaches	<a href="https://ccas.nd.edu/research/#hrust-3">https://ccas.nd.edu/research/#hrust-3</a>

Scheme 1. Some Key Areas along the Drug Discovery Continuum Where Computational Applications and Tools Are Employed



on the usage and the compute nature of the software, pharma organizations should carefully plan the right architecture for software deployment, as it directly affects the users as well the IT organizations responsible for managing such software. Another aspect that is important to consider is software-as-a-service (SaaS), platform-as-a-service (PaaS), or on-premises installation of software or technology. In the past several years, many software providers have offered their solutions as either SaaS or PaaS. Pharma organizations, when making the decision to license or utilize a solution, must review this prior to capitalizing on the technology.

Both the number and the degree of computational applications and tools available to the medicinal chemist is expanding rapidly, with the rate expected to continue in the coming years. Computational tools are integral to the drug discovery continuum, and they have largely transformed the way medicinal chemistry is practiced, particularly in the past decade. In the tables we have highlighted well over 150 tools or applications from many different vendors, with over 50 focused predominantly on SBDD and over 80 focused on CI-based applications, with at least 10 having utility for both. Many of the vendors mentioned have relationships with most of the top pharmaceutical companies, and thus the drug discovery industry shares many of the same tools. It is unlikely, however, that any one organization has access to all the tools mentioned in the tables; therefore, we hope this will give a perspective on the landscape of computational tools as it stands presently. We know this list will likely be outdated by the time this article is published, but we feel that many important computational applications have been captured and that these tools will continue to help the medicinal chemist successfully navigate the drug discovery continuum for years to come.

## AUTHOR INFORMATION

### Corresponding Authors

**Philip B. Cox** – Drug Discovery Science and Technology, AbbVie, North Chicago, Illinois 60064-6217, United States; [orcid.org/0000-0001-5205-7744](https://orcid.org/0000-0001-5205-7744); Email: [phil.cox@abbvie.com](mailto:phil.cox@abbvie.com)

**Rishi Gupta** – Drug Discovery Science and Technology, AbbVie, North Chicago, Illinois 60064-6217, United States; Present Address: Novartis Institutes for BioMedical Research, 181 Massachusetts Ave., Cambridge, MA 02139, USA; Email: [rishi-1.gupta@novartis.com](mailto:rishi-1.gupta@novartis.com)

Complete contact information is available at:

<https://pubs.acs.org/10.1021/acsmmedchemlett.1c00662>

## Notes

The authors declare the following competing financial interest(s): Both P.C. and R.G. are/were employees of AbbVie at the time of this study. The design, study content, and financial support for the study were provided by AbbVie. AbbVie participated in the interpretation of data, review, and approval of the publication.

## ABBREVIATIONS

CADD, computer-aided drug design; DMTA, design–make–test–analyze; AI/ML, artificial intelligence/machine learning; SBDD, structure-based drug design; CI, cheminformatics; PK/PD, pharmacokinetics/pharmacodynamics; MPO, multiple-parameter optimization; QSAR, quantitative structure–activity relationship; QSPR, quantitative structure–property relationship; ADMET, absorption, distribution, metabolism, excretion, and toxicity; PPI, protein–protein interaction; USPTO, United States Patent and Trademarks Office; ASKCOS, automated system for knowledge-based continuous organic synthesis; ELN, electronic laboratory notebook; HPC, high-performance computing; SaaS, software as a service; PaaS, platform as a service; VLS, virtual ligand screening; QM/MM, quantum mechanics/molecular mechanics

## REFERENCES

- Warr, W. A. A CADD-alog of strategies in pharma. *J. Comput. Aided Mol. Des.* **2017**, *31* (3), 245–247. (All articles in this issue, including the ones listed here, were from various pharma organizations.) (a) McGaughey, G.; Patrick Walters, W. Modeling & Informatics at Vertex Pharmaceuticals Incorporated: our philosophy for sustained impact. *J. Comput. Aided Mol. Des.* **2017**, *31* (3), 293–300. (b) Manas, E. S.; Green, D. V. S. CADD medicine: design is the potion that can cure my disease. *J. Comput. Aided Mol. Des.* **2017**, *31* (3), 249–253. (c) Brown, F. K.; Sherer, E. C.; Johnson, S. A.; Holloway, K. M.; Sherborne, B. S. The evolution of drug design at Merck Research Laboratories. *J. Comput. Aided Mol. Des.* **2017**, *31* (3), 255–266. (d) van Vlijmen, H.; Desjarlais, R. L.; Mirzadegan, T. Computational chemistry at Janssen. *J. Comput. Aided Mol. Des.* **2017**, *31* (3), 267–273. (e) Muegge, I.; Bergner, A.; Kriegl, J. M. Computer-aided drug design at Boehringer Ingelheim. *J. Comput. Aided Mol. Des.* **2017**, *31* (3), 275–285. (f) Tsui, V.; Ortwine, D. F.; Blaney, J. M. Enabling drug discovery project decisions with integrated computa-



- tional chemistry and informatics. *J. Comput. Aided Mol. Des.* **2017**, *31* (3), 287–291.
- (2) Hoffmann, T.; Gastreich, M. The Next Level in Chemical Space Navigation: Going Far Beyond Enumerable Compound Libraries. *Drug Discovery Today* **2019**, *24* (5), 1148–1156.
- (3) Glen, R. C.; Galloway, W. R. J. D.; Spring, D. R.; Liwiki, G. Multi-Parameter Optimization in Drug Discovery: Example of the S-HT1b GPCR. *Molecular Informatics* **2016**, *35*, 599–605. Lusher, S. J.; McGuire, R.; Azevedo, R.; Boiten, J.-W.; van Schaik, R. C.; de Vlieg, J. A Molecular Informatics View on Best Practice in Multi-Parameter Compound Optimization. *Drug Discovery Today* **2011**, *16*, 555–568.
- (4) Mendez-Lucio, O.; Baillif, B.; Clevert, D.-A.; Rouquie, D.; Wichard, J. De novo Generation of hit-like molecules from gene expression signatures using artificial intelligence. *Nat. Commun.* **2020**, *11*, 10.
- (5) Tong, X.; Liu, X.; Tan, X.; Li, X.; Jiang, J.; Xiong, Z.; Xu, T.; Jiang, H.; Qiao, N.; Zheng, M. Generative Models for De Novo Drug Design. *J. Med. Chem.* **2021**, *64*, 14011–14027.
- (6) Williams, C. M.; Dallaston, M. A. The Future of Retrosynthesis and Synthetic Planning: Algorithmic, Humanistic or the Interplay? *Aust. J. Chem.* **2021**, *74*, 291–326.
- (7) Schrödinger Release 2021-1: Maestro; Schrödinger, LLC: New York, 2021.
- (8) Dixon, S. L.; Duan, J.; Smith, E.; Von Bargen, C. D.; Sherman, W.; Repasky, M. P. AutoQSAR: an automated machine-learning tool for best practice QSAR modeling. *Future Med. Chem.* **2016**, *8* (15), 1825–1839.
- (9) Watts, K. S.; Dalal, P.; Murphy, R. B.; Sherman, W.; Friesner, R. A.; Shelley, J. C. ConfGen: A Conformational Search Method for Efficient Generation of Bioactive Conformers. *J. Chem. Inf. Model.* **2010**, *50*, 534–546.
- (10) Bowers, K. J.; Chow, E.; Xu, H.; Dror, R. O.; Eastwood, M. P.; Gregersen, B.A.; Klepeis, J. L.; Kolossvary, I.; Moraes, M. A.; Sacerdoti, F. D.; Salmon, J. K.; Shan, Y.; Shaw, D. E. Scalable Algorithms for Molecular Dynamics Simulations on Commodity Clusters. Proceedings of the ACM/IEEE Conference on Supercomputing (SC06), Tampa, FL, Nov 11–17, 2006.
- (11) Abel, R.; Wang, L.; Harder, E. D.; Berne, B. J.; Friesner, R. A. Advancing Drug Discovery through Enhanced Free Energy Calculations. *Acc. Chem. Res.* **2017**, *50* (7), 1625–1632.
- (12) Friesner, R. A.; Murphy, R. B.; Repasky, M. P.; Frye, L. L.; Greenwood, J. R.; Halgren, T. A.; Sanschagrin, P. C.; Mainz, D. T. Extra Precision Glide: Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for Protein-Ligand Complexes. *J. Med. Chem.* **2006**, *49*, 6177–6196.
- (13) Miller, E.; Murphy, R.; Sindhikara, D.; Borrelli, K.; Grisewood, M.; Ranalli, F.; Dixon, S. L.; Jerome, S.; Boyles, N. A.; Day, T.; Ghanakota, P.; Mondal, S.; Rafi, S. B.; Troast, D. M.; Abel, R.; Friesner, R. A. A Reliable and Accurate Solution to the Induced Fit Docking Problem for Protein-Ligand Binding. *ChemRxiv Preprint*, **2002**, DOI: 10.26434/chemrxiv.11983845.v1.
- (14) Giardina, S. F.; Werner, D. S.; Pingle, M.; Feinberg, P. B.; Foreman, K. W.; Bergstrom, D. E.; Arnold, L. D.; Barany, F. *J. Med. Chem.* **2020**, *63* (6), 3004–3027.
- (15) Schrödinger Release 2021-1: MacroModel; Schrödinger, LLC: New York, 2021.
- (16) Dixon, S. L.; Smondyrev, A. M.; Knoll, E. H.; Rao, S. N.; Shaw, D. E.; Friesner, R. A. PHASE: A New Engine for Pharmacophore Perception, 3D QSAR Model Development, and 3D Database Screening. 1. Methodology and Preliminary Results. *J. Comput. Aided Mol. Des.* **2006**, *20*, 647–671.
- (17) PyMOL Molecular Graphics System, Version 2.0; Schrödinger, LLC: New York, 2017.
- (18) Murphy, R. B.; Philipp, D. M.; Friesner, R. A. A mixed quantum mechanics/molecular mechanics (QM/MM) method for large-scale modeling of chemistry in protein environments. *J. Comput. Chem.* **2000**, *21*, 1442–1457.
- (19) Sastry, G. M.; Dixon, S. L.; Sherman, W. Rapid Shape-Based Ligand Alignment and Virtual Screening Method Based on Atom/Feature-Pair Similarities and Volume Overlap Scoring. *J. Chem. Inf. Model.* **2011**, *51*, 2455–2466.
- (20) Schrödinger Release 2021-1: Core Hopping; Schrödinger, LLC: New York, 2021.
- (21) Salam, N. K.; Nuti, R.; Sherman, W. Novel Method for Generating Structure-Based Pharmacophores Using Energetic Analysis. *J. Chem. Inf. Model.* **2009**, *49*, 2356–2368.
- (22) Schrödinger Release 2021-1: Field-based QSAR; Schrödinger, LLC: New York, 2021.
- (23) Bochevarov, A. D.; Harder, E.; Hughes, T. F.; Greenwood, J. R.; Braden, D. A.; Philipp, D. M.; Rinaldo, D.; Halls, M. D.; Zhang, J.; Friesner, R. A. Jaguar: A high-performance quantum chemistry software program with strengths in life and materials sciences. *Int. J. Quantum Chem.* **2013**, *113* (18), 2110–2142.
- (24) Scarbath-Evers, K.; Cappel, D.; Weiser, J. Digitalisierung: molekulares Design plattformisieren. *Nachr. Chem.* **2020**, *130*, 34–36.
- (25) Schrödinger Release 2021-1: QikProp; Schrödinger, LLC: New York, 2021.
- (26) Abel, R.; Young, T.; Farid, R.; Berne, B. J.; Friesner, R. A. Role of the active-site solvent in the thermodynamics of Factor Xa ligand binding. *J. Am. Chem. Soc.* **2008**, *130*, 2817–2831.
- (27) Halgren, T. Identifying and Characterizing Binding Sites and Assessing Druggability. *J. Chem. Inf. Model.* **2009**, *49*, 377–389.
- (28) Schrödinger Release 2021-1: Canvas; Schrödinger, LLC: New York, 2021.
- (29) Greenwood, J. R.; Calkins, D.; Sullivan, A. P.; Shelley, J. C. Towards the comprehensive, rapid, and accurate prediction of the favorable tautomeric states of drug-like molecules in aqueous solution. *J. Comput. Aided Mol. Des.* **2010**, *24*, 591–604.
- (30) Zhu, K.; Borrelli, K. W.; Greenwood, J. R.; Day, T.; Abel, R.; Farid, R. S.; Harder, E. Docking covalent inhibitors: A parameter free approach to pose prediction and scoring. *J. Chem. Inf. Model.* **2014**, *54*, 1932–1940.
- (31) Davis, A.; Warrington, B. H.; Vinter, J. G. Strategic approaches to drug design II. Modelling studies on phosphodiesterase substrates and inhibitors. *J. Comput. Aided Mol. Des.* **1987**, *1*, 97–119.
- (32) Cheeseright, T.; Mackey, M.; Rose, S.; Vinter, A. Molecular Field Extrema as Descriptors of Biological Activity: Definition and Validation. *J. Chem. Inf. Model.* **2006**, *46* (2), 665–676.
- (33) Bauer, M. R.; Mackey, M. D. Electrostatic Complementarity as a Fast and Effective Tool to Optimize Binding and Selectivity of Protein–Ligand Complexes. *J. Med. Chem.* **2019**, *62* (6), 3036–3050.
- (34) Cheeseright, T. J.; Mackey, M. D.; Melville, J. L.; Vinter, J. G. FieldScreen: Virtual Screening Using Molecular Fields. Application to the DUD Data Set. *J. Chem. Inf. Model.* **2008**, *48* (11), 2108–2117.
- (35) Cresset Software, Lead Finder. BioMolTech, Toronto, Canada; <http://www.cresset-group.com/lead-finder/>
- (36) Rush, T. S.; Grant, J. A.; Mosyak, L.; Nicholls, A. A shape-based 3-D scaffold hopping method and its application to a bacterial protein-protein interaction. *J. Med. Chem.* **2005**, *48* (5), 1489–1495.
- (37) Grant, J. A.; Haigh, J. A.; Pickup, B. T.; Nicholls, A.; Sayle, R. A. Lingos, Finite State Machines, and Fast Similarity Searching. *J. Chem. Inf. Model.* **2006**, *46* (5), 1912–1918.
- (38) Vidal, D.; Thormann, M.; Pons, M. LINGO, an Efficient Holographic Text Based Method to Calculate Biophysical Properties and Intermolecular Similarities. *J. Chem. Inf. Model.* **2005**, *45* (2), 386–393.
- (39) Cannon, E. O. New Benchmark for Chemical Nomenclature Software. *J. Chem. Inf. Model.* **2012**, *52* (5), 1124–1131.
- (40) Halgren, T. A. Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94. *J. Comput. Chem.* **1996**, *17*, 490–519.
- (41) McGann, M. FRED Pose Prediction and Virtual Screening Accuracy. *J. Chem. Inf. Model.* **2011**, *51* (3), 578–596.
- (42) Hawkins, P. C. D.; Skillman, A. G.; Warren, G. L.; Ellingson, B. A.; Stahl, M. T. Conformer Generation with OMEGA: Algorithm and Validation Using High Quality Structures from the Protein Databank and Cambridge Structural Database. *J. Chem. Inf. Model.* **2010**, *50*, 572–584.



- (43) Haigh, J. A.; Pickup, B. T.; Grant, J. A.; Nicholls, A. Small molecule shape-fingerprints. *J. Chem. Inf. Model.* **2005**, *45*, 673.
- (44) Grant, J. A.; Pickup, B.; Nicholls, A. Smooth Permittivity Function for Poisson-Boltzmann Solvation Methods. *J. Comput. Chem.* **2001**, *22*, 608–640.
- (45) Wlodek, S.; Skillman, A. G.; Nicholls, A. Ligand entropy in gas-phase, upon solvation and protein complexation. Fast estimation with Quasi-Newton Hessian. *J. Chem. Theory Comput.* **2010**, *6* (7), 2140–2152.
- (46) SeeSAR, Version 10.3.3; BioSolveIT GmbH: Sankt Augustin, Germany, 2021. [www.biosolveit.de/SeeSAR](http://www.biosolveit.de/SeeSAR)
- (47) infiniSee, Version 2.2.2; BioSolveIT GmbH: Sankt Augustin, Germany, 2021. [www.biosolveit.de/infiniSee](http://www.biosolveit.de/infiniSee)
- (48) Warren, G. L.; Andrews, C. W.; Capelli, A. M.; Clarke, B.; LaLonde, J.; Lambert, M. H.; Lindvall, M.; Nevins, N.; Semus, S. F.; Senger, S.; Tedesco, G.; Wall, I. D.; Woolven, J. M.; Peishoff, C. E.; Head, M. S. A Critical Assessment of Docking Programs and Scoring Functions. *J. Med. Chem.* **2006**, *49* (20), 5912–5931.
- (49) Reulecke, I.; Lange, G.; Albrecht, J.; Klein, R.; Rarey, M. Towards an Integrated Description of Hydrogen Bonding and Dehydration: Decreasing False Positives in Virtual Screening with the HYDE Scoring Function. *ChemMedChem* **2008**, *3* (6), 885–897.
- (50) Lemmen, C.; Hiller, C.; Lengauer, T. RigFit: A New Approach to Superimposing Ligand Molecules. *J. Comput. Aided Mol. Des.* **1998**, *12* (5), 491–502.
- (51) Boehm, M.; Wu, T.-Y.; Claussen, H.; Lemmen, C. Similarity Searching and Scaffold Hopping in Synthetically Accessible Combinatorial Chemistry Spaces. *J. Med. Chem.* **2008**, *51*, 2468–2480.
- (52) Molecular Operating Environment (MOE) 2019.01; Chemical Computing Group: Montreal, Canada, 2021.
- (53) Mayfield, J.; Sayle, R. A. The Secrets of Fast SMARTS Matching. 8th Joint Sheffield Conference on Chemoinformatics, Sheffield, UK, June 17–10, 2019.
- (54) O'Boyle, N. M.; Boström, J.; Sayle, R. A.; Gill, A. Using Matched Molecular Series as a Predictive Tool to Optimize Biological Activity. *J. Med. Chem.* **2014**, *57*, 2704.
- (55) Patsy. Presented at the 9th International Conference on Chemical Structures (ICCS), Noordwijkerhout, The Netherlands, June 5–9, 2011.
- (56) Smallworld. Presented at SCI's "What Can Big Data Do for Chemistry?", Royal Society of Chemistry, London, UK, Oct 11, 2017.
- (57) CaffeineFix v2.0. Presented at the 244th American Chemical Society National Meeting, Philadelphia, PA, Aug 19–23, 2012.
- (58) LeadMine v2.0. Presented at the 244th American Chemical Society National Meeting, Philadelphia, PA, Aug 19–23, 2012.
- (59) Sayle, R. A.; May, J. Pharmaceutical industry best practices in lessons learned: ELN implementation of Merck's reaction review policy. Presented at the 254th American Chemical Society National Meeting, Washington, DC, Aug 20–24, 2017.
- (60) HazELNut. Presented at the 250th American Chemical Society National Meeting, Boston, MA, Aug 16–19, 2015.
- (61) Schneider, N.; Lowe, D. N.; Sayle, R.; Tarselli, M.; Landrum, G. Big Data from Pharmaceutical Patents: A Computational Analysis of Medicinal Chemists' Bread and Butter. *J. Med. Chem.* **2016**, *59* (9), 4385–4402.
- (62) Mayfield, J. Pistachio: Search and faceting of large reaction databases. Presented at the 254th American Chemical Society National Meeting, Washington, DC, Aug 20–24, 2017.
- (63) Brodney, M. D.; Brosius, A. D.; Gregory, T.; Heck, S. D.; Klug-McLeod, J. L.; Poss, C. S. Project-focused activity and knowledge tracker: a unified data analysis, collaboration, and workflow tool for medicinal chemistry project teams. *J. Chem. Inf. Model.* **2009**, *49* (12), 2639–2649.
- (64) Hu, Q.; Peng, Z.; Sutton, S. C.; Na, J.; Kostrowicki, J.; Yang, B.; Thacher, T.; Kong, X.; Mattaparti, S.; Zhou, J. Z.; Gonzalez, J.; Ramirez-Weinhouse, M.; Kuki, A. Pfizer Global Virtual Library (PGVL): a chemistry design tool powered by experimentally validated parallel synthesis information. *ACS Comb. Sci.* **2012**, *14* (11), 579–89.
- (65) Luty, B.; Rose, P. W. The need for scientific software engineering in the pharmaceutical industry. *J. Comput. Aided Mol. Des.* **2017**, *31*, 301–304.
- (66) Stiefl, N.; Gedeck, P.; Chin, D.; Hunt, P.; Lindvall, M.; Spiegel, K.; Springer, C.; Biller, S.; Buenemann, C.; Kanazawa, T.; Kato, M.; Lewis, R.; Martin, E.; Polyakov, V.; Tommasi, R.; van Drie, J.; Vash, B.; Whitehead, L.; Xu, Y. J.; Abagyan, R.; Rausch, E.; Totrov, M. FOCUS - development of a global communication and modeling platform for applied and computational medicinal chemists. *J. Chem. Inf. Model.* **2015**, *55*, 896–908.
- (67) Sander, T.; Freyss, J.; von Korff, M.; Rufener, C. DataWarrior: An Open-Source Program for Chemistry Aware Data Visualization and Analysis. *J. Chem. Inf. Model.* **2015**, *55*, 460–47.
- (68) Sud, M. MayaChemTools: an open-source package for computational drug discovery. *J. Chem. Inf. Model.* **2016**, *56*, 2292–2297.
- (69) 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.* **2010**, *31*, 455–461.
- (70) Wójcikowski, M.; Zielenkiewicz, P.; Siedlecki, P. Open Drug Discovery Toolkit (ODDT): a new open-source player in the drug discovery field. *J. Cheminform.* **2015**, *7*, 26.
- (71) Wallach, I.; Dzamba, M.; Heifets, A. Atomnet: A deep convolutional neural network for bioactivity prediction in structure-based drug discovery. *arXiv Preprint* **2015**, DOI: [10.48550/arXiv.1510.02855](https://doi.org/10.48550/arXiv.1510.02855).
- (72) Somody, J. C.; MacKinnon, S. S.; Windemuth, A. Structural coverage of the proteome for pharmaceutical applications. *Drug Discovery Today* **2017**, *22*, 1792–1799.
- (73) Madhukar, N. S.; Khade, P. K.; Huang, L.; Gayvert; Galletti, G.; Stogniew, M.; Allen, J. E.; Giannakakou, P.; Elemento, O. A. Bayesian machine learning approach for drug target identification using diverse data types. *Nat. Commun.* **2019**, *10*, 5221.