

# Computational Modeling of $\beta$ -Secretase 1 (BACE-1) Inhibitors Using Ligand Based Approaches

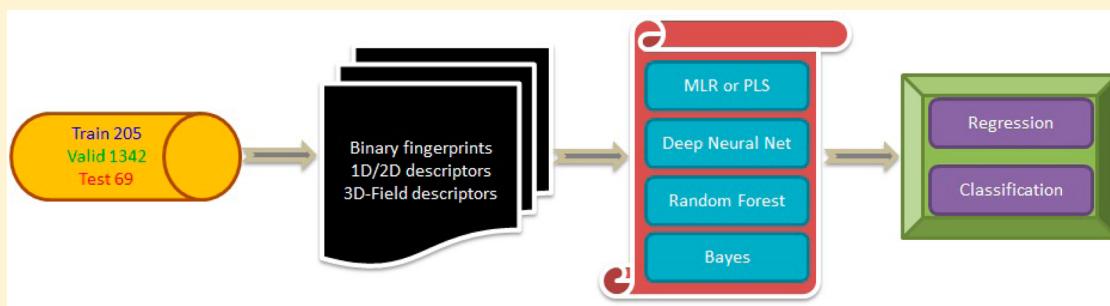
Govindan Subramanian,<sup>\*§</sup> Bharath Ramsundar,<sup>‡</sup> Vijay Pande,<sup>‡</sup> and Rajiah Aldrin Denny<sup>†</sup>

<sup>§</sup>VMRD Global Discovery, Zoetis, 333 Portage Street, Kalamazoo, Michigan 49007, United States

<sup>‡</sup>Department of Computer Science and <sup>‡</sup>Department of Chemistry, Stanford University, 318 Campus Drive, Stanford, California 94305, United States

<sup>†</sup>Worldwide Medicinal Chemistry, Pfizer Inc., 610 Main Street, Cambridge, Massachusetts 02139, United States

## Supporting Information



**ABSTRACT:** The binding affinities ( $IC_{50}$ ) reported for diverse structural and chemical classes of human  $\beta$ -secretase 1 (BACE-1) inhibitors in literature were modeled using multiple in silico ligand based modeling approaches and statistical techniques. The descriptor space encompasses simple binary molecular fingerprint, one- and two-dimensional constitutional, physicochemical, and topological descriptors, and sophisticated three-dimensional molecular fields that require appropriate structural alignments of varied chemical scaffolds in one universal chemical space. The affinities were modeled using qualitative classification or quantitative regression schemes involving linear, nonlinear, and deep neural network (DNN) machine-learning methods used in the scientific literature for quantitative–structure activity relationships (QSAR). In a departure from tradition, ~20% of the chemically diverse data set (205 compounds) was used to train the model with the remaining ~80% of the structural and chemical analogs used as part of an external validation (1273 compounds) and prospective test (69 compounds) sets respectively to ascertain the model performance. The machine-learning methods investigated herein performed well in both the qualitative classification (~70% accuracy) and quantitative  $IC_{50}$  predictions (RMSE  $\sim 1 \log$ ). The success of the 2D descriptor based machine learning approach when compared against the 3D field based technique pursued for hBACE-1 inhibitors provides a strong impetus for systematically applying such methods during the lead identification and optimization efforts for other protein families as well.

## INTRODUCTION

In silico computer assisted drug design (CADD) plays an integral part in several areas of preclinical small molecule drug discovery.<sup>1</sup> Ligand- and structure-based modeling approaches enable small molecule scientific discovery research teams to prioritize the compound progression pathways from the initial hit identification to the final clinical candidate nomination in the organization's research and development portfolio.<sup>2–4</sup> One of the key parameters in the optimization path to the development candidate involves the small molecule affinity improvement to its intended therapeutic target.

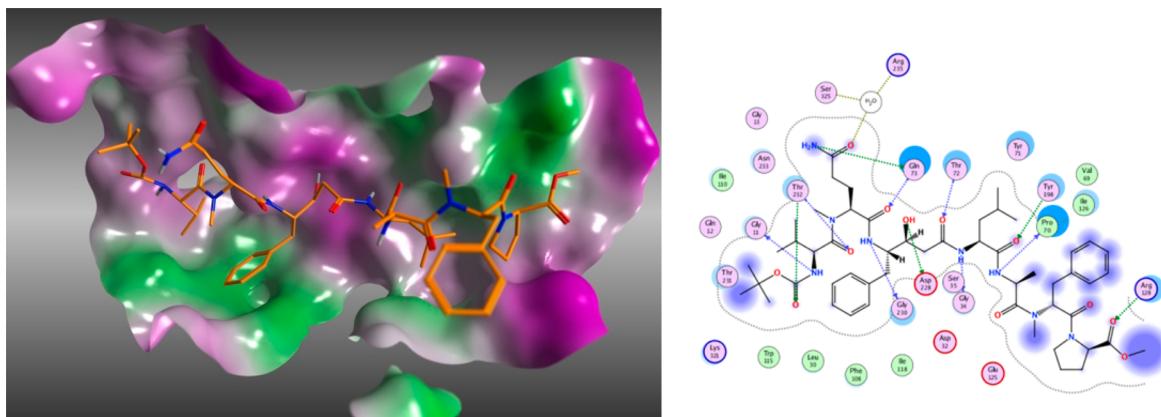
When the structure of the small molecule bound to its protein is available, high bias is often used to apply structure-based design (SBD) techniques<sup>5</sup> to guide further affinity improvement efforts, since the X-ray structure of the complex provides a great deal of insight into the key drivers modulating affinity. In fact, SBD approaches like docking work under the

premise of the identified or hypothesized binding site.<sup>6,7</sup> The shape and curvature of the binding pocket as well as the computed molecular grids representing them determine the plausible orientation and fit of the synthetic compound under consideration, resulting in qualitative rank ordering of in silico screened small molecule ligands binding to the target.<sup>8</sup> Thereafter, high end molecular dynamics simulation is used to assess the binding affinity quantitatively and more accurately ( $\Delta G_{bind}$ ),<sup>9–12</sup> with implicit solvation methods currently offering a trade-off for computational cost and accuracy in predicting binding affinity. Among the shortcomings of the SBD approach, the plasticity of the protein active site residues upon ligand binding (induced fit) creates uncertainties that could not be handled straightforwardly by the use of a single protein–ligand

Received: May 23, 2016

Published: September 30, 2016

**Scheme 1.** Depiction of BACE-1 Binding Site (left) Using the Ligand from PDB Code 3UQP along with the Protein–Ligand Interaction (right)



complex structure.<sup>13,14</sup> This necessitates the use of multiple protein structures (ensemble docking) to understand and predict ligand affinities, a process that often adds noise and computational cost when compared to the potential benefit.<sup>15</sup> Even though there has been a steady increase in the use of SBD approaches for quantitative affinity prediction,<sup>16,17</sup> the approaches are still considered to be in their infancy and are often unable to provide consistent accuracy across a range of targets to be applied routinely for rank ordering compounds based on their affinity.<sup>10</sup>

On the other hand, ligand-based design (LBD) methods with QSAR<sup>18,19</sup> statistical fits dominated the field during the last six decades due to the scarcity of solved ligand bound protein X-ray structures. These methods use common features in the chosen ligand set without any *a priori* knowledge of the environment surrounding the small molecule. Traditional LBD approaches like the pharmacophore<sup>20,21</sup> and molecular field<sup>22</sup> based model employing molecular alignments uses known experimental activity information from a preselected compound set to guess the potential microenvironment around the ligand and applies a statistical fit to predict the activity of the unknowns or new design modifications. The sheer publication and citation volumes using these techniques stand as a testament to the success and acceptance of such *pseudoreceptor* approaches by the CADD community.

Although the strengths and limitations of the two major computational approaches described above and routinely used in *in silico* drug discovery could be debated, the complementarity of the SBD and LBD approaches can enable the field to progress further by using their hybrid.<sup>23–26</sup> For instance, the ligands binding mode to their intended target can be identified using SBD methods or X-ray structures of ligand bound protein complex and a quantitative evaluation of the affinities computed using established 3D QSAR approaches<sup>27,28</sup> by overlaying related structural analogs (using maximum common substructure) to the bound orientations of the ligand in the X-ray structure or via docking protocols. We and other groups have started to realize enhanced benefits using such combinations with high success rates in prediction.<sup>29–33</sup> Herein, we demonstrate that such an approach can now be applied to systems with a very limited data set size and diverse chemical scaffolds. To compare the effectiveness of such hybrid approaches, multiple LBD techniques, starting from simple Bayesian methods to sophisticated machine-learning methods, were applied for the same data set to assess the computational

ease vs prediction accuracy as the feature complexity (1D to 3D) and information content (fingerprints to molecular fields) were incremented gradually.

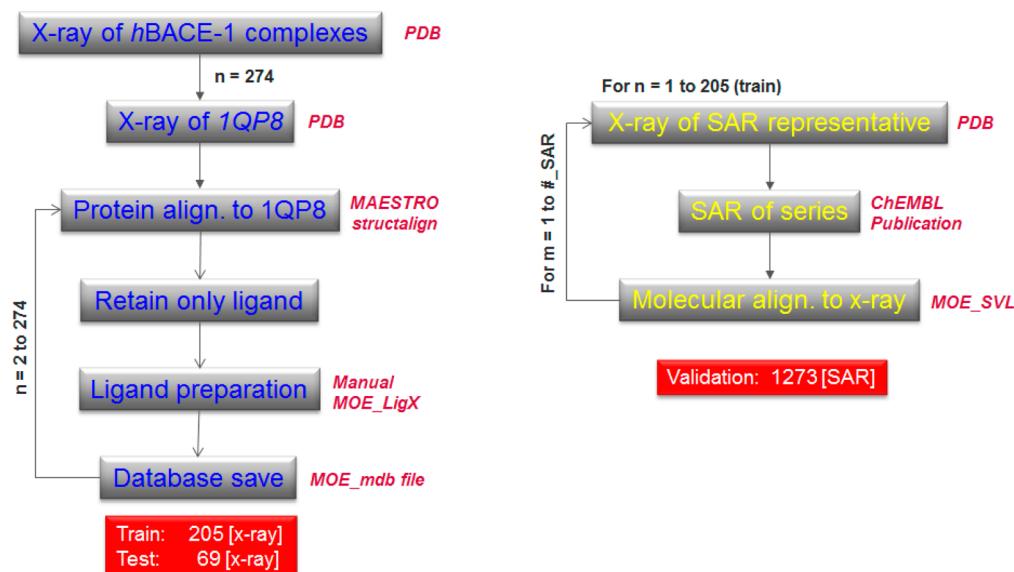
The target selected for this computational exercise is BACE-1, a long sought potential therapeutic target for Alzheimer's.<sup>34,35</sup> This aspartyl protease is considered to be the key driver for the brain amyloid  $\beta$  accumulation and has eluded pharmaceutical and biotechnology companies alike in delivering a small molecule drug for over a decade.<sup>36,37</sup> The BACE-1 orthosteric active site utilizes two aspartate residues (D32 and D228) to anchor endogenous substrates or synthetic ligands (**Scheme 1**) and is located in between the N- and C-terminal domain.<sup>38</sup> The shallow and highly exposed binding site is shielded by an extremely flexible  $\beta$  hairpin loop (Y68–A77) that adapts multiple flap conformations<sup>39</sup> depending on the bound ligand that either disturbs or perturbs the active site water network based on the nature (closed/open) of the flap configuration.<sup>40,41</sup> In addition, the BACE-1 activity shifts depending on the pH as well as the protonation states of the catalytic dyad.<sup>42,43</sup> Such complexities have rendered it very difficult to model and interpret the effect of the thermodynamic contributions (enthalpy, entropy, etc.) obtained through biophysical experimental measurements to the small molecule BACE-1 inhibitor binding.<sup>44</sup>

Although scientific rigor alongside computational and experimental interplay can enable a much deeper understanding of the key factors influencing binding, they are more often restricted to the specific chemical series that was used for interrogation in the binding affinity optimization cycle. It is seldom that such learning can be extended widely and hence provides limited value to the broader scientific community. To circumvent such limitations and expand the scope to address several BACE-1 inhibitor chemical and structural classes, multiple LBD approaches were pursued to identify the sweet spot between computational rigor and trade-off for meaningful compromise that would allow rapid and routine evaluation of new compound ideas in the day-to-day iterative lead optimization cycle.

## METHODS

The current data set includes 1547 synthetic human BACE-1 inhibitors reported in scientific literature over the past decade with protein–ligand X-ray structures available for over 250 of them in the PDB.<sup>45</sup> Experimental IC<sub>50</sub>'s are also reported for 1486 compounds, and K<sub>d</sub>/K<sub>i</sub> values are reported for another 43.

**Scheme 2.** Workflow for the Training, Test, and Validation Set Compound Alignment Used for 3D-Field Based Approaches



The data set covers compounds originating from at least 30 different laboratories that include academia, biotechnology, and pharmaceutical companies and spans across the globe in terms of their origin, cautioning on the inhomogeneity of experimental observation<sup>46</sup> used for statistical modeling ([Supporting Information S1](#)). The structures and reported activities for all the compounds were retrieved from either ChEMBL<sup>47</sup> or from the original source publication using the citation information from PDB. In assembling the ligand data set, only publications that reported a solved X-ray ligand–BACE-1 structure was considered as this provides a reasonable premise to guess on the potential binding mode (bioactive conformation) of the closely related medicinal chemistry derived analogs reported in the same article as part of the structure–activity relationship (SAR). We have made the ligand data set publicaly available in a convenient CSV format.<sup>48</sup>

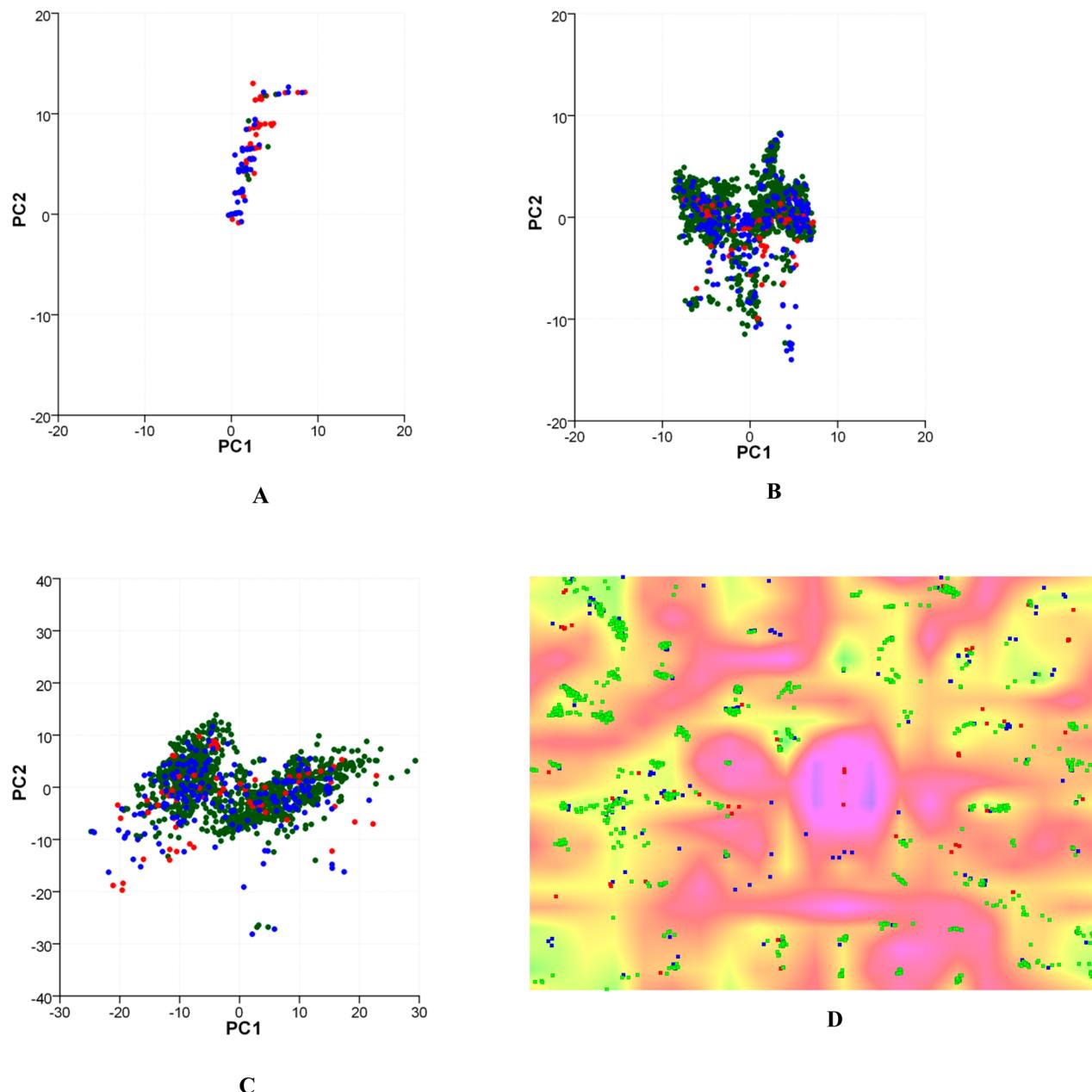
The physicochemical attributes for the 1547 ligand data set were computed using BIOVIA's Pipeline Pilot<sup>49</sup> and vary between 138 and 1350 for molecular weight (MW); -6.6 and 7.4 for logD; -5.1 and 7.6 for AlogP; 16 and 525 for polar surface area (PSA); 0 and 18 for H-bond acceptor (HBA); 1 and 15 for H-bond donor (HBD); 0 and 40 rotatable bonds; 10 and 90 for heavy atom counts for the reported experimental pIC<sub>50</sub>'s ranging from 2.5 to 9.5. As would be apparent, the chemistry captures very small synthetic fragments and large molecular weight oligopeptide and their mimics. The ligand efficiency (LE) varies between 0.2 and 0.5, binding efficiency index (BEI) ranges from 8 to 22, and the surface efficiency index (SEI) mostly between 4 and 12.<sup>50</sup> A distribution of the various physicochemical property profiles (*Supporting Information S2*) reveals that the data set attributes are not skewed and span the individual space with a Gaussian spread.

The global 3D alignment of all the analogs from multiple chemical scaffolds was achieved by following the steps outlined in **Scheme 2**. Chain A of 2QP8 (PDB code) was used as the reference, as this structure had one of the best available resolutions (1.5 Å) reported when this work was initiated. Each of the reported PDB structures for the BACE-1··inhibitor X-ray complex was structurally aligned to the above reference using the *structalign* script available within the Schrödinger's

MAESTRO suite of programs.<sup>51</sup> The protein, metal ions, and solvent were subsequently removed from the aligned structures and only one ligand structure from each PDB entry was retained for further processing. The structure data file built based on protein superposition (containing all the aligned ligands) was uploaded to the molecular database spreadsheet within MOE<sup>52</sup> from Chemical Computing Group. The ligands were subsequently cleaned (bond orders fixed) and prepared by the addition of hydrogens and molecular charges as deemed appropriate. This procedure now provides an alignment of the diverse ligands in a universal chemical space as would be seen by the protein target. It also minimizes the pharmacophore and shape bias elicited by the molecular superposition algorithms, and underpins the notion that structurally similar ligands need not bind in a similar way.<sup>29</sup>

The 274 compounds assembled in this manner (X-ray of aligned hBACE-1...ligand complex) formed the basis of the training (205) and test (69) sets respectively and were used for the 3D molecular alignment dependent approaches pursued in this study. Since the *bioactive* conformation for each ligand is now identified (based on the X-ray complex), all medicinal chemistry derived structural analogs within each scaffold reported as part of the SAR exploration in the respective publication were aligned to each individual reference X-ray structures (**Scheme 2**, right) to yield the final 1273 validation set analogs that did not have an experimental X-ray structure disclosed in public literature. No similarity cutoff threshold was used for the SAR compounds since these were standard R-group analogs belonging to the same chemotype that is represented by the X-ray structure reported in the same publication.

The alignment protocol follows the procedure described earlier in literature.<sup>29</sup> Specifically, the ligand coordinates extracted from the BACE-1 complex X-ray structure reported in each individual publication was used as the reference frame (**Scheme 2**, right) and the SAR analogs reported in that publication were superposed to the chosen reference using the *fragment\_superpose* and *flexalign\_manual* svl scripts available from MOE. Further refinement was achieved through subtle constrained minimization of each of the aligned ligand to



**Figure 1.** PCA plot for the 1547 BACE-1 inhibitors. Training, validation, and test set compounds are colored in blue, green, and red, respectively. (A) Tier 2 Bemis–Murcko framework. (B) ECFP6 fingerprints. (C) Canvas descriptors. (D) Scaffold grouping on a SOM for the training (blue), validation (green), and test (red) set compounds.

achieve maximum overlay with the reference structure per the *refine* option implemented within the *flexalign\_manual.svl* script. This procedure of choosing the individual X-ray reference and SAR analogs was iterated for each publication (see S1 Supporting Information) and the consolidated alignment of all the reference compounds (now part of the training set) and the SAR analogs (now part of the validation set) comprise the full data set used for the modeling exercise.

An additional 69 ligand bound X-ray structures were grouped as the prospective test set since these complexes do not have reported IC<sub>50</sub> values, but some of them have either K<sub>i</sub>/K<sub>d</sub> disclosed instead. The predicted binding affinities would then offer a decent starting point if scaffold hopping exercise need to be pursued for any of these chemical series. Succinctly, the published X-ray of BACE-1···ligand complex constituted the

source of the 205 training set compounds and the reported SAR in scientific literature formed the basis of the 1273 validation set molecules such that the latter has at least one close representative in the training set.

In order to ascertain if such strenuous computational rigor described above is required to obtain statistically meaningful and robust predictive QSAR models, additional LBD approaches were also pursued. These include the standard QSAR approaches that use the traditional binary fingerprints or molecular descriptors as computed by Canvas from Schrödinger.<sup>53</sup> For instance, the 1000 most informative bits from linear, radial, dendritic, and MolPrint2D fingerprints<sup>54,55</sup> were used to develop qualitative classification models while the constitutional, physicochemical, and topological descriptors computed using Canvas were used to build quantitative

regression models. Both the qualitative and quantitative models included linear, nonlinear, and deep learning techniques.<sup>56</sup> Such a thorough coverage of the interdependence of data set, descriptor, and statistical approach for one protein target represented by diverse chemical scaffolds should provide standard baselines when approaching similar problems in a lead identification or optimization cycle within a preclinical drug discovery setting.

## ■ RESULTS

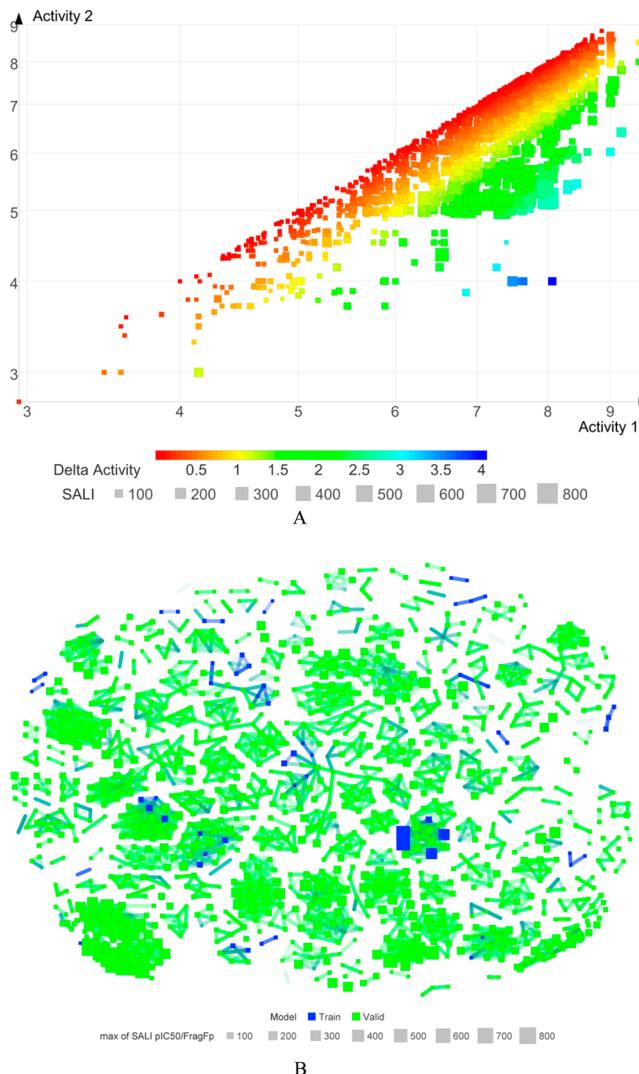
**Chemistry Space.** Prior to analyzing the outcomes from the different approaches and models, it is pertinent to chart out the chemical space occupied by the training and validation set BACE-1 inhibitors as this would provide some granularity for the well-recognized model applicability domain concerns.<sup>57</sup> Since the model generation involved multiple descriptor features, the chemical space was generated using the same feature space. The first three principal components (PC) were computed using extended connectivity fingerprint (ECFP6),<sup>58</sup> Canvas constitutional, physicochemical, and topological descriptors,<sup>54</sup> as well as Bemis–Murcko Tier-1 (topography with *atomtype* annotation) or Tier-2 (topography) molecular framework scheme.<sup>59</sup>

Figure 1 depicts the scatter plot of the PC1 and PC2 space derived from some of the procedures mentioned above. As would be expected, the Tier-2 reflects the minimal diversity points in the compound space since this represents the 2-dimensional (2D) shape and geographical topology of the varying chemotypes considered in this study. The 2D skeleton terrain sheds light on the chemical framework heterogeneity (Figure 1a) within a dimensionality reduction visualization context, with the expected homogeneous distribution (overlap) for the validation set around the training set compounds. An alternate perspective obtained using the radial ECFP6 fingerprints captures the subtle and varied contours and represents a few clusters (Figure 1b). A pairwise comparison of the training set compounds using ECFP6 fingerprint and Tanimoto similarity computed via Canvas is provided in Supporting Information S3. The macroscopic chemistry space view obtained by computing the various constitutional, physicochemical, and topological properties together reflect on the broad property based diversity of the inhibitors considered. The scatter distribution, being continuous and not overly islandic, also provides enough motivation to have confidence in the models being developed. The plots also highlight that the positioning of the validation set in comparison to the training set compounds is not unreasonable to raise concerns in the predictive capability (model applicability domain) of the models being developed using either the fingerprint or descriptor space (Figure 1c).

Two dimensional self-organizing map (SOM)<sup>60</sup> is another dimensionality reduction technique that provides an orthogonal viewpoint compared to the previous approaches.<sup>61</sup> The outcome from SOM using the Canvas descriptors is shown in Figure 1d. It should be noted that the SOM implementation in DataWarrior,<sup>62</sup> uses *Rubberbanding force field* approach that translates the similarity better than the traditional PCA. Since the SOM clustering is based on similar compound pairs and neighborhood relations, the diversity of the chemical scaffolds considered in this study is reminiscent in the distinction shown between the topological neighbors and close neighbors. The SOM depiction in Figure 1d also shows that majority of the training and validation set compounds have clear topographical

neighbors, while the training and test set compounds offer additional close neighbors that can potentially be exploited for SAR exploration, if needed.

**Biological Property Landscape.** Although the unsupervised approaches for modeling the chemical space present a clean picture of confidence for pursuing modeling activities, additional evaluations were undertaken on the data set to delineate potential disconnects in the biological property space being modeled when relatively minimal changes are incorporated to the structure. The *SkeletonSphere* descriptor scheme implemented in DataWarrior<sup>62</sup> was used for the chemical descriptions of the training and validation data sets. The structure activity landscape index (SALI)<sup>63</sup> that provides a disconnect measure between the biological observation ( $\Delta pIC50$ ) and the extent of chemical diversity (1 – similarity) for each compound pair was subsequently computed and incorporated within Figure 2a of the affinity ( $pIC50$ ) scatter plot for each of the molecule pair being compared.



**Figure 2.** (a) SALI plot of compound pairs (training and validation) with >95% similarity. The  $x$  and  $y$  axis represent the experimental  $pIC50$  activity values of the compound pairs being plotted (see S4 Supporting Information). (b) Activity cliffs (marker size) for the training and validation set grouped based on their neighborhood similarity relationships.

Table 1. Statistical Measures for the Various Classification Models Developed in This Work

Descriptor <sup>c</sup>	Approach	Training Set <sup>a</sup>				Validation Set <sup>b</sup>			
		Accuracy <sup>d</sup>	Sensitivity <sup>e</sup>	Specificity <sup>f</sup>	MCC <sup>g</sup>	Accuracy	Sensitivity	Specificity	MCC
Linear	Bayes <sup>h</sup>	0.75	0.98	0.51	0.56	0.56	0.98	0.24	0.31
ECFP6	Bayes <sup>h</sup>	0.83	0.94	0.73	0.69	0.6	0.61	0.59	0.20
Dendritic	Bayes <sup>h</sup>	0.74	0.97	0.51	0.54	0.55	0.98	0.23	0.30
MolPrint2D	Bayes <sup>h</sup>	0.93	0.97	0.89	0.87	0.64	0.87	0.46	0.35
Canvas <sup>i</sup>	RP <sup>j</sup>	0.90	0.95	0.85	0.81	0.62	0.78	0.5	0.29
Canvas <sup>i</sup>	RF <sup>k</sup>	0.99	1.00	0.98	0.98	0.68	0.93	0.49	0.46
Canvas <sup>i</sup>	DNN <sup>l</sup>	0.90	0.95	0.86	0.81	0.69	0.92	0.52	0.46
ECFP6	Bayes <sup>h,m</sup>	0.79	0.90	0.70	0.60	0.76	0.87	0.67	0.54
Canvas <sup>i</sup>	RF <sup>k,m</sup>	0.99	0.99	0.99	0.99	0.81	0.77	0.84	0.61
Canvas <sup>i</sup>	DNN <sup>l,m</sup>	0.88	0.85	0.90	0.75	0.82	0.75	0.88	0.64

<sup>a</sup>Training set (205): experimentally active (102) with  $IC_{50} \leq 100$  nM; experimentally inactive (103). <sup>b</sup>Validation set (1273): experimentally active (551) with  $IC_{50} \leq 100$  nM; experimentally inactive (722). <sup>c</sup>Fingerprint and descriptors as implemented within Canvas modeling suite from Schrödinger. <sup>d</sup>(TP + TN)/total no. molecules where TP and TN correspond to true positives and true negatives. <sup>e</sup>TP/(TP + FN) where FN correspond to false negatives. <sup>f</sup>TN/(TN + FP) where FP correspond to false positives. <sup>g</sup>Matthews correlation coefficient, MCC = [(TP\*TN - FP\*FN)/ $\sqrt{((TP+FP)(TP+FN)(TN+FP)(TN+FN))}$ ]. <sup>h</sup>Model developed using Bayesian approach as implemented within Canvas modeling suite from Schrödinger. <sup>i</sup>Constitutional, physicochemical, and topological descriptors as implemented within Canvas modeling suite from Schrödinger. <sup>j</sup>Model developed using recursive partitioning (RP) using the Canvas modeling suite from Schrödinger. <sup>k</sup>Random forest (RF) model developed using DEEPCHEM package. <sup>l</sup>Deep neural net (DNN) model developed using DEEPCHEM package. <sup>m</sup>Reverse split (yellow highlight). Training set (1180): experimentally active (521) with  $IC_{50} \leq 100$  nM; experimentally inactive (659). Validation set (295): experimentally active (130) with  $IC_{50} \leq 100$  nM; experimentally inactive (165).

Approximately 2765 pairs were identified using a similarity cutoff threshold of 95% ([Supporting Information S4](#)). Among these, only 51 pairs can be attributed solely to the training set and this provides another inference for the compound diversity within the training data set. Among the 2714 chemical similarity pairs within the validation set, activity cliffs >2 log units are exhibited by at least 102 pairs representing ~4% of the validation set. These experimental observations suggest instances where subtle changes to structure has profound impact on the  $IC_{50}$  being measured in spite of the conserved modifications to the structure. While such biological property space ruggedness cautions on the predictive capabilities of the *in silico* models being developed, it should be emphasized that the overall terrain is pretty smooth with ~2200 pairs exhibiting  $\Delta pIC50 < 1$  log unit. Irrespective of the model performance on the molecules that demonstrate activity cliffs, the experimental observations shed light on the expected microenvironment (steric, electrostatic, H-bond, etc.) hotspots that need to be paid attention to, besides the black box modeling outcome.

[Figure 2b](#) provides a visual depiction of the relationship between the observed experimental affinities with respect to the similarities in the chemical space computed using *SkelSpheres* descriptors implemented in DataWarrior.<sup>62</sup> The key observation is the finding that training set compounds (blue squares) are spread across the validation set compounds, although not being blown out extensively. The near neighbors in the validation set that surrounds the training set based on similarity and the SALI index (square size) shows that the biological environment is also smooth to a major extent albeit the anticipated cliffs observed in some instances within the clustered compounds.

Thus, the analysis of the chemical and biological space reveals that the data set and modeling space [fingerprints/descriptors] should result in productive and predictive models that could guide affinity optimization efforts for *hBACE-1* protein target.

**Classification.** With the confidence provided by the chemistry and biology spaces described above, the modeling domain is expected to perform reasonably well and the same can be gleaned from [Table 1](#) ([Supporting Information S5 and S5a](#)). The data set was split into a binary active ( $IC_{50} \leq 100$  nM) and inactive class, using a self-imposed cutoff for the experimental  $IC_{50}$ . Apart from the radial fingerprint (ECFP6) described above, fingerprint schemes like the linear, dendritic, and MolPrint2D implemented in Canvas were used to develop additional Bayesian classification models. These models achieve between 74 and 93% classification accuracy on the training set, but only around 55–65% prediction accuracy on the validation set.

In order to construct models with increased predictive power, additional classification models were developed via recursive partitioning (RP), random forest (RF), and deep neural net (DNN) statistical techniques using the Canvas constitutional, physicochemical, and topological descriptors. The RF and DNN models were constructed using the DeepChem library.<sup>64</sup> Hyperparameters for RFs and DNNs were selected by model performance on the validation set. The RF models had 10 or 100 trees, and used either square-root, log, and full cutoffs for the number of features used when computing tree split. The number of trees and feature-cutoff were selected with grid hyperparameter search on the validation set. The DNN models had a single hidden layer with 1000 units, and dropout 0.5. Learning rate and decay rate for the

Table 2. Statistical Parameters for the Various Quantitative Models Developed in This Work

Approach <sup>c</sup>	Descriptors	Software	Training Set <sup>a</sup>				Validation Set <sup>b</sup>			
			r <sup>d</sup>	RMSE <sup>e</sup>	MAE <sup>f</sup>	SE <sup>g</sup>	r <sup>2</sup>	RMSE	MAE	SE
MLR	Canvas <sup>h</sup>	Canvas	0.567	1.007	0.835	0.762	0.086	1.223	0.945	0.750
RF	Canvas <sup>h</sup>	DeepChem	0.947	0.351	0.265	0.307	0.261	1.099	0.802	0.671
DNN	Canvas <sup>h</sup>	DeepChem	0.901	0.481	0.371	0.375	0.351	1.030	0.763	0.632
CoMFA	3D-fields <sup>i</sup>	Sybyl	0.919	0.436	0.348	0.420	0.341	1.038	0.813	0.693
COMSIA	3D-fields <sup>i</sup>	Sybyl	0.777	0.723	0.579	0.641	0.143	1.184	0.917	0.838
ABM	3D-fields <sup>i</sup>	MAESTRO	0.917	0.442	0.354	0.425	0.372	1.014	0.785	0.675
FQSAR_gau	3D-fields <sup>i</sup>	MAESTRO	0.884	0.522	0.415	0.493	0.320	1.055	0.801	0.701
FQSAR_ff	3D-fields <sup>i</sup>	MAESTRO	0.585	0.986	0.780	0.784	0.115	1.203	0.935	0.953
MLR <sup>j</sup>	Canvas <sup>h</sup>	Canvas	0.313	1.089	0.890	0.607	0.191	1.186	0.941	0.572
RF <sup>j</sup>	Canvas <sup>h</sup>	DeepChem	0.949	0.296	0.217	0.240	0.721	0.697	0.530	0.520
DNN <sup>j</sup>	Canvas <sup>h</sup>	DeepChem	0.792	0.599	0.462	0.509	0.740	0.672	0.520	0.538
CoMFA <sup>j</sup>	3D-fields <sup>i</sup>	Sybyl	0.776	0.622	0.475	0.548	0.691	0.733	0.566	0.611
ABM <sup>j</sup>	3D-fields <sup>i</sup>	MAESTRO	0.729	0.684	0.527	0.585	0.680	0.747	0.585	0.599

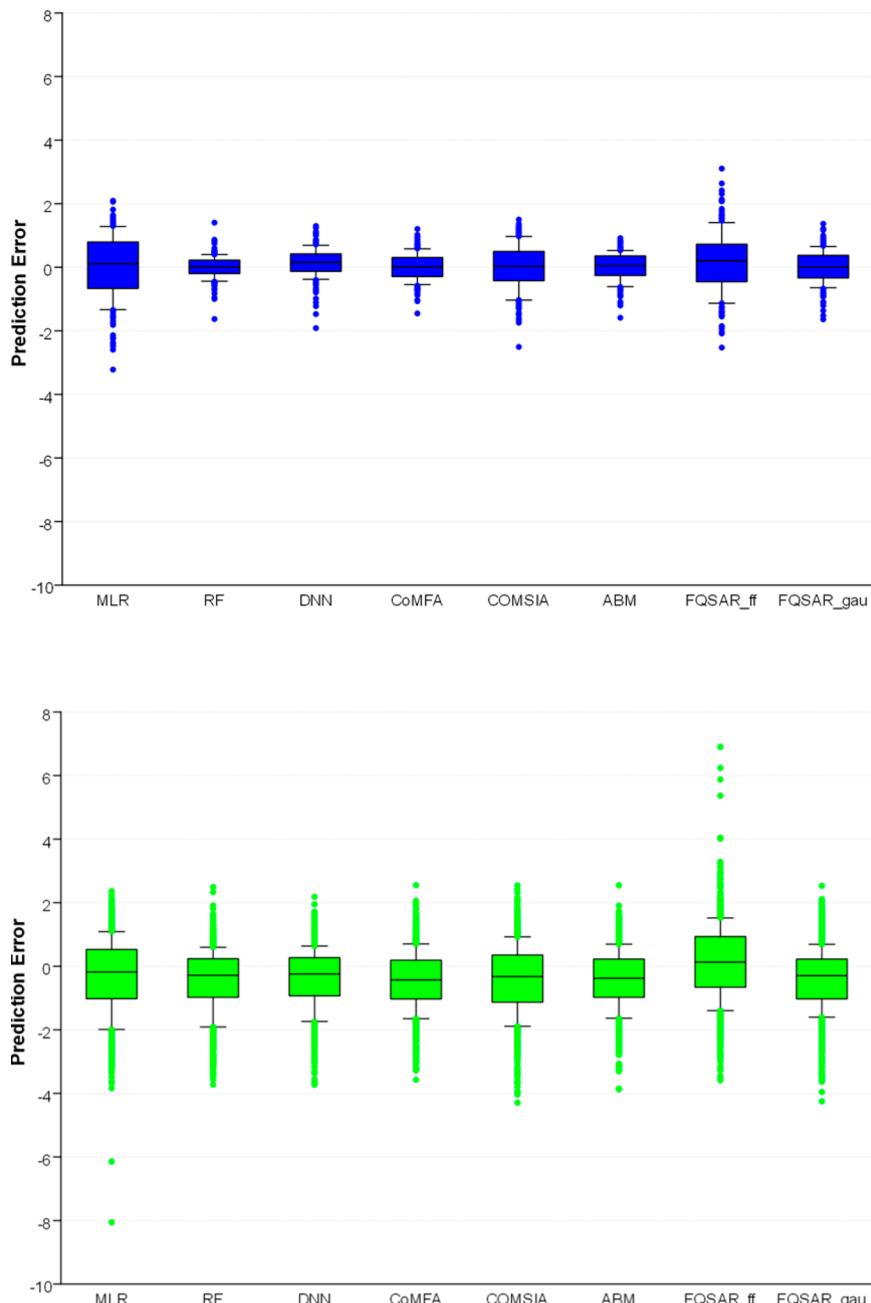
<sup>a</sup>Training set (205): experimentally active (102) with IC<sub>50</sub> ≤ 100 nM; experimentally inactive (103). <sup>b</sup>Validation set (1273): experimentally active (551) with IC<sub>50</sub> ≤ 100 nM; experimentally inactive (722). <sup>c</sup>Statistical technique employed. See the Abbreviations section for definitions. <sup>d</sup>Coefficient of the fit of a linear regression. <sup>e</sup>Root-mean-square error. <sup>f</sup>Mean absolute error. <sup>g</sup>Standard error. <sup>h</sup>1D and 2D constitutional, physiochemical, and topological descriptors as implemented within Canvas modeling suite from Schrödinger. <sup>i</sup>3D-grid based field descriptors utilizing hydrophobic, H-bond donor, and acceptor probes as implemented by the individual approaches implemented within Schrödinger and Sybyl modeling packages. <sup>j</sup>Reverse split (yellow highlighting). Training set (1180): experimentally active (521) with IC<sub>50</sub> ≤ 100 nM; experimentally inactive (659). Validation set (295): experimentally active (130) with IC<sub>50</sub> ≤ 100 nM; experimentally inactive (165).

DNNs were selected with random hyperparameter search on the validation set.<sup>65</sup> Both RFs and DNNs demonstrated greater representational power and learned models with classification accuracy in the 90% range for the training set. On the validation set, both RF and DNN models achieved stronger predictive power with almost 70% accuracy, and offered a maximum of 15% accuracy enrichment compared to the fingerprint based schemes and Bayesian modeling. The performance of the RP model was comparable to the Bayesian model using MolPrint2D fingerprint scheme (Table 1).

**Regression.** Multiple approaches with increasing sophistication were pursued as part of the quantitative modeling exercise and statistical outcomes are provided in Table 2 (Supporting Information S6). At a first glance, the results are underwhelming. The consistently >1 log unit root-mean-square (RMS) error for the validation set seem to caution against the application of such approaches within a lead optimization setting in a discovery program. However, the box plot of the error distribution in Figure 3 reveals that most models performed decently for a majority of the compounds with the exception of some outliers (Supporting Information S7). No single approach outperforms the rest, but CoMFA seem to have an edge over COMSIA within the SYBYL application suite and the atom based QSAR modeling (ABM) appears superior compared to the force field and Gaussian approximations offered by the MAESTRO suite when 3D-field, -grid based approaches alone are compared. As expected, MLR that used descriptor feature selection prior to model development performed the worst, but the RF and DNN models built with DeepChem using the Canvas descriptors exhibited comparable performance to the 3D alignment dependent approaches

mentioned above (Table 2). The results also highlight that there is only minor prediction enrichment offered by the use of 3D based approaches when compared against 2D descriptor based statistical techniques such as DNNs or RFs, suggesting that LBD with sophisticated statistical techniques recovers many of the advantages of 3D-field based techniques. The prediction error of the models is reasonable, considering that the experimental observations are drawn from multiple laboratories in various countries.

**Reverse Split.** To assess the sensitivity of the various methods on the data set size, an alternative 80/20 split of the training and validation set was performed by selecting every fifth compound as a validation set member after sorting the full data set based on the experimental affinity (pIC<sub>50</sub>). Such a split would remove any chemotype bias and represent the biological affinity axis uniformly. This alternate split resulted in 1180 compounds in the training and 295 compounds in the validation set and only methods that performed well for the original data set was selected (Supporting Information S10). As should be expected, the performance accuracy of the selected methods increases, with the machine learning models maintaining a slight edge over their counterparts in classification and regression (Tables 1 and 2). The statistical fit ( $R^2$ ) for the validation set is enhanced dramatically and the RMSE show a 3-fold improvement in error containment compared to the prediction errors for the original 20/80 data set split, albeit the minor loss in the training set performance (Table 2). Notably, the machine learning models now slightly outperform the 3D-field based techniques, matching the expectation that learning algorithms grow increasingly powerful with additional data. The results from the classification models



**Figure 3.** Box plot of the predicted rms error for the training (blue) and validation (green) set compounds from all the approaches considered.

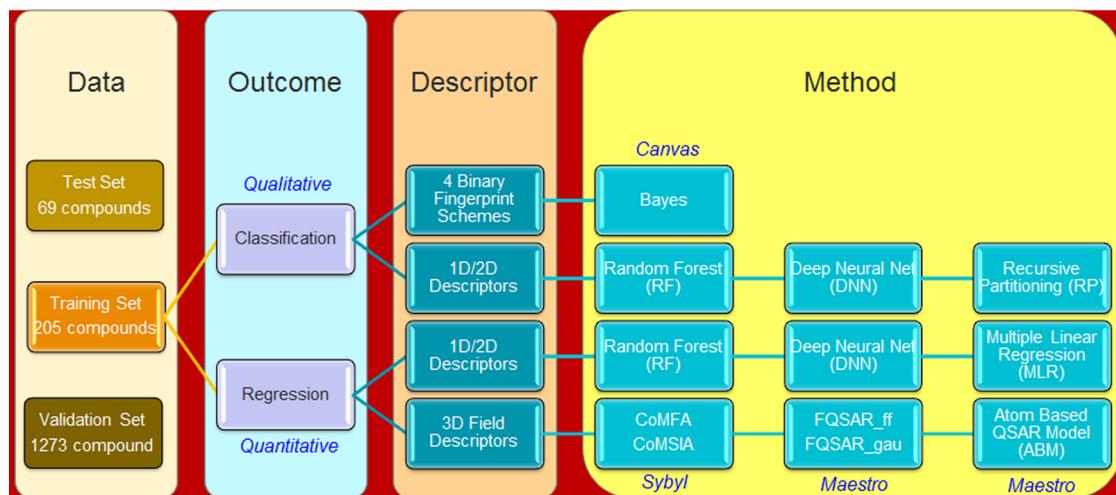
were not dramatic, but did show a tendency of 10–15% accuracy enrichment for the validation set.

## ■ DISCUSSION

In comparing the relationship between the physicochemical attributes and the experimental affinity for the BACE-1 inhibitors, no distinct patterns emerged as beacons for new affinity improvement designs (*Supporting Information S2*). For instance, AlogP did not provide clear probabilities between active and inactive compounds since the distribution of actives at the AlogP value around  $-7$  is similar to the ones observed at AlogP values of  $+7$ . There is a near equal frequency of actives and inactives at AlogP values in  $1\text{--}4$  range as well. Contrary to the lipophilicity trends, the data set suggests that the minimum molecular weight (MW) should be above 300 for compounds to exhibit  $\text{IC}_{50} \leq 100$  nM. However, no discrimination is

observed at the higher MW ranges. The computed ligand BEI<sup>66,67</sup> reflects the expected inverse correlation as the molecular weight increases and the SEI reveals an exponential decline as the ligand PSA increases imitating on the shallow and exposed binding site occupied by BACE-1 inhibitors (*Scheme 1*). The distribution of BEI and SEI histograms [*Supporting Information S2*] also reveal that most of the small molecules are not clustered around their acceptable index values suggesting the suboptimal use of ligand atoms to achieve affinity.<sup>68</sup> Other features like the H-bond donor or acceptor count also does not guide in setting appropriate filters for affinity improvement activities. A correlation plot of AlogP vs PSA advocate a minimal requirement of at least  $60 \text{ \AA}^2$  for PSA and an AlogP value  $>2$ , but is indeed a scatter in distinguishing actives from inactives when viewed from the whole data set. Unfortunately, efforts to draw inferences on affinity based on physicochemical

## Scheme 3. Modeling Scheme and Workflow



parameters (akin to in silico BBB, membrane permeability ADME models) did not yield fruition and this necessitated the development of statistical models reported herein (**Scheme 3**) to capture the experimental trends with decent predictive power becoming an outcome.

Simple Bayesian classification models (**Table 1**) provide a robust framework for filtering ideas, but there is a significant loss in capturing the inactives (specificity) by the use of linear and dendritic fingerprints for the training set compounds. This loss is evident when the validation set results are inspected and points to a preference for ECFP6 or MolPrint2D among the fingerprint schemes evaluated. Fortunately, model accuracies could be improved by the use of RF and DNN techniques with more elaborate descriptors. This enhanced accuracy is likely due to the greater representational capacity of modern machine learning techniques over simpler statistical methods. At a qualitative level, the modeling outcome for this data set revealed the capability of machine-learning technique to construct predictive models, given the relatively small 205 training set compounds. This effectiveness for the DNN models is likely due to the use of statistical regularization techniques such as dropout.<sup>69,70</sup>

The loss in specificity across the different classification models and approaches prompted the consideration of a consensus model to see if aggregating the learning from multiple techniques would improve on the prediction capabilities and minimize the noise. **Table 3** provides the performance enrichment by selecting 5/7 models (results from linear and dendritic fingerprints ignored). In examining the results, all the statistical techniques and descriptor schemes

seem to learn equally well for the actives, since 95% and 75% of the training and validation set is retrieved by at least 4/5 models developed. However, the classification models still do not seem to have captured all the subtle microenvironment effects from the inactives as the consensus prediction from at least 4/5 models retrieves only ~40% of the true negatives and is inferior to the predictions from the machine learning methods (**Table 1**).

Among all the quantitative regression models developed, the outcome from MLR and FQSAR\_ff can be ignored from further consideration (**Table 2**) due to their weakest performance amid the 2D and 3D descriptor based approaches. Very good correlations between the different approaches provided in **Table 4** (**Supporting Information S8**) clearly indicate that the data set is neither sensitive to any specific modeling approach nor is there an advantage for the 3D vs 2D descriptor space used here. The RF and DNN models are highly correlated (**Table 4**), indicating that both techniques learn similar patterns. Similarly, the correlation between CoMFA and FQSAR\_gau suggests the 3D-grid based techniques by both vendors perform comparably, albeit the use of different partial atomic charges schemes. The pIC<sub>50</sub> prediction self-consistency exhibited by the various modeling approaches also encourages re-examination of presumed mispredictions (>2 log unit) when compared against the reported experimental observation. A recent study reported an average 3-fold variability in the observed experimental activity when looking at assay sensitivity over time from a single laboratory.<sup>46</sup> Hence prediction errors close to 1 log unit (10-fold) should still be treated as fairly reasonable, given the speed and throughput of predictions compared to free energy based methods<sup>10</sup> that strive to achieve predictions within experimental error bars. Further analysis of the error distribution from RF, CoMFA, and ABM (**Supporting Information S7**) suggest that ~40% of the validation set compounds are predicted with a rms error <0.5 log units compared to experimentally determined IC<sub>50</sub> values, and another ~30% when the threshold is moved to 1 log unit.

Closer inspection of the regression plots from ABM, CoMFA, and DNN shows that the prediction error is relatively high for inactive compounds (**Figure 4**). A reassessment of the errors showed an improvement of 3- to 5-fold for the validation set compounds when only molecules with experimental IC<sub>50</sub> ≤ 1 μM are considered (**Figure 5**; **Supporting Information S9**).

**Table 3. Performance of the Consensus Models Showcasing Classification Accuracy Enrichment Using Bayesian (ECFP6, MolPrint2D), RP, RF, and DNN Methods**

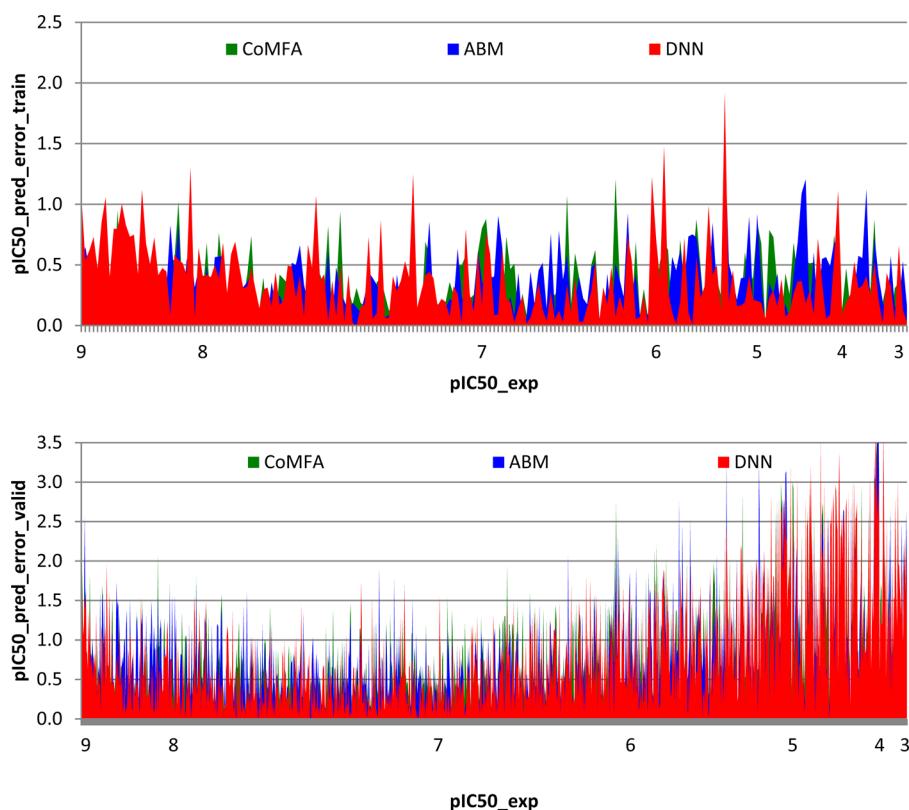
data set	no. molecules	5/5 <sup>a</sup>	4/5	3/5	2/5	1/5	0/5 <sup>b</sup>
Train_active	102	88	12	1	1		
Valid_active	551	247	188	78	17	16	5
Train_inactive	103	66	19	10	4	2	2
Valid_inactive	722	177	107	67	98	126	147

<sup>a</sup>Number of molecules classified correctly by all five of the models.

<sup>b</sup>Number of molecules classified incorrectly by all five of the models.

**Table 4.** Correlation Matrix for the pIC50 Predicted Using Various Approaches for Training (top off-diagonal; *italics*) and Validation (bottom off-diagonal; yellow highlight) Sets

<i>r_Training</i>	MLR	RF	DNN	CoMFA	COMSIA	ABM	FQSAR_gau	FQSAR_ff
<b>MLR</b>		0.78	0.736	0.745	0.814	0.748	0.775	0.736
<b>RF</b>	<b>0.751</b>		0.955	0.946	0.857	0.956	0.94	0.809
<b>DNN</b>	<b>0.684</b>	<b>0.869</b>		0.927	0.831	0.943	0.92	0.796
<b>CoMFA</b>	0.553	0.661	0.707		0.877	0.962	0.955	0.775
<b>COMSIA</b>	0.713	0.728	0.695	<b>0.718</b>		0.895	0.914	0.764
<b>ABM</b>	0.578	0.695	0.727	<b>0.885</b>	<b>0.781</b>		0.967	0.792
<b>FQSAR_gau</b>	0.382	0.558	0.53	0.516	0.501	<b>0.616</b>		0.808
<b>FQSAR_ff</b>	0.617	0.697	0.72	0.859	0.794	0.913	<b>0.545</b>	



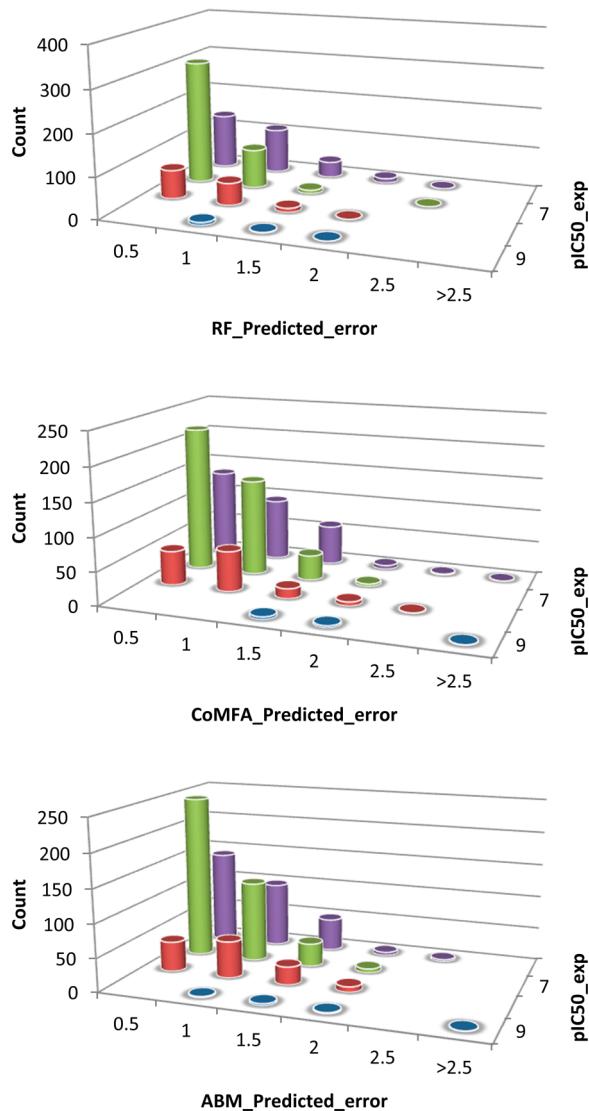
**Figure 4.** RMSE plot for each of the training (top) and validation set (bottom) compounds based on DNN (red), CoMFA (green), and ABM (blue). Compounds sorted based on experimental pIC50 values with each vertical line representing the individual molecule's RMSE with respect to the experimentally observed values.

Incidentally for this subset, the 2D based RF method outperforms the 3D based techniques and reiterates the fact that reliable models can still be developed using traditional descriptors, but with the use of machine learning methods. Overall, quantitative modeling of the data set still offers strong prediction accuracy and can successfully be used in a lead optimization setting.

Redeveloping some of the models using a reverse split of 80/20 training to validation set ratio did result in enhanced statistical fit for the quantitative models and demonstrated a 3-fold improvement in minimizing the prediction errors for the validation set. Similarly, the classification models were able to achieve accuracy and specificity improvements. However, the conclusions from the qualitative (**Table 1**) and quantitative

(**Table 2**) approaches were not game changers to suggest that the data set size considered here could influence the modeling outcome. The results support the view that reasonable models with good prediction capabilities can still be developed with a relatively diverse and modest data set size. This would become more apparent, if parallels can be drawn from the relatively good success of the implemented global *in silico* ADME models using relatively small data sets<sup>71,72</sup> as well as the secondary structure prediction based on just the sequence information<sup>73,74</sup> from a fairly small sample size that stand as a testament of time.<sup>75</sup>

The test set data reported in this study paves the way to verify the predictive capabilities of the models developed as they represent additional chemical space (**Figure 1a** and **d**).



**Figure 5.** Predicted RMSE distribution spread based on RF, CoMFA, and ABM for the validation compound subset with experimental  $IC_{50} \leq 1 \mu M$ .

Although the kind of correlation expected between experimental  $IC_{50}$  and observed  $K_i$  is not established for BACE-1 inhibitors considered here, it could still be assumed to be close enough for comparing the predictions. Analysis of the 35 compounds with reported  $K_i$  in the test set shows that the DNN method retrieves ~80% of the actives and ~77% of the inactives. These results are comparable to that achieved for the validation set and provide confidence to the extension of the model for an external data set.

The strong model accuracy on the test set suggests a simple strategy for using predictive machine learning models during prospective design: evaluate proposed compounds with learned models to ensure that synthesized compounds are predicted to be active. Given the low hit rate of early stage screens, this strategy could help boost the number of actives discovered in a lead identification scenario. This motivated the inclusion of the prediction results for the remaining 25 compounds in the test set for prospective design activities. Prediction results from quantitative modeling for part of the test set are not compared (Supporting Information S11) since the experimental measures

are different, but are provided as guidance to prioritize future design strategies expanding on these molecules and scaffold.

Finally, it is rather difficult to compare the performance of the currently developed models for *h*BACE-1 with reported LBD and SBD literature due to the different choice of the data set size, chemical scaffolds considered, and the approaches pursued.<sup>76,77</sup> The modeling outcome published recently by Cramer<sup>78</sup> using template CoMFA and neural net ( $SD \sim 0.88$ ) suggests the present CoMFA and RF/DNN outcome (Table 2) to be better and can serve as future reference comparator for the 2D and 3D based LBD approaches. Similarly, the prediction errors achieved here are also comparable to that reported recently for BACE-1 using modern free energy calculations.<sup>10,79</sup>

## CONCLUSION

This study analyzed ligands known to bind human BACE-1 through various LBD approaches starting from modest fingerprints to sophisticated 3D fields. The constitutional, topological, and physicochemical descriptors projected in a low dimensional chemical space for visual analysis show a continuous spread of the data set and assured reliability in the data set being modeled. Correlation of the observed experimental binding affinity with appropriate compound feature scheme reveals the biological landscape to exhibit some activity cliffs (~4%) that can potentially be used to understand affinity directions. The mapping of the chemical and biological spaces offered confidence that descriptor based techniques can yield predictive models.

Preliminary attempts to correlate physicochemical properties to affinity were unsuccessful. Qualitative classification models developed subsequently identified ECFP6 and MolPrint2D to be better fingerprint schemes among the ones considered. However, DNN and RF machine learning methods with Canvas descriptors resulted in robust classification accuracy and are shown to exhibit superior performance compared to traditional Bayesian techniques. Quantitative regression models developed also suggest that DNN and RF using Canvas descriptors can achieve statistical accuracy similar to 3D field based techniques that often require molecular alignment of diverse chemical scaffolds in one universal chemical space. Preliminary confidence on the application of such approaches is supported through the prediction outcome for the external test set compounds used in this study.

The current results provide a strong pretext for the systematic use of advanced machine-learning approaches in computational drug discovery,<sup>80–82</sup> when high end molecular dynamics based free energy approaches serve as a limiting factor. The brevity in choosing an unconventional 20/80 training to validation set split reflecting real life project scenario demonstrates that productive and predictive models can still be achieved, if the modeling problem at hand is thought through properly and executed appropriately.

Tests on the reversed and traditional 80/20 split reveal that machine-learning methods become even more effective given additional data. Despite the fact that the original 20/80 split is an *extreme* scenario for machine-learning, DNNs and RFs still achieve strong performance. To encourage the use of machine-learning methods in practice, we have open-sourced all code used to construct our machine-learning models as part of the DeepChem library (see associated content). Full details about the machine learning models are available in the open-sourced code.

The good correlation of the predictions between different methods and descriptor schemes removes the chance correlation liability and emphasizes the predictive power of the models developed. As with any QSAR modeling, the outcome is always measured in terms of the model guidance achieved for prospective designs. The current endeavor offers the best of both practitioners world<sup>83</sup> since the 1D/2D descriptors and machine learning methods provide the prediction strength and 3D field based results convey an understanding about the kind of changes needed to modulate affinity. The work's intention is not to exhaustively compare the strengths and limitations of the fingerprint scheme, descriptor set used or approaches employed, but to provide a systematic and rational QSAR workflow that could be pursued for enabling affinity improvement activities in lead optimization efforts as part of the iterative everyday medicinal chemistry activities within a preclinical drug discovery project. The current work also offers additional confidence to the preliminary study reported for another aspartyl protease, renin,<sup>29</sup> in arriving at a generic generalization on the use of LBD approaches for quantitative affinity predictions with a high throughput compared to the expensive and low throughput free energy simulation based techniques reported recently,<sup>10</sup> while offering comparable performance attributes.

Finally, as mentioned above, it is increasingly becoming evident that deep learning techniques can play a key role in predictive modeling<sup>84</sup> especially during the lead optimization step when computational models need to be generated or implemented for multiple experimental end points that include physicochemical attributes such as solubility and lipophilicity; early ADMET properties like microsomal stability, hERG liability, and cytochrome P-450 inhibition/induction, etc. All these experimental or predicted parameters are typically used as part of the multiparameter optimization (MPO) score.<sup>85</sup> Within this context, if the affinity prediction models such as the ones developed here are embedded as part of the MPO scoring system, much higher efficiency could be achieved in design ideas that can go forward toward synthesis prioritization. The strong performance of the machine learning methods on the test set provides preliminary evidence that such approaches can be fully integrated within the MPO activities in the lead optimization endeavor.

## ■ ASSOCIATED CONTENT

### S Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: [10.1021/acs.jcim.6b00290](https://doi.org/10.1021/acs.jcim.6b00290).

SD file with all the 3D chemical structures, experimental and predicted pIC50s, source literature information, and PDB codes, along with PUBMED and ChEMBL identification numbers, where annotated. Microsoft excel file with results and analysis tabs (S1–11) for the outcomes presented in the manuscript ([ZIP](#))

## ■ AUTHOR INFORMATION

### Corresponding Author

\*Phone: +1 269-359-9528. E-mail: [govindan.subramanian@zoetis.com](mailto:govindan.subramanian@zoetis.com).

### Funding

B.R. was supported by the Fannie and John Hertz Foundation.

### Notes

The authors declare no competing financial interest.

DeepChem BACE-1 inhibitor examples are available at <https://github.com/deepchem/deepchem/tree/master/examples/bace>

## ■ ACKNOWLEDGMENTS

G.S. dedicates this work to Professor Eluvathingal D. Jemmis on the occasion of his 65th birthday. G.S. also thanks Shashidhar N. Rao for help with CoMFA and COMSIA calculations.

## ■ ABBREVIATIONS

BACE-1,  $\beta$ -secretase 1; 3D, 3-dimensional; 2D, 2-dimensional; RF, random forest; DNN, deep neural nets; RP, recursive partitioning; CoMFA, comparative molecular field analysis; COMSIA, comparative molecular similarity index analysis; MLR, multiple linear regression; ABM, atom-based QSAR model; FQSAR-ff, field QSAR using force fields; FQSAR\_gau, field QSAR using Gaussian approximation

## ■ REFERENCES

- (1) Sliwoski, G.; Kothiwale, S.; Meiler, J.; Lowe, E. W., Jr. Computational Methods in Drug Discovery. *Pharmacol. Rev.* **2014**, *66*, 334–395.
- (2) Ekins, S.; Mestres, J.; Testa, B. *In silico* Pharmacology for Drug Discovery: Methods for Virtual Ligand Screening and Profiling. *Br. J. Pharmacol.* **2007**, *152*, 9–20.
- (3) Ekins, S.; Mestres, J.; Testa, B. *In silico* Pharmacology for Drug Discovery: Applications to Targets and Beyond. *Br. J. Pharmacol.* **2007**, *152*, 21–37.
- (4) Macalino, S. J. Y.; Gosu, V.; Hong, S.; Choi, S. Role of Computer-aided Drug Design in Modern Drug Discovery. *Arch. Pharmacal Res.* **2015**, *38*, 1686–1702.
- (5) Anderson, A. C. The Process of Structure-Based Drug Design. *Chem. Biol.* **2003**, *10*, 787–797.
- (6) Kuntz, I. D.; Blaney, J. M.; Oatley, S. J.; Langridge, R.; Ferrin, T. E. A Geometric Approach to Macromolecule-ligand Interactions. *J. Mol. Biol.* **1982**, *161*, 269–288.
- (7) Kitchen, D. B.; Decornez, H.; Furr, J. R.; Bajorath, J. Docking and Scoring in Virtual Screening for Drug Discovery: Methods and Applications. *Nat. Rev. Drug Discovery* **2004**, *3*, 935–949.
- (8) 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.; et al. A Critical Assessment of Docking Programs and Scoring Functions. *J. Med. Chem.* **2006**, *49*, 5912–5931.
- (9) Homeyer, N.; Stoll, F.; Hillisch, A.; Gohlke, H. Binding Free Energy Calculations for Lead Optimization: Assessment of Their Accuracy in an Industrial Drug Design Context. *J. Chem. Theory Comput.* **2014**, *10*, 3331–3344.
- (10) Wang, L.; Wu, Y.; Deng, Y.; Kim, B.; Pierce, L.; Krilov, G.; Lupyan, D.; Robinson, S.; Dahlgren, M. K.; Greenwood, J.; et al. Accurate and Reliable Prediction of Relative Ligand Binding Potency in Prospective Drug Discovery by Way of a Modern Free-Energy Calculation Protocol and Force Field. *J. Am. Chem. Soc.* **2015**, *137*, 2695–2703.
- (11) Aldeghi, M.; Heifetz, A.; Bodkin, M. J.; Knapp, S.; Biggin, P. Accurate Calculation of the Absolute Free Energy of Binding for Drug Molecules. *Chem. Sci.* **2016**, *7*, 207–218.
- (12) De Vivo, M.; Masetti, M.; Bottegoni, G.; Cavalli, A. Role of Molecular Dynamics and Related Methods in Drug Discovery. *J. Med. Chem.* **2016**, *59*, 4035–4061.
- (13) Craig, I. R.; Essex, J. W.; Spiegel, K. Ensemble Docking into Multiple Crystallographically Derived Protein Structures: An Evaluation Based on the Statistical Analysis of Enrichments. *J. Chem. Inf. Model.* **2010**, *50*, 511–524.
- (14) Korb, O.; Olsson, T. S. G.; Bowden, S. J.; Hall, R. J.; Verdonk, M. L.; Liebeschuetz, J. W.; Cole, J. C. Potential and Limitations of Ensemble Docking. *J. Chem. Inf. Model.* **2012**, *52*, 1262–1274.

- (15) Grebner, C.; Iegre, J.; Ulander, J.; Edman, K.; Hogner, A.; Tyrchan, C. Binding Mode and Induced Fit Predictions for Prospective Computational Drug Design. *J. Chem. Inf. Model.* **2016**, *S6*, 774–787.
- (16) Hansen, N.; van Gunsteren, W. F. Practical Aspects of Free-energy Calculations: A Review. *J. Chem. Theory Comput.* **2014**, *10*, 2632–47.
- (17) Marillet, S.; Boudinot, P.; Cazals, F. High-resolution Crystal Structures Leverage Protein Binding Affinity Predictions. *Proteins: Struct., Funct., Genet.* **2016**, *84*, 9–20.
- (18) Hansch, C.; Maloney, P.; Fujita, T.; Muir, R. Correlation of Biological Activity of Phenoxyacetic Acids with Hammett Substituent Constants and Partition Coefficients. *Nature* **1962**, *194*, 178–180.
- (19) Cherkasov, A.; Muratov, E. N.; Fourches, D.; Varnek, A.; Baskin, I. I.; Cronin, M.; Dearden, J.; Gramatica, P.; Martin, Y. C.; Todeschini, R.; et al. QSAR Modeling: Where Have you Been? Where are you Going to. *J. Med. Chem.* **2014**, *57*, 4977–5010.
- (20) Sanders, M. P. A.; Barbosa, A. J. M.; Zarzycka, B.; Nicolaes, G. A. F.; Klomp, J. P. G.; Vlieg, J.; Rio, A. D. Comparative Analysis of Pharmacophore Screening Tools. *J. Chem. Inf. Model.* **2012**, *52*, 1607–1620.
- (21) Wolber, G.; Sippl, W. Pharmacophore Identification and Pseudo-receptor Modeling. In *The Practice of Medicinal Chemistry*, Fourth ed.; Wermuth, C. G., Aldous, D., Raboissone, P., Rognan, D., Eds.; Academic Press: London, United Kingdom, 2015; pp 489–510.
- (22) Cramer, R. D., III; 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*, 5959–5967.
- (23) Drwal, M. N.; Griffith, R. Combination of Ligand- and Structure-based Methods in Virtual Screening. *Drug Discovery Today: Technol.* **2013**, *10*, e395–e401.
- (24) Broccatelli, F.; Brown, N. Best of Both Worlds: On the Complementarity of Ligand-based and Structure-based Virtual Screening. *J. Chem. Inf. Model.* **2014**, *S4*, 1634–1641.
- (25) Huang, S.-Y.; Li, M.; Wang, J.; Pan, Y. HybridDock: A Hybrid Protein-Ligand Docking Protocol Integrating Protein- and Ligand-based Approaches. *J. Chem. Inf. Model.* **2016**, *S6*, 1078–1087.
- (26) Prathipati, P.; Mizuguchi, K. Integration of Ligand and Structure Based Approaches for CSAR-2014. *J. Chem. Inf. Model.* **2016**, *S6*, 974–987.
- (27) Klebe, G.; Abraham, U. Comparative Molecular Similarity Index Analysis (CoMSIA) to Study Hydrogen-bonding Properties and to Score Combinatorial Libraries. *J. Comput.-Aided Mol. Des.* **1999**, *13*, 1–10.
- (28) Cappel, D.; Dixon, S. L.; Sherman, W.; Duan, J. Exploring Conformational Search Protocols for Ligand-based Virtual Screening and 3-D QSAR Modeling. *J. Comput.-Aided Mol. Des.* **2015**, *29*, 165–182.
- (29) Subramanian, G.; Rao, S. N. An Integrated Computational Workflow for Efficient and Quantitative Modeling of Renin Inhibitors. *Bioorg. Med. Chem.* **2012**, *20*, 851–858.
- (30) Urniaż, R. D.; Jóźwiak, K. X-ray Crystallographic Structures as a Source of Ligand Alignment in 3D-QSAR. *J. Chem. Inf. Model.* **2013**, *S3*, 1406–1414.
- (31) Therese, P. J.; Manvar, D.; Kondepudi, S.; Battu, M. B.; Sriram, D.; Basu, A.; Yogeeshwari, P.; Kaushik-Basu, N. Multiple e-Pharmacophore Modeling, 3D-QSAR, and High-throughput Virtual Screening of Hepatitis C Virus NS5B Polymerase Inhibitors. *J. Chem. Inf. Model.* **2014**, *S4*, 539–552.
- (32) Cramer, R. D.; Wendt, B. Template CoMFA: The 3D-QSAR Grail? *J. Chem. Inf. Model.* **2014**, *S4*, 660–671.
- (33) Cramer, R. D. Template CoMFA Applied to 116 Biological Targets. *J. Chem. Inf. Model.* **2014**, *S4*, 2147–2156.
- (34) Haass, C.; Selkoe, D. J. Soluble Protein Oligomers in Neurodegeneration: Lessons from the Alzheimer's Amyloid beta-peptide. *Nat. Rev. Mol. Cell Biol.* **2007**, *8*, 101–112.
- (35) Walsh, D. M.; Selkoe, D. J. A beta Oligomers – a Decade of Discovery. *J. Neurochem.* **2007**, *101*, 1172–1184.
- (36) Cummings, J. L.; Morstorf, T.; Zhong, K. Alzheimer's Disease Drug-development Pipeline: Few Candidates, Frequent Failures. *Alzheimer's Res. Ther.* **2014**, *6*, 37.
- (37) Jarvis, L. M. Alzheimer's Next Chapter. *Chem. Eng. News* **2015**, *93* (22), 11–15.
- (38) Hong, L.; Koelsch, G.; Lin, X.; Wu, S.; Terzyan, S.; Ghosh, A. K.; Zhang, X. C.; Tang, J. Structure of the Protease Domain of Memapsin 2 ( $\beta$ -secretase) Complexed with Inhibitor. *Science* **2000**, *290*, 150–153.
- (39) Xu, Y.; Li, M. J.; Greenblatt, H.; Chen, W.; Paz, A.; Dym, O.; Peleg, Y.; Chen, T.; Shen, X.; He, J.; et al. Flexibility of the Flap in the Active Site of BACE1 as Revealed by Crystal Structures and Molecular Dynamics Simulations. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2012**, *D68*, 13–25.
- (40) Shimizu, H.; Tosaki, A.; Kaneko, K.; Hisano, T.; Sakurai, T.; Nukina, N. Crystal Structure of an Active Form of BACE1, an Enzyme Responsible for Amyloid  $\beta$  Protein Production. *Mol. Cell. Biol.* **2008**, *28*, 3663–3671.
- (41) Manoharan, P.; Chennoju, K.; Ghoshal, N. Target Specific Proteochemometric Model Development for BACE1 – Protein Flexibility and Structural Water are Critical in Virtual Screening. *Mol. BioSyst.* **2015**, *11*, 1955–1972.
- (42) Dominguez, J. L.; Christopeit, T.; Villaverde, M. C.; Gossas, T.; Otero, J. M.; Nyström, S.; Baraznenok, V.; Lindström, E.; Danielson, U. H.; Sussman, F. Effect of the Protonation State of the Titratable Residues on the Inhibitor Affinity of BACE-1. *Biochemistry* **2010**, *49*, 7255–7263.
- (43) Ellis, C. R.; Shen, J. pH-Dependent Population Shift Regulates BACE1 Activity and Inhibition. *J. Am. Chem. Soc.* **2015**, *137*, 9543–9546.
- (44) Geschwindner, S.; Ulander, J.; Johansson, P. Ligand Binding Thermodynamics in Drug Discovery: Still a Hot Tip? *J. Med. Chem.* **2015**, *58*, 6321–6335.
- (45) RCSB Protein Data Bank. <http://www.rcsb.org/> (accessed Jan 30, 2014).
- (46) Kramer, C.; Dahl, G.; Tyrchan, C.; Ulander, J. A Comprehensive Company Database Analysis of Biological Assay Variability. *Drug Discovery Today* **2016**, *21*, 1213–1221.
- (47) ChEMBL. <https://www.ebi.ac.uk/chembl> (accessed Mar 30, 2014).
- (48) DeepChem datasets. <https://github.com/deepchem/deepchem/tree/master/datasets> (accessed May 3, 2016).
- (49) BIOVIA Pipeline Pilot. <http://accelrys.com/products/collaborative-science/biovia-pipeline-pilot/> (accessed May 3, 2016).
- (50) Abad-Zapatero, C.; et al. Ligand Efficiency Indices for Effective mapping of Chemico-biological Space: The concept of an Atlas-like Representation. *Drug Discovery Today* **2010**, *15*, 804–811.
- (51) MAESTRO; Schrödinger, LLC: New York, NY, USA, 2014.
- (52) MOE; Chemical Computing Group: Montreal, Canada, 2014.
- (53) Canvas; Schrödinger, LLC: New York, NY, USA, 2014.
- (54) Sastry, M.; Lowrie, J. F.; Dixon, S. L.; Sherman, W. Large-Scale Systematic Analysis of 2D Fingerprint Methods and Parameters to Improve Virtual Screening Enrichments. *J. Chem. Inf. Model.* **2010**, *50*, 771–784.
- (55) An, Y.; Sherman, W.; Dixon, S. L. Kernel-based Partial Least Squares: Application to Fingerprint-based QSAR with Model Visualization. *J. Chem. Inf. Model.* **2013**, *S3*, 2312–2321.
- (56) Ramsundar, B.; Kearnes, S.; Riley, P.; Webster, D.; Konerding, D.; Pande, V. Massively Multitask Networks for Drug Discovery. Cornell University Library. <http://arxiv.org/abs/1502.02072> (accessed May 3, 2016).
- (57) Sheridan, R. P. The Relative Importance of Domain Applicability Metrics for Estimating Prediction Errors in QSAR Varies with Training Set Diversity. *J. Chem. Inf. Model.* **2015**, *55*, 1098–1107.
- (58) Rogers, D.; Hahn, M. Extended-connectivity Fingerprints. *J. Chem. Inf. Model.* **2010**, *50*, 742–754.
- (59) Bemis, B. W.; Murcko, M. A. The Properties of Known Drugs. 1. Molecular Frameworks. *J. Med. Chem.* **1996**, *39*, 2887–2893.

- (60) Kohonen, T. *Self-Organizing Maps*, 3rd ed.; Springer-Verlag: Berlin, 2001.
- (61) Osolodkin, D. I.; Radchenko, E. V.; Orlov, A. A.; Voronkov, A. E.; Palyulin, V. A.; Zefirov, N. S. Progress in Visual Representations of Chemical Space. *Expert Opin. Drug Discovery* **2015**, *10*, 959–973.
- (62) 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*, 460–473.
- (63) Guha, R.; van Drie, J. H. Structure-activity Landscape Index: Identifying and Quantifying Activity Cliffs. *J. Chem. Inf. Model.* **2008**, *48*, 646–658.
- (64) Deepchem Home Page. <https://github.com/deepchem/deepchem> (accessed May 3, 2016).
- (65) Bergstra, J.; Bengio, Y. Random Search for Hyper-Parameter Optimization. *J. Machine Learn. Res.* **2012**, *12*, 281–305.
- (66) Abad-Zapatero, C.; Metz, J. T. Ligand Efficiency Indices as Guideposts for Drug Discovery. *Drug Discovery Today* **2005**, *10*, 464–469.
- (67) Abad-Zapatero, C. Ligand Efficiency Indices for Effective Drug Discovery. *Expert Opin. Drug Discovery* **2007**, *2*, 469–488.
- (68) Mignani, S.; Huber, S.; Tomás, H.; Rodrigues, J.; Majoral, J.-P. Compound High-quality Criteria: A New Vision to Guide the Development of Drugs, Current Situation. *Drug Discovery Today* **2016**, *21*, 573–584.
- (69) Srivastava, N.; Hinton, G.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *J. Mach. Learn. Res.* **2014**, *15*, 1929–1958.
- (70) Mendenhall, J.; Meiler, J. Improving Quantitative Structure-activity Relationship Models using Artificial Neural Networks Trained with Dropout. *J. Comput.-Aided Mol. Des.* **2016**, *30*, 177–189.
- (71) Gleeson, M. P.; Hersey, A.; Hannongbua, S. In-silico ADME Models: A General Assessment of their Utility in Drug Discovery Applications. *Curr. Top. Med. Chem.* **2011**, *11*, 358–381.
- (72) Peach, M. L.; Zakharov, A. V.; Liu, R.; Pugliese, A.; Tawa, G.; Wallqvist, A.; Nicklaus, M. C. Computational Tools and Resources for Metabolism-related Property Predictions. 1. Overview of Publicly Available (free and commercial) Databases and Software. *Future Med. Chem.* **2012**, *4*, 1907–1932.
- (73) Kabsch, W.; Sander, C. How Good are Predictions of Protein Secondary Structure? *FEBS Lett.* **1983**, *155*, 179–182.
- (74) Drozdetskiy, A.; Cole, C.; Procter, J.; Barton, G. J. JPred4: a protein Secondary Structure Prediction Server. *Nucleic Acids Res.* **2015**, *43*, W389–W394.
- (75) Chou, P. Y.; Fasman, G. D. Conformational Parameters for Amino Acids in Helical,  $\beta$ -sheet, and Random Coil Regions Calculated from Proteins. *Biochemistry* **1974**, *13*, 211–222.
- (76) Ambre, P.; Roy, K. Advances in Quantitative Structure-activity relationship models of anti-Alzheimer's Agents. *Expert Opin. Drug Discovery* **2014**, *9*, 697–723 and references cited therein.
- (77) Wu, Q.; Li, X.; Gao, Q.; Wang, J.; Li, Y.; Yang, L. Interaction mechanism Exploration of HEA Derivatives as BACE1 Inhibitors by *in silico* Analysis. *Mol. BioSyst.* **2016**, *12*, 1151–1165 and references cited therein.
- (78) 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*, e0129307.
- (79) Ciordia, M.; Perez-Benito, L.; Delgado, F.; Trabanco, A. A.; Tresadern, G. Application of Free Energy Perturbation for the Design of BACE1 Inhibitors. *J. Chem. Inf. Model.* **2016**, *56*, 1856.
- (80) Xu, Y.; Dai, Z.; Chen, F.; Gao, S.; Pei, J.; Lai, L. Deep Learning for Drug-induced Liver Injury. *J. Chem. Inf. Model.* **2015**, *55*, 2085–2093.
- (81) Gawehn, E.; Hiss, J. A.; Schneider, G. Deep Learning in Drug Discovery. *Mol. Inf.* **2016**, *35*, 3–14.
- (82) Mamoshina, P.; Vieira, A.; Putin, E.; Zhavoronkov, A. Applications of Deep Learning in Biomedicine. *Mol. Pharmaceutics* **2016**, *13*, 1445–1454.
- (83) Fujita, T.; Winkler, D. A. Understanding the Roles of the “Two QSARs. *J. Chem. Inf. Model.* **2016**, *56*, 269–274.
- (84) Xu, Y.; Dai, Z.; Chen, F.; Gao, S.; Pei, J.; Lai, L. Deep Learning for Drug-Induced Liver Injury. *J. Chem. Inf. Model.* **2015**, *55*, 2085–2093.
- (85) Segall, M. D.; Yusof, I.; Champness, E. J. Avoiding Missed Opportunities by Analyzing the Sensitivity of Our Decisions. *J. Med. Chem.* **2016**, *59*, 4267–4277.