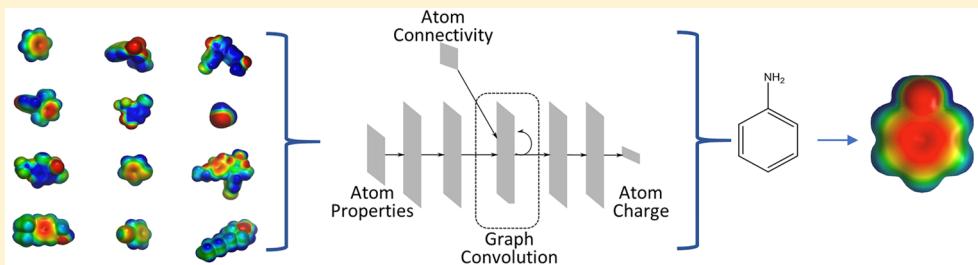


Practical High-Quality Electrostatic Potential Surfaces for Drug Discovery Using a Graph-Convolutional Deep Neural Network

Prakash Chandra Rathi, R. Frederick Ludlow, and Marcel L. Verdonk*[†]

Astex Pharmaceuticals, 436 Cambridge Science Park, Milton Road, Cambridge CB4 0QA, United Kingdom

Supporting Information



ABSTRACT: Inspecting protein and ligand electrostatic potential (ESP) surfaces in order to optimize electrostatic complementarity is a key activity in drug design. These ESP surfaces need to reflect the true electrostatic nature of the molecules, which typically means time-consuming high-level quantum mechanics (QM) calculations are required. For interactive design much faster alternative methods are required. Here, we present a graph convolutional deep neural network (DNN) model, trained on ESP surfaces derived from high quality QM calculations, that generates ESP surfaces for ligands in a fraction of a second. Additionally, we describe a method for constructing fast QM-trained ESP surfaces for proteins. We show that the DNN model generates ESP surfaces that are in good agreement with QM and that the ESP values correlate well with experimental properties relevant to medicinal chemistry. We believe that these high-quality, interactive ESP surfaces form a powerful tool for driving drug discovery programs forward. The trained model and associated code are available from https://github.com/AstexUK/ESP_DNN.

INTRODUCTION

The value of electrostatic calculations for understanding and predicting molecular properties has been recognized for decades.^{1–3} In particular, it is well-known that molecular electrostatics can be predictive of a molecule's chemical reactivity and its ability to form certain types of interactions. A popular way of visualizing the electrostatic nature of molecules is through electrostatic potential (ESP) surfaces.⁴ Molecular ESP surfaces are generated by systematically moving a unit charge probe around the molecule's surface. At each surface point, the ESP energy of the probe is calculated and the molecular surface is then colored accordingly.

In drug discovery, it is widely accepted that electrostatic complementarity between protein and ligand is critically important in order to obtain optimal affinity and selectivity. Typically, the strategy is to generate ESP surfaces of different ligands and visually inspect and compare them, where possible in the context of the ESP surface of the target protein.⁴ Additionally, several groups have proposed metrics that assess protein–ligand electrostatic complementarity by computationally comparing the ESP surfaces of protein and ligand in the bound state. A good example of such an electrostatic complementarity measure, as well as a review of this field, was reported by Bauer and Mackey recently.⁵

The utility to drug discovery of ESP surfaces and any derived metrics strongly relies on the quality of the underlying

electrostatic model. However, the electrostatics of protein–ligand interactions is complex for a variety of reasons. First, the types of interactions for which electrostatics play a role are not limited to well-known ones like hydrogen bonds but also include more subtle interactions involving aromatic π clouds or σ holes in halogen and sulfur atoms.^{6–8} Also, the local environment within a molecule (e.g., electron withdrawing substituents on an aromatic ring) can have a significant effect on an atom's electrostatic properties. These effects can only be captured accurately with quantum mechanical (QM) calculations, but these methods are computationally expensive, preventing their use as a tool for interactive drug design. This issue can be partially resolved by using precalculated point charges that have been derived from QM calculations. However, this only addresses the calculation of protein ESP surfaces, as a higher-level (slow) calculation is still required for each new ligand.

Recently, Bleiziffer and colleagues used random forest regression models, trained on QM calculations for a large set of compounds, to predict atomic point charges for small molecules.⁹ The aim of this study was to generate partial

Special Issue: Artificial Intelligence in Drug Discovery

Received: July 22, 2019

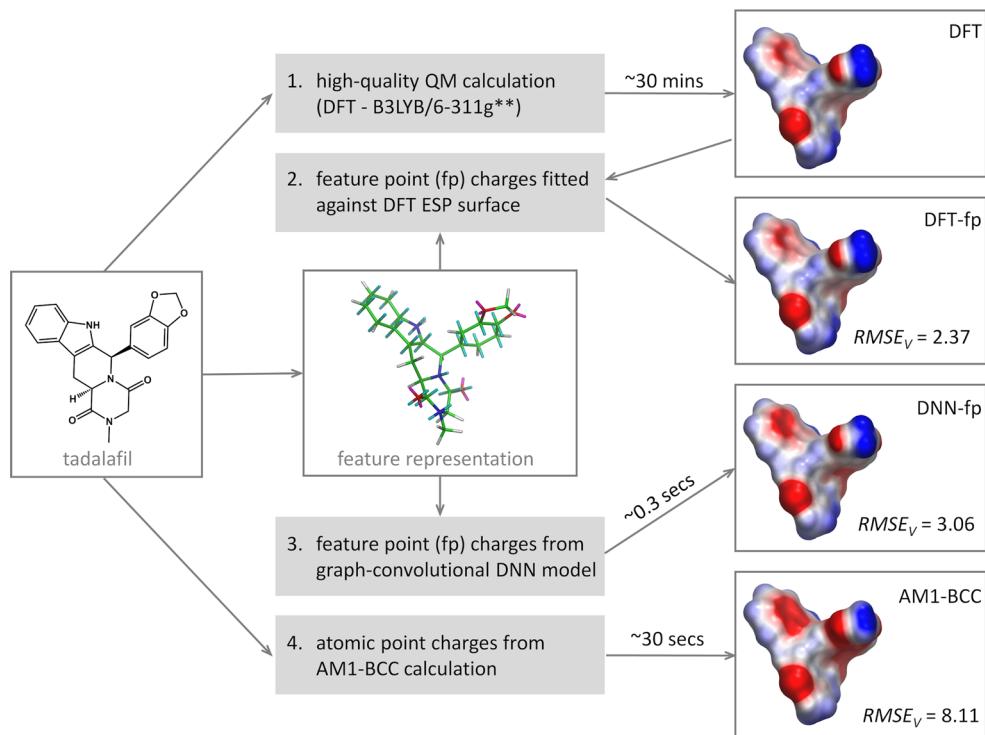


Figure 1. Overview of the methods for generating molecular ESP surfaces discussed in this paper, using tadalafil as a representative example of the relative quality of DFT-fp, DNN-fp, and AM1-BCC ESP surfaces, compared to a DFT ESP surface. The DNN-fp ESP surface approaches the quality of the DFT-fp ESP surface, which in turn is a very good approximation of the DFT ESP surface. The AM1-BCC ESP surface, although qualitatively similar to the other ESP surfaces, has significant differences to the DFT ESP surface, particularly around the benzodioxole ring of the tadalafil molecule. $RMSE_V$ values are in kcal/mol.

charges suitable for molecular mechanics (MM) force fields rather than for producing practical high-quality ESP surfaces. Only atom-centered point charges were used, so features like σ holes and lone pairs were not modeled. As the aim was to derive force field point charges, formally charged groups were modeled without taking into account any shielding of charge in an aqueous environment.⁴ Another recent study used machine learning models, trained on QM calculations, to predict the electrostatic properties of σ holes.¹⁰ The derived models accurately predict the maximum ESP value of a σ hole, but the analysis is limited to halogen atoms.

Here, we introduce a method for generating near-QM quality ESP surfaces for ligand molecules in a time frame suitable for interactive drug design; i.e. as a chemist or modeler edits a structure in a molecular visualizer, the ESP surfaces can be updated essentially instantaneously, with minimal loss in precision. We achieve this by training a graph-convolutional deep neural network (DNN) against high-quality QM calculations for over 100 000 small molecules. The key features of our model are the following: (1) the electrostatics are described by atom-centered point charges as well as off-centered point charges representing p orbitals, lone pairs, and σ holes; (2) point charges are trained specifically to optimally reproduce high-quality QM ESP surfaces; (3) a simple empirical correction is applied to formally charged groups, mimicking the shielding of the charge in an aqueous environment. We believe this represents an ideal application for machine learning methods because we are in full control of the training set data. Additionally, unlike more context-dependent properties like protein–ligand binding affinity, the predicted point charges only depend on the chemical

environment within a molecule. In particular graph-convolutional DNNs should be ideal for learning the rules that govern the distribution of charge within molecules.

In addition, we describe a method for generating protein ESP surfaces that are fully compatible with the ligand ESP surfaces described above. Using an independent test set, we show that the ligand ESP surfaces predicted by the DNN model are in good agreement with high-quality QM and significantly better than ESP surfaces calculated from AM1-BCC¹¹ point charges. In addition, we show that there is a good correlation between the DNN ESP values and experimental properties relevant to drug design, illustrating how this method could be successfully employed by medicinal chemists and modelers.

RESULTS AND DISCUSSION

The **Methods** section explains how we have developed methods for generating fast, high-quality ESP surfaces for ligands and proteins. In this study, we focus on the validation of the method for generating ligand ESP surfaces, which uses a graph-convolutional DNN model, trained on high-quality QM calculations for >100 000 diverse molecules. Validating the protein ESP surfaces is beyond the scope of the current paper.

All QM calculations were run using density functional theory (DFT) at the B3LYP/6-311G** level of theory using B3LYP/6-31G* optimized geometries. This is a popular DFT method¹² and also the default QM method used by modelers at Astex for investigating molecular conformation and molecular electrostatics. In the remainder of this manuscript, we will refer to this method simply as “DFT”. The aim of the DNN model is to produce near-DFT-quality ESP surfaces,

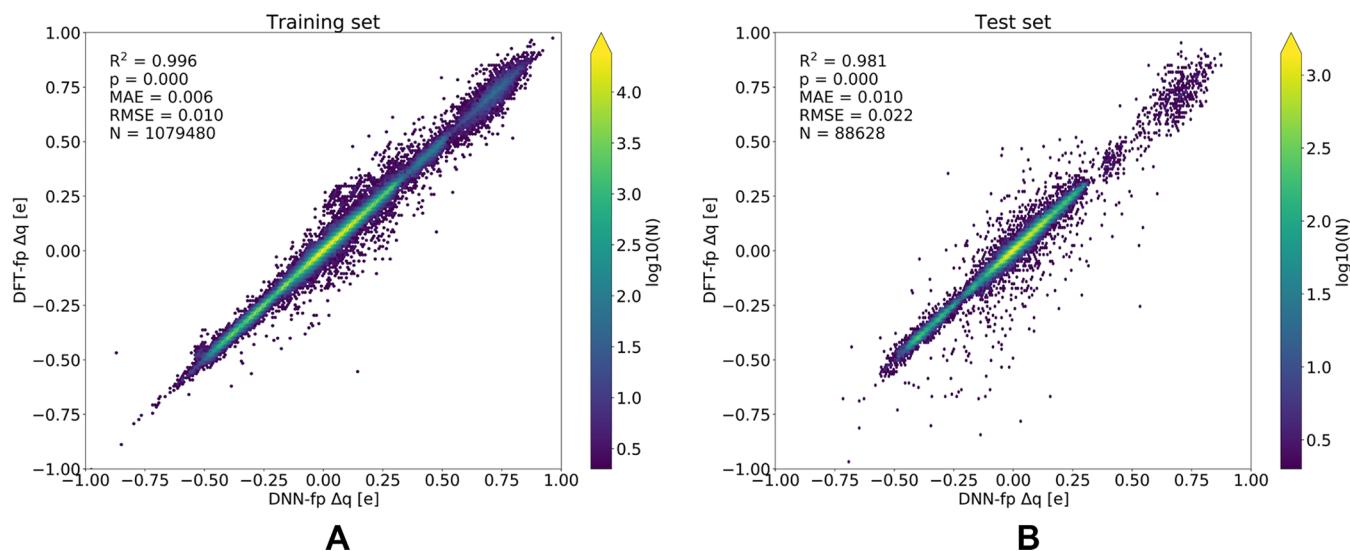


Figure 2. DFT-fp Δq values are plotted against the DNN-fp Δq values for (A) the training set and (B) the test set. Color of a data point on the correlation plots indicates the number of observations (see the corresponding color bar). The plots are generated using the hexbin function of the Python matplotlib library¹⁴ with a grid size of 200.

Table 1. Performance of Point Charge Models in Terms of Reproducing ESP Values on the Molecular Surface across the Compounds in the Test Set^a

	DFT-fp	DNN-fp	AM1-BCC
$\langle \text{RMSE}_V \rangle$ molecular surface	2.81 (0.44)	3.27 (0.62)	7.86 (1.04)
RMSE_V hydrogen features	2.89	3.41	8.52
RMSE_V lone-pair features	3.23	3.97	6.48
RMSE_V p-orbital features	2.57	3.23	9.62
RMSE_V σ -hole features (all)	2.92	3.46	15.68
RMSE_V σ -hole features (Cl, Br, I)	2.92	3.34	18.00

^aThe table shows the average RMSE_V value (kcal/mol) across the entire molecular surface, as well as the RMSE_V values (kcal/mol) for specific feature types, across the test set. Results are shown for DFT-fp ESP surfaces, DNN-fp ESP surfaces, and AM1-BCC ESP surfaces. In each case the DFT ESP surfaces are used as the reference.

using a point-charge model to represent molecular electrostatics.

We will discuss four different methods for generating molecular ESP surfaces (see Figure 1): (1) DFT ESP surface, derived directly from the DFT electron density map and nuclear positions; this will be used as a benchmark for the other methods; (2) DFT-fp ESP surface, a feature-point (fp) charge model optimized to reproduce the DFT ESP surface of a specific molecule; this shows how well DFT ESP surfaces can be described with our point charge model and represents the best possible quality that the DNN-fp model may achieve; (3) DNN-fp ESP surface, a feature-point-charge model predicted by our graph-convolutional DNN model; (4) AM1-BCC ESP surface, a point charge model based on the AM1 method;¹³ AM1 is a semiempirical method and AM1-BCC¹¹ is a popular model for generating point charges to be used in force fields. The DFT-fp and DNN-fp models use point charges on atoms as well as on lone-pair, p-orbital, and σ -hole features (see Methods section). The AM1-BCC model only has atomic point charges. To assess the quality of an ESP surface produced by a particular method, we use RMSE_V , the root-mean-square error (RMSE) in the ESP values across the molecular surface, relative to the DFT ESP surface.

Training and Test Set Construction. In order to derive and validate the graph-convolutional DNN model for the calculation of ESP surfaces for small molecules, we constructed

a training set of molecules, as well as an independent test set. For both sets, we ensured broad chemical coverage and relevance to medicinal chemistry (see Methods section). For each molecule in the training and test set we (1) generated the molecule's electron density map from the DFT calculation, (2) generated a molecular surface, (3) calculated ESP values on the molecular surface from the DFT electron densities and nuclear positions, and (4) fitted excess point charges, Δq (see Methods section), against the DFT ESP values on the molecular surface to create the DFT-fp model.

For the DFT-fp model, the median RMSE_V value across all molecules in the training set is 3.15 kcal/mol (see Supporting Information). We believe this level of precision is more than sufficient for the ESP surfaces to be of utility for interactive drug design, for deriving quantitative structure–activity relationships (QSAR), or for scoring protein–ligand interactions. For comparison, ESP values on surfaces of neutral molecules range from approximately -60 kcal/mol for strong hydrogen-bond acceptors to +60 kcal/mol for strong donors.

Deep Neural Network Model. A DNN model was trained against the DFT-fp charges generated for the training set. Figure 2A shows a plot of the DFT-fp Δq values against the predicted DNN-fp values for the training set. It is clear that the model is able to describe the point charges of the training set molecules well, with a coefficient of determination, $R^2 = 0.996$ and RMSE in the predicted Δq values, $\text{RMSE}_{\Delta q} =$

$0.010e$. Additionally, with an average RMSE_V of 3.30 kcal/mol across the training set (see [Supporting Information](#)), the DNN-fp ESP surfaces approximate the quality of the DFT-fp ESP surfaces. The DNN-fp model was then used to calculate point charges for all molecules in the external test set. [Figure 2B](#) shows that there is good correlation between DNN-fp and DFT-fp Δq values, with $R^2 = 0.981$ and $\text{RMSE}_{\Delta q} = 0.022e$.

For the current work, the point charges predicted by the model are only of indirect relevance, as the primary objective is to generate high-quality ESP surfaces. [Table 1](#) compares the ESP values produced by the DNN-fp model to those obtained directly from the DFT calculation (AM1-BCC results listed in [Table 1](#) will be discussed below). The average molecular surface RMSE_V of 3.27 kcal/mol across all compounds in the test set is comparable to the value obtained for the training set and only slightly above that obtained for DFT-fp point charges. It is clear from the results in [Table 1](#) that all feature types are predicted equally well across the test set. [Figure 3](#) compares

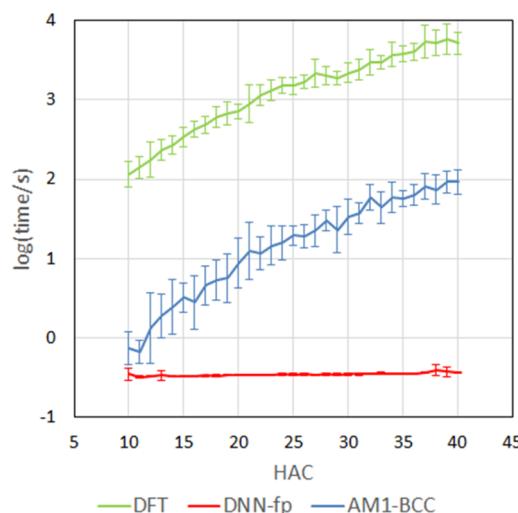


Figure 3. Timings for DFT, DNN-fp, and AM1-BCC calculations as a function of molecular size. The average and standard deviation at each data point are calculated for 10 molecules randomly selected from our internal compound registry. The DFT timings correspond to the runtime for a single point calculation using 6-31G** basis sets. For AM1, the time required for generating a mol2 file with AM1-BCC charges using ANTECHAMBER¹⁵ is reported, and for DNN-fp, the time required for generating a PQR file with charges on atoms and atomic features is reported (the average DNN-fp timing over all molecules is 0.34 s). All calculations were performed on a single core of an Intel Xeon processor E5 v4 (3.00 GHz).

the CPU timings for the DFT calculations and the DNN-fp model calculations as a function of molecular size. For drug-sized molecules ($\text{HAC} \approx 30$) the DNN-fp model achieves a speedup of approximately 4 orders of magnitude, compared to the DFT method. The DNN-fp model is fast enough for interactive design, and we believe there is significant scope for speeding up the method further should this be required.

The DNN-fp model produces Δq values for which the DNN-fp ESP surfaces are calculated. We also tested two much simpler formalisms for generating the Δq values as baseline models in order to further understand the relative performance of the DNN-fp model. In the first of these baseline models, we simply set all Δq values to zero. The $\langle \text{RMSE}_V \rangle$ value we obtain for this baseline model against the test set is 7.66 (2.81) kcal/mol . In the second baseline model, we derive Δq values for all

feature group types as the average Δq value obtained from a constrained point charge fitting procedure across all feature groups of that type in the training set (see [Methods](#) section). For this second baseline model, the RMSE_V value against the test set is 8.43 (3.09) kcal/mol . It is clear that the DNN-fp model significantly outperforms the two baseline models.

With any machine-learning model it is important to test for any evidence of overfitting, in this case due to similarity between compounds in the test and training sets. Here we took two separate approaches to this. First, for each compound in the test set we identified the most similar compound in the training set (see [Supporting Information](#)). We then examined the RMSE_V values as a function of the similarity to the nearest training set compound. This showed that, as expected, the quality of the predicted ESP surfaces does improve slightly with increasing similarity to a compound in the training set (see [Figure S1](#)). However, even for test set molecules that are distant from molecules in the training set, the predictions are excellent. Second, we assigned a scaffold to each compound in the training and test set (see [Figure S2](#)). We then assessed the quality of the ESP surfaces produced for the molecules containing scaffolds that have a representative in the training set vs those molecules whose scaffolds do not. We saw essentially no difference in RMSE_V values between these two groups of compounds (see [Figure S2](#)). We believe these results show that our model can extrapolate to new molecules and new scaffolds.

We are aware of two other studies where researchers have used machine learning methods to derive models predicting partial charges or electrostatic potentials. In their study predicting force field point charges for ligands, Bleiziffer and colleagues reported an RMSE of $0.016e$ on their external test set.⁹ However, it is difficult to compare this to our $\text{RMSE}_{\Delta q}$ value of $0.022e$ directly for two main reasons. First, the two compound sets are very different; our study uses 4814 diverse medicinal chemistry compounds from ChEMBL, whereas Bleiziffer et al. use a set of 1081 approved drug molecules. Second, whereas Bleiziffer et al. predict point charges for each atom, our study predicts Δq values for feature groups (see [Methods](#) section). For comparison, in our study the RMSE error in the point charges placed on individual features is $0.014e$ for atomic (C, N, O, S, P, F, Cl, Br, I, H) features and $0.0045e$ for nonatomic (lone-pair, p-orbital, σ -hole) features. The reason why the errors in the individual feature point charges are smaller is because the error in the feature group Δq values is effectively divided among the feature points in the feature group in the same way Δq is (see [Methods](#) section).

In another study, Heidrich and colleagues derived a machine learning model specifically to predict the potential of σ holes of halogen atoms.¹⁶ The authors obtain $\text{RMSE}_V = 3.82 \text{ kcal/mol}$ for the potential in the direction of the σ hole in chloro, bromo, and iodo substituents. [Figure 4](#) shows the correlation between the DFT ESP value and the ESP value predicted by the DNN-fp model, in the direction of the σ hole, for all chloro-, bromo-, and iodo-containing compounds in our test set, with $\text{RMSE}_V = 3.34 \text{ kcal/mol}$ (see [Table 1](#); similar plots are provided for other feature types in the [Supporting Information](#)). Again, we cannot compare these two results directly, in this case because of differences in the compound sets (Heidrich et al. do not make their sets available), the level of theory, and the method for calculating the molecular surface. However, it is clear that our global model is able to predict the potential at halogen σ holes with a similar precision

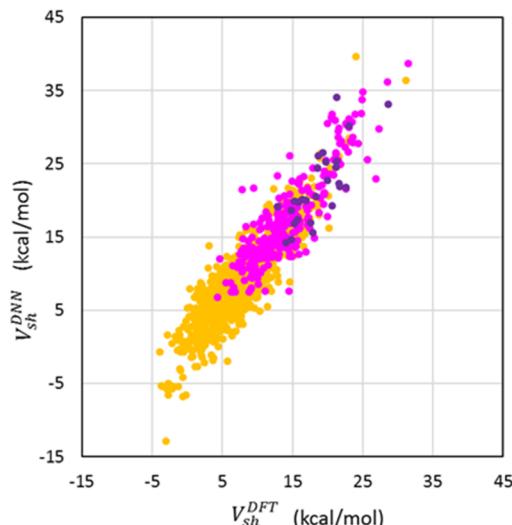


Figure 4. DNN-fp ESP values plotted against the corresponding DFT ESP values for σ holes in chloro (orange), bromo (pink), and iodo (purple) substituents in the test set.

to the highly specific local models obtained by Heidrich et al. for halogen atoms.

For comparison, we also ran AM1¹³ calculations and derived AM1-BCC point charges¹¹ (see *Methods* section) for all molecules in the training set and the test set. While AM1 calculations are significantly faster than DFT, they are still too slow to be used interactively, particularly for drug-sized compounds (see *Figure 3*). By use of DFT as a benchmark, in terms of generating high-quality ESP surfaces, the point charges produced by our model significantly outperform the AM1-BCC point charges (see *Table 1*). The particularly high RMSE_V values for σ holes could be due to the fact that AM1-BCC only places point charges on atoms. It should be possible to address this by introducing, within the AM1-BCC protocol, off-center point charges representing σ holes for halogens and sulfur atoms. It is worth noting that the higher quality DNN-fp ESP surfaces are obtained at a fraction of the computational cost required for AM1-BCC calculations (see *Figure 3*); for drug-sized molecules, the DNN-fp model is approximately 100 times faster than the AM1-BCC method.

Property Predictions. Apart from the ability to deliver high-quality molecular ESP surfaces in a time frame useful for interactive design, the utility of the ESP surfaces is ultimately defined by their ability to provide rational interpretations and correlations for key molecular properties. Here, we will look at four examples where molecular ESP surfaces provide correlations with experimentally determined molecular properties and we will assess the performance of our DNN-fp model in each case.

A. Hydrogen-Bond Basicity. The first of these properties is the relative strength of hydrogen bond acceptors. Hydrogen bonding is of significant importance to drug design and understanding the relative abilities of different acceptors to interact with hydrogen bond donors is highly desirable within the context of rational drug discovery. Here we will use the hydrogen bond basicity pK_{BHX} ¹⁶ which provides an experimentally determined scale of the relative strengths of different acceptors in terms of their ability to form hydrogen bonds. It was shown that molecular electrostatic potentials can be predictive of hydrogen bond basicity.^{17,18} In particular, Kenny

and co-workers published a detailed study on a range of acceptor types,¹⁹ and we will use data from this paper to test if our DNN-fp ESP surfaces also correlate with hydrogen bond basicity. It is outside the scope of the current study to repeat the work reported by Kenny et al. in its entirety. Instead, we selected a few compounds (see *Supporting Information*) from the data sets used in that study, representing four series of compounds: pyridines, acetophenones, nitriles, and ethers. *Figure 5A* shows the pK_{BHX} values plotted against the lone-pair potentials generated using the DNN-fp model for each of these four series of compounds. It is clear that good correlations are observed between the predicted potential for the lone pair and the pK_{BHX} values for each of the chemical series. It is interesting to note that we have simply used the ESP value in the direction of the lone pair on the molecular surface rather than identifying the minimum value of the ESP around each acceptor, which was the approach adopted by Kenny et al. Our approach still provides good correlations but may have led to bigger differences between chemical series.

B. pK_a Values. The second property we investigate is acidity of functional groups, which can affect key medicinal chemistry parameters like solubility and affinity. It was shown that pK_a values of acidic groups correlate with electrostatic potentials, either at the hydrogen atom of the acidic group or at the lone pair of the conjugate base.^{20–22} Here, we used data for a number of 4-substituted pyridines and aliphatic carboxylic acids (see *Supporting Information*), extracted from a pK_a database,²³ to assess whether we see a similar correlation for ESP values generated using our DNN-fp model. *Figure 5B* shows that there is a clear correlation between the lone pair ESP value of the (conjugate) base and the corresponding pK_a value for both the pyridines and carboxylic acids. Similar correlations are also seen when we use the potential of the hydrogen atom of the protonated species (see *Table 2*). As we observed for the hydrogen bond basicity values, separate regression models are required for each functional group. Although outside the scope of the current work, it would seem that graph-convolutional DNN methods could be ideally suited to train atomic environments directly against pK_a values or pK_{BHX} values to produce global models for these properties.

C. Factor Xa Affinity. The third example involves SAR observed for binding affinities against coagulation factor Xa (fXa). *Figure 6A* shows the protein ESP surface for the fXa active site. We generate protein ESP surfaces using a completely analogous point charge model, trained on DFT calculations at the same level of theory as the ligand ESP surfaces (see *Methods* section). As a result, protein ESP surfaces can also be generated interactively and are completely compatible with the ligand surfaces. Our focus here is on the P4 pocket, which is faced by the π clouds of three aromatic residues. Apart from being lipophilic in nature, this pocket is also quite electronegative and this is clearly visible from the protein ESP surface. The pocket is well-known to interact favorably with electron-poor or even positively charged lipophilic groups on ligands.²⁴ Young and colleagues observed this clearly on a series of closely related compounds where a substituted phenyl is placed in the P4 pocket.²⁵ The variable R-substituent on the phenyl ring (see *Figure 6A*) is mostly pointing toward solvent, which makes this a good test case to assess the effect of the electrostatic nature of the phenyl ring on affinity. The authors hypothesized that the SAR they observed could be driven by the close contact between the electronegative π cloud of Trp215 and the aromatic proton on

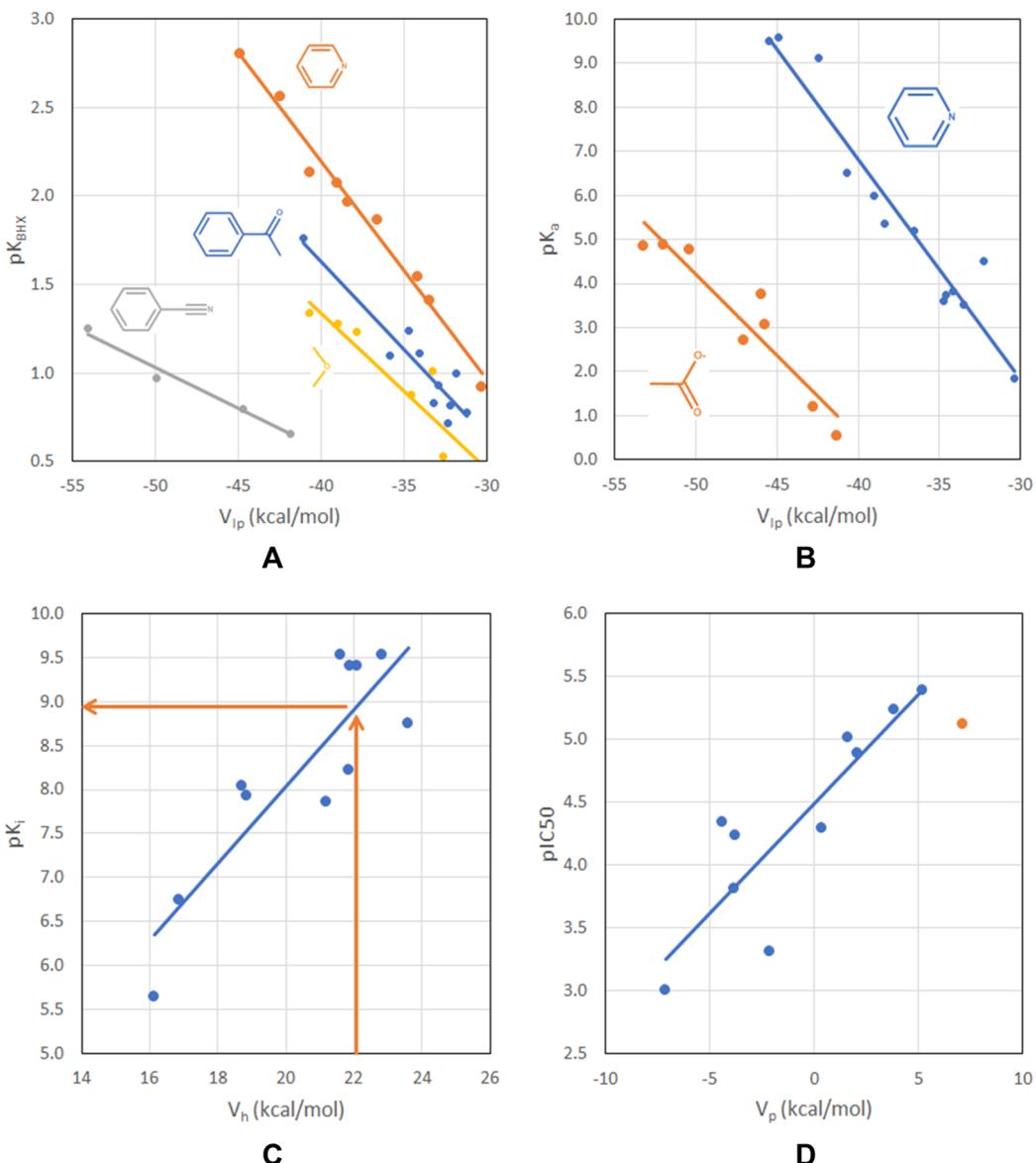


Figure 5. Correlations between DNN-fp ESP values and four experimental properties: (A) hydrogen bond basicity pK_{BHX} of pyridines (orange), acetophenones (blue), nitriles (gray), and ethers (yellow); (B) acidity constant pK_a for pyridines (blue) and carboxylic acids (orange); (C) fXa affinity pK_i , where the orange arrows indicate the prediction for a compound where the ligand phenyl ring in Figure 7A is replaced by a 2-pyridine ring; (D) XIAP affinity pIC_{50} , where only indoline compounds (blue points) were used in the regression and where the orange point represents the chloroazaindoline derivate (compound 17 in the study by Chessari et al.²⁷).

the ligand in the meta position to the R-substituent (see Figure 6A). Hence, using our DNN-fp model, we calculated the ESP values V_h in the direction of the C–H proton on the molecular surface of the ligands investigated by Young et al. and plotted them against the observed fXa affinity of the compounds (Figure 5C). Supporting the authors' hypothesis, a good correlation is obtained, with $R^2 = 0.76$ and an RMSE between the fitted line and the experimental pK_i values of 0.59. Figure 7A illustrates the effect that the substituents have on the ESP of the C–H proton pointing at Trp215. However, Figure 7A also shows that the entire phenyl ring becomes more electropositive as the R-substituent gets more electron withdrawing. Therefore, we also tested the correlation of the average ESP value $\langle V \rangle$ of the three C–H groups interacting with the electronegative bottom of the S4 pocket (see Figure 6A and Figure 7A). Interestingly, this ring-average potential provides a comparable R^2 value of 0.80 when plotted against

the fXa affinity. Hence, it is difficult to conclude that the effect of the R-substituent on fXa affinity is driven entirely by the electrostatic nature of the meta aromatic proton.

D. XIAP Binding Affinity. The final property we investigate is binding affinity for the X-linked inhibitor of apoptosis protein (XIAP). Figure 6B shows the ESP surface of the XIAP SMAC binding site. During a fragment-to-leads campaign against this target, Chessari and co-workers identified the weakly binding ($IC_{50} \approx 500 \mu M$) early lead compound shown in Figure 6B.²⁷ The authors noted that the π system of the indoline ring contacts an electronegative patch on the protein ESP surface, which was considered an electrostatic clash; this negative patch is also clearly visible in our DFT-derived ESP surface for XIAP (Figure 6B). Chessari et al. showed that the XIAP affinity can be greatly improved by optimizing the electrostatics of the indoline ring and that affinity tracked very well with the electron donating/withdrawing nature of the R-

Table 2. Correlation between ESP Values on the Molecular Surface against Hydrogen Bond Basicity, Acidity Constants, and Protein–Ligand Binding Affinity^a

Hydrogen Bond Basicity (pK_{BHx})						
	N	R_{DFT}^2	R_{DNN}^2	R_{AM1}^2	$R_{DNN-DFT}^2$	$R_{AM1-DFT}^2$
pyridines V_{lp}^b	9	0.981	0.983	0.930	0.969	0.967
acetophenones V_{lp}^b	10	0.896	0.869	0.807	0.976	0.959
nitriles V_{lp}^b	4	0.986	0.978	0.929	0.979	0.962
ethers V_{lp}^b	8	0.971	0.941	0.927	0.981	0.866
Acidity Constant (pK_a)						
	N	R_{DFT}^2	R_{DNN}^2	R_{AM1}^2	$R_{DNN-DFT}^2$	$R_{AM1-DFT}^2$
pyridines V_{lp}^b	13	0.929	0.931	0.900	0.974	0.967
pyridines V_h^c	13	0.898	0.869	0.055	0.982	0.093
carboxylic acids V_{lp}^b	8	0.674	0.886	0.770	0.470	0.959
carboxylic acids V_h^c	8	0.963	0.941	0.948	0.945	0.975
Protein–Ligand Binding Affinity (pK_t/pIC_{50})						
	N	R_{DFT}^2	R_{DNN}^2	R_{AM1}^2	$R_{\sigma}^2 h$	$R_{DNN-DFT}^2$
fXa V_h^d	11	0.720	0.757	0.454	0.897	0.955
fXa $\langle V \rangle^e$	11	0.855	0.791	0.655		0.957
XIAP V_p^f	10	0.715	0.761	0.707	0.774	0.985
XIAP V_h^g	10	0.798	0.789	0.721		0.958
						$R_{AM1-DFT}^2$

^a R^2 values are given between the experimental property and the ESP values for DFT, DNN-fp, and AM1-BCC. Additionally, for the DNN-fp and AM1-BCC ESPs, R^2 values are given when plotted against the DFT ESPs. Feature ESP values were calculated as outlined in the Methods section.

^bESP value in the direction of the lone pair. For carboxylic acids, these calculations were on negatively charged molecules. ^cESP value in the direction of the acidic proton. For pyridines, these calculations were on positively charged molecules. ^dESP value in the direction of the aromatic proton on the phenyl ring (solid circle in Figure 6A and Figure 7A). ^eAverage ESP value of the aromatic protons and p orbitals for three atoms on the phenyl ring (dashed ellipsoid in Figure 6A and Figure 7A). ^fAverage ESP value of the aromatic p orbitals for two atoms in the indoline ring (solid ellipsoid in Figure 6B and Figure 7B). ^gAverage ESP value of two protons on a CH₂ group in the indoline ring (dashed circle in Figure 6B and Figure 7B).

^hHammett substituent constants (σ_m for fXa and σ_p for XIAP). We were unable to find some of the Hammett constants used by Young et al.²⁵ in the provided literature reference,²⁶ so the R^2 value for fXa was taken directly from their paper. For XIAP, we used the σ_p values listed in Table 2 of Chessari et al.'s paper²⁷ to calculate R^2 .

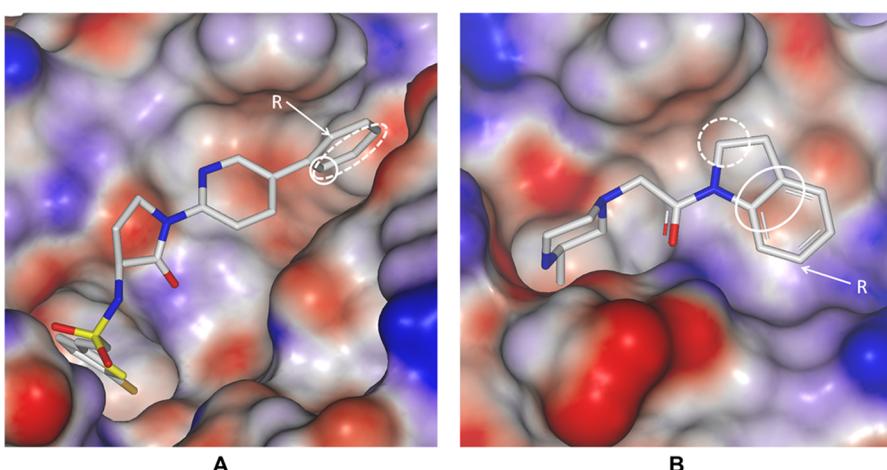


Figure 6. DFT-trained protein ESP surfaces (see Methods section) for (A) the fXa active site (PDB code 2VH6) and (B) the XIAP SMAC binding site (PDB code 5C7A). For both targets a ligand is shown and the R-substituent location is indicated. Also, the areas where literature hypotheses are focused in terms of how affinity is affected by electrostatics are highlighted as solid white ellipsoids (V_h for fXa and V_p for XIAP in Table 2). The alternative/additional hypotheses tested in the current work are highlighted as dashed white ellipsoids ($\langle V \rangle$ for fXa and V_h for XIAP in Table 2).

substituent (see Figure 6B). Hence, using our DNN-fp model, we calculated the average ESP value V_p of the two central carbon atoms' p orbitals in the indoline ring (which directly contact the negative patch on the protein ESP surface) for all R-substituents reported by Chessari et al. and plotted them against the XIAP affinity (see Figure 5D). Again, consistent with the postulated hypothesis, we see a good correlation with $R^2 = 0.76$ (see Table 2) and an RMSE between the fitted line and the experimental pIC_{50} values of 0.37. Figure 7B illustrates the effect of the R-substituent on the DNN-fp ESP surface of

the ligands. It is interesting to note that, as well as affecting the electronics of the aromatic ring, the R-substituent also significantly modulates the potential around the pyrrolidine CH₂ groups (see Figure 7B). It is impressive that the DNN-fp model is able to capture this relatively long-range effect, which is also observed in the DFT ESP surfaces (not shown). One of the pyrrolidine CH₂ groups stacks directly against the π cloud of Trp323 (see Figure 6B), so any electrostatic changes in this part of the ligand may also be reflected in the XIAP affinity of the compounds. Therefore, we also plotted the ESP value V_h of

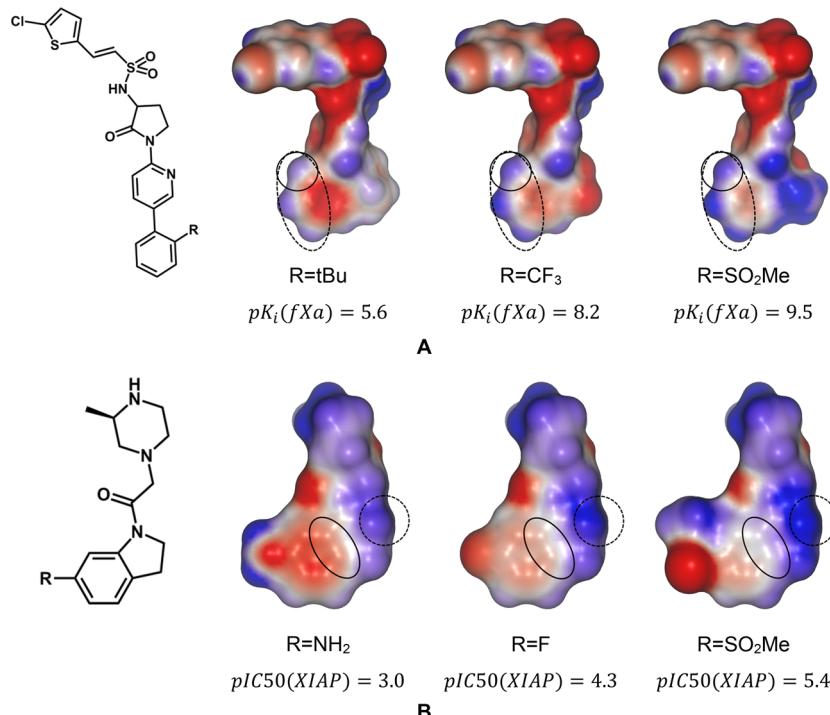


Figure 7. ESP surfaces generated by our DNN-fp model for (A) three of the fXa compounds reported by Young et al.²⁵ and (B) three compounds reported by Chessari et al.²⁷ Solid line circles/ellipsoids indicate the areas the authors proposed to be linked to the correlation between ESPs and affinity. Dashed line circles/ellipsoids represent areas of alternative/additional hypotheses we test here.

the CH_2 hydrogens against the XIAP affinity and obtained a similar R^2 value of 0.79 (see Table 2). It is difficult to know whether the interaction with the negative patch or with the Trp323 π cloud has a larger effect on the XIAP affinity, but it is likely that both contribute.

For all molecules used for the property-ESP hypotheses/correlations listed in Table 2, we also ran a DFT calculation and repeated all the correlations using DFT ESP values. In nearly all cases, the DFT calculations provide comparable correlation coefficients to the DNN-fp model. The only exception is the pK_a prediction from the lone pair ESP values for carboxylic acids, where DFT performs significantly worse ($R^2 = 0.67$) than the DNN-fp model ($R^2 = 0.89$). We are not sure what the exact reason behind this is, but it might be related to the fact that the DFT calculations were done on formally negatively charged molecules, whereas the DNN-fp model uses an empirical model to correct the partial charges it predicts for the neutral species (see Methods section). Importantly, excellent correlation is observed between the DNN-fp ESP values and the DFT ESP values in all cases except, again, the pK_a prediction from the lone pair ESP values for carboxylic acids.

The same exercise was repeated using ESP values calculated using AM1-BCC charges (see Table 2). In most cases, AM1-BCC provides useful models with correlation coefficients similar to DFT and DNN-fp. One notable exception to this is the pK_a prediction from the proton ESP values for pyridines, where AM1-BCC performs very poorly. The correlation plot (see Figure S9 in the Supporting Information) shows that there appear to be two separate groups of compounds that display distinct correlations. It is beyond the scope of this work to investigate this further, but it is interesting to note that again we are measuring ESP values on formally charged species. Closer inspection of the data in Table 2 reveals that for 11 out

of the 12 hypotheses tested our DNN-fp model provides a higher R^2 value than the AM1-BCC method ($p = 0.003$; exact binomial test). We believe that this is a direct consequence of the higher precision achieved by the DNN-fp model. This is supported by the fact that for 10 out of the 12 cases, the correlation with DFT ESP values is better for the DNN-fp model than it is for the AM1-BCC method ($p = 0.02$).

In order to test to what extent molecular conformation affects the quality of the ESP surfaces produced by our DNN-fp model, we generated 100 conformers for the *tert*-butyl fXa compound in Figure 7A. Next, we generated DFT ESP surfaces as well as DNN-fp ESP surfaces for all conformers and calculated the corresponding RMSE_V values. Across the 100 conformers, the average RMSE_V value was 4.03 kcal/mol with a standard deviation of 0.22 kcal/mol. For the two most “folded up” conformers, the RMSE_V values were 4.45 and 4.02 kcal/mol. For comparison, the RMSE_V value for the Corina geometry is 3.90 kcal/mol. These results indicate that the quality of the ESP surfaces is not particularly dependent on the molecular conformation.

Correlation plots such as those in Figure 5C and Figure 5D can be used to predict the SAR of closely related series of compounds. These examples illustrate how we believe this methodology will be at its most powerful: for the understanding and prediction of local SAR, e.g., modifying a single small substituent or variation of heteroatoms in an aromatic ring. We do not expect the ESP surfaces in isolation to be particularly useful as a general scoring function for predicting protein–ligand binding, because of confounding factors like lipophilic interactions, water molecules and ligand conformational effects.

Both Young and colleagues and Chessari et al. reported good correlations between affinity and Hammett substituent constants^{26,28} for the fXa and XIAP examples, respectively.^{25,27}

It is clear from Table 2 that the R^2 values obtained for the Hammett constants are comparable to the ones we obtain using ESP values. However, although Hammett constants are clearly very valuable for assessing SAR, there are key advantages to using ESPs. Most importantly, the use of ESPs is not restricted to studying substituents on a single aromatic ring type. For example, if we consider a virtual compound designed against fXa where the ligand phenyl ring in Figure 7A is replaced by a 2-pyridine ring, i.e., where the aromatic carbon forming the attachment point for the R-group is replaced by a nitrogen atom, we cannot use standard benzoic acid derived Hammett constants to make predictions about its affinity. However, using the DNN-fp model, we can simply calculate V_h (orange line in Figure 5C) and use it to predict that this compound would have $pK_i(fXa) \approx 9$. Similarly, Chessari et al. also reported the XIAP affinity of the chloroazaindoline derivate (compound 17 in their study). This compound could not be included in the correlation plot the authors reported against the Hammett constant. However, using the DNN-fp ESPs, we can add the compound to the correlation plot in Figure 5D (orange circle) and demonstrate that it fits the trend observed for the other compounds. Another advantage of the use of ESPs over Hammett constants is that the ESP surfaces provide a much more detailed spatial view of the electrostatics of the ligand in the context of those of the protein, allowing more specific hypothesis generation.

The correlations observed for protein–ligand affinity are weaker than they are for hydrogen bond basicity and pK_a values. This is to be expected as the latter two properties depend more directly on the electrostatic properties of the molecules. For protein–ligand binding, other factors will generally play a role. For example, for fXa the R-substituent may form additional interactions with the protein, affect the conformation of the ligand, or change the water structure around that part of the active site, which will perturb the correlation between affinity and the electrostatics of the ligand phenyl ring. In fact, for most protein–ligand affinity SAR examples, these perturbing factors will be more significant, and the relatively clean and direct correlations we observe between affinity and ESP values for fXa and XIAP are probably quite unusual. However, that in no way diminishes the utility to drug discovery of high-quality ESP surfaces for proteins and ligands. Even in a multifactorial SAR scenario, these surfaces provide an excellent indication of the electrostatic complementarity between protein and ligand, and how the ligand may be altered in order to optimize affinity for the protein.

CONCLUSIONS

We have developed methods for generating fast, high-quality ESP surfaces for proteins and ligands. For ligands, we have trained a graph-convolutional DNN model against DFT ESP surfaces for over 100 000 diverse molecules. Testing the model against an independent test set of molecules relevant to medicinal chemistry showed that it provides good agreement with DFT ESP surfaces. We showed that in terms of the quality of the ESP surfaces, the DNN-fp model significantly outperforms the AM1-BCC model. The DNN-fp model generates ESP surfaces in a fraction of a second, thus fully enabling their use for interactive drug design. For a drug-sized molecule, the AM1-BCC method is approximately 100-fold slower and DFT ESP surfaces are 4 orders of magnitude slower to generate. For proteins we developed a method for generating fast DFT-trained ESP surfaces that are fully

compatible with the ligand DNN-fp ESP surfaces. Next, we assessed the performance of our DNN-fp model in terms of generating ligand ESPs that correlate with experimentally measured molecular properties. These included hydrogen bond basicity, acidity constants, and two examples of protein–ligand binding affinity. In each case, a strong correlation was observed between the DNN-fp ESP values and the experimental property. We showed that correlation coefficients were generally comparable to those obtained with DFT ESP values and better than those observed for AM1-BCC ESP values. We believe this new method, which we make freely available via https://github.com/AstexUK/ESP_DNN, will form a powerful tool for interactive drug design.

METHODS

Constructing Data Sets. Training set molecules were selected from a superset of commercially available compounds in eMolecules²⁹ and the Sigma-Aldrich³⁰ catalogue in a way to maximize the coverage of different chemical environments. Molecules with $HAC \leq 17$ and consisting exclusively of elements B, Br, C, Cl, F, H, I, N, O, P, and S were considered. All molecules were neutralized, and up to eight tautomers were generated for each molecule using the MolVS³¹ Python library built upon the RDkit chemoinformatics toolkit (version 2017.09.1).³² This resulted in a set of 1 336 480 molecules for further filtering. The local chemical environment of an atom was represented as atom-centered fragments with a varying radius for including neighboring atoms. As such, fragments of increasing radii between zero (only the central atom) and six (the central atom and all atoms that are up to six bonds away) were created for all molecules in the set. All such fragments that occur at least 10 times across the set were considered further. The smallest molecule that contained a particular fragment was set as representative for the fragment type. Next, the smallest set of representative molecules that includes all chemical environments of radius up to 2 was selected. As the number of chemical environments rapidly increases with the radius of the fragment, it was impractical to include all chemical environments with a radius larger than 2, and hence only a subset of chemical environments was included. We selected a subset of molecules that cover 90% of all the atoms in the set in terms of atomic environments with radius 3. Similarly, we selected subsets of molecules covering 85%, 60%, and 50% of all atoms in the set in terms of chemical environments of radii 4, 5, and 6, respectively. This resulted in a training set of 100 500 molecules.

An external test set of diverse lead-like molecules was constructed from the ChEMBL database version 20.³³ Lead-sized ChEMBL molecules ($10 \leq HAC \leq 25$) that have binding affinity of $<1 \mu M$ against at least one target were candidates for the test set. Molecules were grouped according to HAC, and for each group K-medoids clustering was performed to create a predetermined number of clusters using the pyclust library for Python.³⁴ Pairwise Dice-distance between the Morgan fingerprints (radius = 3) of two molecules, calculated using the RDkit toolkit, was used as the distance criteria for K-medoid clustering. For each group of molecules (based on HAC), the number of clusters was set to 2.5% of the total number of molecules in the group with lower and upper bounds on the number of clusters set to 100 and 500, respectively. Cluster representatives from each cluster formed the test set. Similar to the training set, up to eight tautomers for each molecule in the test set were generated. This resulted in a test set of 4815 molecules.

Feature Groups. Our approach describes each non-hydrogen atom in a molecule as a feature group. These feature groups consist of a set of features: the parent atom itself, its bound hydrogen atoms, and additional features representing lone pairs, p-orbitals, or σ -holes, each of which carries a point charge. Atom typing is achieved via feature group definitions. Each feature group definition uses a SMARTS³⁵ pattern to assign a feature group type to any atom that matches that pattern. A feature group type provides the rules for turning a matching

atom into a feature group; i.e., it specifies which features are to be added and what their relative geometries are.

A set of small neutral model compounds (see *Supporting Information*) was used to prepare all feature group types and feature group definitions. For each model compound, DFT electron density isosurfaces were generated using the method described below. Each feature group type is defined by a specific model compound. For example, ethane defines the “ $-\text{CH}_3$ ” feature group type, methanol the “ $-\text{OH}$ ” feature group type, etc. Parent atom features and hydrogen atom features are placed exactly on the corresponding atoms. For each feature group type, we then decided which additional features to include, as well as the directions they point in, taking into account atomic orbitals, local symmetry, and the location of electrostatic extrema on the DFT ESP isosurface.

Having decided on the features’ relative directions, we then identified the optimal separation distance between parent atom and nonatomic features. In order to achieve this, we first measured the distance D_f from the parent atom to the ESP surface in the direction of each of the features. Next, we placed all nonatomic features at a distance $L_f = C \cdot D_f$ from the parent atom, in the predefined directions. We then systematically varied C , calculating RMSE_V using unconstrained point charges (see below) for all model compounds and found that $C = 0.5$ is optimal. Hence, for all results presented here, we used $L_f = 0.5D_f$. The model compounds were also used to set features’ van der Waals radii R_f for all feature group types. For each feature in a feature group type, except the parent atom feature, we used $R_f = D_f - L_f$. For the parent atom feature itself, we used the distance from the atom to the closest isosurface point that is at least 40° away from every nonatomic feature in the feature group.

For our DNN-fp representation, point charges of individual features are constrained to the point charges of the other features within the same feature group. This is achieved by defining the charge on feature i in a feature group as follows:

$$q_i = q_i^0 + f_i \Delta q$$

where q_i^0 is the reference charge of feature i , Δq is the excess charge on the feature group, and f_i is the fraction of Δq that is assigned to feature i . For all the work we describe here, we used $f_i = 1/n$ where n is the number of features in the feature group. This way, the excess charge is distributed uniformly across the features in the feature group. We have found that constraining the features’ point charges within a feature group, relative to each other, prevents overfitting and reduces noise levels in the fitted point charges, thus providing the machine learning method the optimal opportunity to detect the true signal in the training set.

Reference charges are normalized such that

$$\sum_{i=1}^n q_i^0 = 0$$

Reference charges are constants that form part of feature group definitions and were derived from the molecules in the training set. First, unconstrained point charges were generated for all molecules in the training set (see below). Next, for each feature group type, e.g., “ $-\text{CH}_3$ ”, we calculated the median point charge placed on each feature type (“C” and “H”) across the training set. These median charges were then normalized such that they sum up to zero across the feature group.

As the model compounds are all neutral molecules, we still need to assign reference charges to formally charged feature group types. For some feature group types we can simply copy the reference charges from the corresponding neutral feature group. For example, the “ $-\text{Op}$ ” feature group type (with associated reference charges) is assigned to both neutral carbonyl atoms, as well as negatively charged carboxylate oxygen atoms. However, a number of formally charged feature group types do not have a direct neutral equivalent. For such feature group types, we derived reference charges from the corresponding neutral species. In this simple two-step, highly empirical process, using the “ $-\text{NH}_3$ ” feature group as an example, we first take the reference charges for the nitrogen (-0.26) and the

hydrogens ($+0.19$) from the “ $-\text{NH}_3$ ” feature group definition. We cannot use these charges as reference charges for the “ $-\text{NH}_3$ ” feature group directly because they do not add up to zero but instead to $\Delta q^0 = -0.26 + 3 \cdot 0.19 = 0.31$. Therefore, in the second step, the reference charges are renormalized by simply subtracting $0.25 \cdot \Delta q^0 = 0.0775$ from each of the feature point reference charges such that the sum over the new reference charges is zero. The newly derived reference charges for the NH_3 feature group are then -0.33 for the nitrogen and $+0.11$ for the hydrogens.

ESP Surfaces. 3D structures of the model compounds, as well as all molecules in the training set and test set, were generated using the CORINA program (version 4.2.0) with default settings.³⁶ Next, 3D DFT ESP maps were generated for all molecules using the Q-Chem software package (version 5.0),³⁷ using the DFT method at the B3LYP/6-311G** level of theory and the B3LYP/6-31G* optimized geometries. The grid size used for the ESP maps was 0.25 \AA . For all model compounds, we defined the molecular surface directly from the DFT maps as the electron density isosurface at 0.002 au . For the training and test set molecules and all molecules used in the examples we discuss, we generated PQR files³⁸ containing the positions and radii of all atomic and nonatomic features. Next, we converted the PQR files into Connolly surfaces³⁹ using the NGL⁴⁰ viewer’s “av” surface mode.⁴¹

Figure 1 shows the four different methods for producing molecular ESP surfaces that are discussed in this work. DFT ESP surfaces were generated by trilinear interpolation of DFT ESP maps for each point of the Connolly surface. The resulting DFT ESP surfaces were considered ground truth for this work.

DFT-fp ESP surfaces are based on feature point (fp) charge representations that provide an optimal fit of the DFT ESP surface. As such, a least-squares fitting to the DFT ESP surfaces was used to derive partial charges on all feature positions in a molecule with a constraint that the total charge of the molecule must add up to zero. The partial charges model we use to describe the ESPs uses the classical Coulomb equation:

$$E = \sum_i \frac{1}{4\pi\epsilon_0} \left(\frac{q_i}{r} \right)$$

where ϵ_0 is the relative permittivity (we used a distance dependent relative permittivity $\epsilon_0 = r$) and q_i is the partial charge on atom i . Our rationale for using a distance dependent relative permittivity here is twofold. First, it provides better fits against the DFT ESP surfaces (for the set of model compounds, $\text{RMSE}_V = 2.64 \text{ kcal/mol}$ using $\epsilon_0 = r$, whereas $\text{RMSE}_V = 4.73 \text{ kcal/mol}$ using $\epsilon_0 = 1$). Second, we want to use a distance-dependent dielectric when the model is used prospectively to generate ESP surfaces for small molecules and proteins (to account for the shielding of charge in an aqueous environment), and the electrostatic function we use to derive the model should be consistent with that.

Two distinct types of DFT-fp surfaces were used in this work. In the unconstrained point charges approach, the partial charge on each feature was fitted independently as a free parameter during the least-squares fitting. For the constrained point charges approach, Δq values for each feature group in a molecule were fitted and partial charges on the features were constrained to those of other features in a feature group (see above). The unconstrained point charges approach was used for the model compounds and for deriving reference charges from the training set. The constrained point charges approach was used everywhere else.

The DNN-fp ESP surfaces use the constrained point charges model (see above), and Δq values are predicted by our graph convolutional DNN-fp model (see below). Since the DNN-fp model was trained on neutral molecules, we adopted the following scheme for deriving Δq values for a charged molecule. We first neutralize all formally charged atoms in the molecule. Next, we use the DNN-fp model to generate Δq values for all atoms in this neutralized molecule; we will refer to these as Δq_n values. We then obtain the final Δq values for all atoms in the original charged molecule, as follows:

$$\Delta q = \Delta q_n + Aq$$

where Q is the formal charge on the atom. A protonated primary amine ("−NH₃⁺"), for example, will have $Q = +1$, but for groups where the formal charge is delocalized, the partial formal charge will be used. For example, an oxygen in a negatively charged carboxylate group will have $Q = -0.5$. A is an attenuation factor, mimicking the shielding of formally charged functional groups in an aqueous environment, which was empirically set to 0.4; this highlights charged parts of the ESP surface, without dominating the regions around them.

DFT-derived ESP surfaces for proteins are calculated in a completely analogous way to ligand DNN-fp ESP surfaces. The only difference is that for protein atoms, the Δq values are taken directly from a precalculated list that was derived from DFT calculations for protein functional groups (see below).

Finally, AM1-BCC ESP surfaces are calculated from AM1-BCC point charges. AM1-BCC is a charge model wherein charges are calculated using AM1 population analysis, to which the bond charge corrections are added, which were parametrized for reproducing the HF/6-31G* ESPs of a set of molecules.¹² For the test set, atomic AM1-BCC partial charges were calculated using the Antechamber module of the AMBER⁴² software. ESP values on the NGL surface points were then calculated using the method described above, with $\epsilon_0 = 1$.

Ligand Point Charges from DNN-fp Model. A molecular graph convolutional DNN as shown in Figure 8 was built using the Keras⁴³

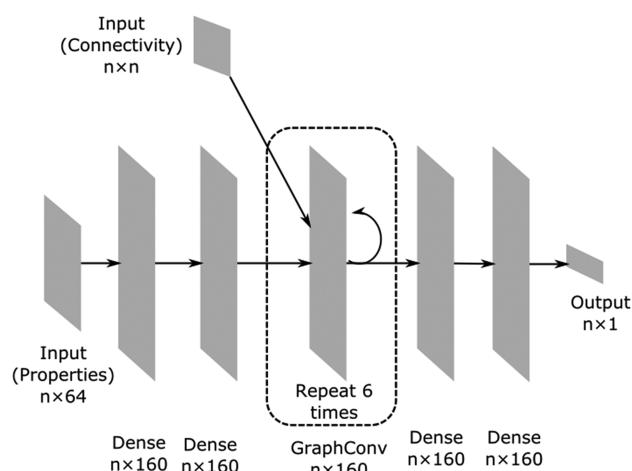


Figure 8. DNN model architecture.

library (version 2.2.4) for Python; TensorFlow⁴⁴ (version 1.7.0) was used as the back end. The model receives 64 simple atomic properties (see Table S6 in the Supporting Information), and atom connectivities were calculated using RDKit.³² Both inputs were arranged as 3D tensors. Given m molecules, the largest of which has n atoms, the property tensor has dimension $m \times n \times 64$. The connectivity is represented by a Boolean tensor of dimension $m \times n \times n$. Both tensors were padded with zeros for molecules with fewer than n atoms. Similarly a 3D tensor of fitted Δq values of dimension $m \times n \times 1$ was used as ground truth, padded with a special "not a number" (NaN) value for molecules with fewer than n atoms; the NaN values were ignored in the loss calculation. A new custom convolutional layer was created, which performs a dot product of the input and a kernel (input width × output width) as the first step. For each atom, output from its neighboring atoms (connected by covalent bonds) is added to it: a graph convolution. Finally an activation function is applied before the values are fed into the next layer. The final model consists of two fully connected layers sequentially connected to input property layer, followed by six custom GraphConv layers, each of which gets the connectivity matrix as an input in addition to the preceding layer. The six GraphConv layers are followed by two fully connected layers.⁴⁵ All these layers have a width of 160 and use rectified linear unit⁴⁵ as

an activation function. The last fully connected layer feeds into the fully connected output layer of width = 1 (1 Δq value for each atom) with a linear activation function. The hyperparameters (width of the hidden layers, type of the activation function, and the number of GraphConv layers) described above were identified using grid search algorithm. For hyperparameter optimization, a random subset of the training set (20%) was set aside for validation. The identified best set of hyperparameters were then used for the final model generation, wherein the whole training set was used for learning and the external test set was used for evaluation of the model. For training, a custom loss function was used, which calculates mean absolute error between true and predicted Δq values ignoring pairs of values with NaNs as the true values. The model was trained using the Adam optimizer⁴⁶ with learning rate of 0.0025, decay of 0.0005, and minibatches of 128 molecules for 1000 steps.

Protein Point Charges. Using the constrained point charges model (see above), we precalculated Δq values for feature groups occurring in protein amino acids. For this purpose we used a set of specific model compounds representing backbone and side chain groups in proteins. For simplicity, we have assumed that there is no charge transfer between side chain and main chain or between adjacent residues. We established empirically that this approximation has very little effect on the ESP surface of the protein.

The model compounds used to generate charges are listed in the Supporting Information. For backbone atoms, we used capped tripeptides, where we constrained the Δq values of the atoms in the central amino acid to add up to zero. Specific tripeptides were used to model glycine and proline. All other amino acid backbone atoms were modeled using an alanine as the central residue and constraining the side-chain methyl to have $\Delta q = 0$. Side chains were modeled by capping them with a methyl group representing the C_δ atom of the protein backbone and constraining this methyl to have $\Delta q = 0$. For example, the serine side chain was modeled as an ethanol molecule, and the Δq value of the methyl group was constrained to be zero.

Calculating Feature ESP Values. For each feature in a molecule, we measure the corresponding feature ESP value on the molecular surface. For a non-parent-atom feature (hydrogen, lone pair, σ-hole, and p-orbitals), the feature ESP value was measured as the average ESP value over all surfaces points that are within angular distance of 10° from the feature vector (a vector between the parent atom and the feature). For an atomic feature, the ESP value on its nearest of all surface points excluding the surface points that are within 40° from any of its nonatomic children features was considered the feature ESP value.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: [10.1021/acs.jmedchem.9b01129](https://doi.org/10.1021/acs.jmedchem.9b01129).

Tables providing data on performance on the training set, feature and feature group type definitions, and atomic properties used for the DNN-fp model; figures showing the effect of similarity to training set compounds on the quality of the ESP surfaces produced by the DNN-fp model; plots showing the correlation between predicted ESP values and DFT ESP values for all feature types; plots showing the correlation between predicted and experimental values for four properties, comparing DFT, DNN-fp, and AM1-BCC (PDF)

AUTHOR INFORMATION

Corresponding Author

*E-mail: marcel.verdonk@astx.com.

ORCID ®

Marcel L. Verdonk: 0000-0002-6484-3328

Notes

The authors declare no competing financial interest.

Data and Models: The training and test data (SMILES, atomic properties and connectivity matrices, and fitted Δq values) used for model building are available at <https://doi.org/10.6084/m9.figshare.9768071.v1>. The trained model and associated code can be found at https://github.com/AstexUK/ESP_DNN/tree/master/esp_dnn.

ACKNOWLEDGMENTS

This work was funded by a postdoctoral fellowship awarded to P.C.R. under the “Sustaining Innovation Postdoctoral Training Program” at Astex Pharmaceuticals. We thank Andreas Bender and Richard Lewis of the Chemistry Department at Cambridge University for getting us started with model training and helpful discussions. We thank the members of the Computational Chemistry and Informatics Department for helpful discussions. In particular we thank Gianni Chessari for discussions around XIAP SAR, Paul Mortenson for assisting us by setting up Q-chem calculations, and Chris Murray and Richard Hall for useful feedback and for reading the manuscript.

ABBREVIATIONS USED

ESP, electrostatic potential; QM, quantum mechanics; DFT, density functional theory; MM, molecular mechanics; DNN, deep neural network; RMSE, root-mean-square error; HAC, heavy atom count; fXa, factor Xa; XIAP, X-linked inhibitor of apoptosis protein

REFERENCES

- (1) Scrocco, E.; Tomasi, J. The Electrostatic Molecular Potential as a Tool for the Interpretation of Molecular Properties. In *New Concepts II*; Springer: Berlin, 1973; Vol. 42, pp 95–170m DOI: [10.1007/3-540-06399-4_6](https://doi.org/10.1007/3-540-06399-4_6).
- (2) Petrolongo, C. Quantum Chemical Study of Isolated and Interacting Molecules with Biological Activity. *Gazz. Chim. Ital.* **1978**, *108*, 445–478.
- (3) Bonaccorsi, R.; Scrocco, E.; Tomasi, J. Molecular SCF Calculations for the Ground State of Some Three Membered Ring Molecules: $(CH_2)_3$, $(CH_2)_2 NH$, $(CH_2)_2 NH_2^+$, $(CH_2)_2 O$, $(CH_2)_2 S$, $(CH_2)_2 CH_2$, and $N_2 CH_2$. *J. Chem. Phys.* **1970**, *52* (10), 5270–5284.
- (4) Weiner, P. K.; Langridge, R.; Blaney, J. M.; Schaefer, R.; Kollman, P. A. Electrostatic Potential Molecular Surfaces. *Proc. Natl. Acad. Sci. U. S. A.* **1982**, *79* (12), 3754–3758.
- (5) 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.
- (6) Wilcken, R.; Zimmermann, M. O.; Lange, A.; Joerger, A. C.; Boeckler, F. M. Principles and Applications of Halogen Bonding in Medicinal Chemistry and Chemical Biology. *J. Med. Chem.* **2013**, *56* (4), 1363–1388.
- (7) Dougherty, D. A. The Cation- π Interaction. *Acc. Chem. Res.* **2013**, *46* (4), 885–893.
- (8) Meyer, E. A.; Castellano, R. K.; Diederich, F. Interactions with Aromatic Rings in Chemical and Biological Recognition. *Angew. Chem., Int. Ed.* **2003**, *42* (11), 1210–1250.
- (9) Bleiziffer, P.; Schaller, K.; Riniker, S. Machine Learning of Partial Charges Derived from High-Quality Quantum-Mechanical Calculations. *J. Chem. Inf. Model.* **2018**, *58* (3), 579–590.
- (10) Heidrich, J.; Exner, T. E.; Boeckler, F. M. Predicting the Magnitude of σ -Holes Using VmaxPred, a Fast and Efficient Tool Supporting the Application of Halogen Bonds in Drug Discovery. *J. Chem. Inf. Model.* **2019**, *59* (2), 636–643.
- (11) Jakalian, A.; Jack, D. B.; Bayly, C. I. Fast, Efficient Generation of High-Quality Atomic Charges. AM1-BCC Model: II. Parameterization and Validation. *J. Comput. Chem.* **2002**, *23* (16), 1623–1641.
- (12) Mardirossian, N.; Head-Gordon, M. Thirty Years of Density Functional Theory in Computational Chemistry: An Overview and Extensive Assessment of 200 Density Functionals. *Mol. Phys.* **2017**, *115* (19), 2315–2372.
- (13) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. Development and Use of Quantum Mechanical Molecular Models. 76. AM1: A New General Purpose Quantum Mechanical Molecular Model. *J. Am. Chem. Soc.* **1985**, *107* (13), 3902–3909.
- (14) Hunter, J. D. Matplotlib: A 2D Graphics Environment. *Comput. Sci. Eng.* **2007**, *9* (3), 90–95.
- (15) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic atom type and bond type perception in molecular mechanical calculations. *J. Mol. Graphics Modell.* **2006**, *25*, 247–260.
- (16) Laurence, C.; Brameld, K. A.; Graton, J.; Le Questel, J.-Y.; Renault, E. The PK(BHX) Database: Toward a Better Understanding of Hydrogen-Bond Basicity for Medicinal Chemists. *J. Med. Chem.* **2009**, *52* (14), 4073–4086.
- (17) Murray, J. S.; Politzer, P. Correlations between the Solvent Hydrogen-Bond-Donating Parameter, Alpha, and the Calculated Molecular Surface Electrostatic Potential. *J. Org. Chem.* **1991**, *56* (23), 6715–6717.
- (18) Kenny, P. W. Prediction of Hydrogen Bond Basicity from Computed Molecular Electrostatic Properties: Implications for Comparative Molecular Field Analysis. *J. Chem. Soc., Perkin Trans. 2* **1994**, *2* (2), 199–202.
- (19) Kenny, P. W.; Montanari, C. A.; Prokopczyk, I. M.; Ribeiro, J. F. R.; Sartori, G. R. Hydrogen Bond Basicity Prediction for Medicinal Chemistry Design. *J. Med. Chem.* **2016**, *59* (9), 4278–4288.
- (20) Ma, Y.; Gross, K.; Hollingsworth, C.; Seybold, P.; Murray, J. Relationships between Aqueous Acidities and Computed Surface-Electrostatic Potentials and Local Ionization Energies of Substituted Phenols and Benzoic Acids. *J. Mol. Model.* **2004**, *10* (4), 235–239.
- (21) Caballero-García, G.; Mondragón-Solórzano, G.; Torres-Cadena, R.; Díaz-García, M.; Sandoval-Lira, J.; Barroso-Flores, J. Calculation of VS_{Max} and Its Use as a Descriptor for the Theoretical Calculation of PK_a Values for Carboxylic Acids. *Molecules* **2019**, *24* (1), 79.
- (22) Virant, M.; Drvarič Talian, S.; Podlipnik, Č.; Hribar-Lee, B. Modelling the Correlation Between Molecular Electrostatic Potential and pKa on Sets of Carboxylic Acids, Phenols and Anilines. *Acta Chim. Slov.* **2017**, *560*–563.
- (23) Slater, A. M. The IUPAC Aqueous and Non-Aqueous Experimental PK_a Data Repositories of Organic Acids and Bases. *J. Comput.-Aided Mol. Des.* **2014**, *28* (10), 1031–1034.
- (24) Pinto, D. J. P.; Smallheer, J. M.; Cheney, D. L.; Knabb, R. M.; Wexler, R. R. Factor Xa Inhibitors: Next-Generation Antithrombotic Agents. *J. Med. Chem.* **2010**, *53* (17), 6243–6274.
- (25) Young, R. J.; Borthwick, A. D.; Brown, D.; Burns-Kurtis, C. L.; Campbell, M.; Chan, C.; Charbaut, M.; Chung, C.; Convery, M. A.; Kelly, H. A.; King, N. P.; Kleanthous, S.; Mason, A. M.; Pateman, A. J.; Patikis, A. N.; Pinto, I. L.; Pollard, D. R.; Senger, S.; Shah, G. P.; Toomey, J. R.; Watson, N. S.; Weston, H. E. Structure and Property Based Design of Factor Xa Inhibitors: Pyrrolidin-2-ones with Biaryl P4 Motifs. *Bioorg. Med. Chem. Lett.* **2008**, *18* (1), 23–27.
- (26) Hansch, C.; Leo, A.; Taft, R. W. A Survey of Hammett Substituent Constants and Resonance and Field Parameters. *Chem. Rev.* **1991**, *91* (2), 165–195.
- (27) Chessari, G.; Buck, I. M.; Day, J. E. H.; Day, P. J.; Iqbal, A.; Johnson, C. N.; Lewis, E. J.; Martins, V.; Miller, D.; Reader, M.; Rees, D. C.; Rich, S. J.; Tamanini, E.; Vitorino, M.; Ward, G. A.; Williams, P. A.; Williams, G.; Wilsher, N. E.; Woolford, J.-A. Fragment-Based Drug Discovery Targeting Inhibitor of Apoptosis Proteins: Discovery of a Non-Alanine Lead Series with Dual Activity Against CIAP1 and XIAP. *J. Med. Chem.* **2015**, *58* (16), 6574–6588.

- (28) Hammett, L. P. The Effect of Structure upon the Reactions of Organic Compounds. Benzene Derivatives. *J. Am. Chem. Soc.* **1937**, *59* (1), 96–103.
- (29) eMolecules database. <https://reaxys.emolecules.com/index.php> (accessed Jan 17, 2019).
- (30) Sigma-Aldrich. <https://www.sigmaaldrich.com/united-kingdom.html> (accessed Jun 3, 2019).
- (31) Swain, M. Molecule Validation and Standardization. <https://github.com/mcs07/MolVS> (accessed Jan 17, 2019).
- (32) RDKit: Open-source cheminformatics. <https://www.rdkit.org/> (accessed Jan 17, 2019).
- (33) Gaulton, A.; Bellis, L. J.; Bento, A. P.; Chambers, J.; Davies, M.; Hersey, A.; Light, Y.; McGlinchey, S.; Michalovich, D.; Al-Lazikani, B.; Overington, J. P. ChEMBL: A Large-Scale Bioactivity Database for Drug Discovery. *Nucleic Acids Res.* **2012**, *40* (D1), D1100–D1107.
- (34) Mirjalili, V. Data Clustering in Python. <https://libraries.io/github/mirjalil/pyclust> (accessed 2019).
- (35) Sayle, R. 1st-class SMARTS patterns. <https://www.daylight.com/meetings/summerschool98/course/basics/ref/sayle/> (accessed May 31, 2019).
- (36) Sadowski, J.; Schwab, C.; Gasteiger, J. CORINA, 3D Structure Generator, version 3.1; Molecular Networks GmbH: Erlangen, Germany, 2003.
- (37) Shao, Y.; Gan, Z.; Epifanovsky, E.; Gilbert, A. T. B.; Wormit, M.; Kussmann, J.; Lange, A. W.; Behn, A.; Deng, J.; Feng, X.; Ghosh, D.; Goldey, M.; Horn, P. R.; Jacobson, L. D.; Kaliman, I.; Khalilullin, R. Z.; Kuš, T.; Landau, A.; Liu, J.; Proynov, E. I.; Rhee, Y. M.; Richard, R. M.; Rohrdanz, M. A.; Steele, R. P.; Sundstrom, E. J.; Woodcock, H. L., III; Zimmerman, P. M.; Zuev, D.; Albrecht, B.; Alguire, E.; Austin, B.; Beran, G. J. O.; Bernard, Y. A.; Berquist, E.; Brandhorst, K.; Bravaya, K. B.; Brown, S. T.; Casanova, D.; Chang, C.-M.; Chen, Y.; Chien, S. H.; Closser, K. D.; Crittenden, D. L.; Diedenhofen, M.; DiStasio, R. A., Jr.; Do, H.; Dutoi, A. D.; Edgar, R. G.; Fatehi, S.; Fusti-Molnar, L.; Ghysels, A.; Golubeva-Zadorozhnaya, A.; Gomes, J.; Hanson-Heine, M. W. D.; Harbach, P. H. P.; Hauser, A. W.; Hohenstein, E. G.; Holden, Z. C.; Jagau, T.-C.; Ji, H.; Kaduk, B.; Khistyayev, K.; Kim, J.; Kim, J.; King, R. A.; Klunzinger, P.; Kosenkov, D.; Kowalczyk, T.; Krauter, C. M.; Lao, K. U.; Laurent, A. D.; Lawler, K. V.; Levchenko, S. V.; Lin, C. Y.; Liu, F.; Livshits, E.; Lochan, R. C.; Luenser, A.; Manohar, P.; Manzer, S. F.; Mao, S.-P.; Mardirossian, N.; Marenich, A. V.; Maurer, S. A.; Mayhall, N. J.; Neuscammman, E.; Oana, C. M.; Olivares-Amaya, R.; O'Neill, D. P.; Parkhill, J. A.; Perrine, T. M.; Peverati, R.; Prociuk, A.; Rehn, D. R.; Rosta, E.; Russ, N. J.; Sharada, S. M.; Sharma, S.; Small, D. W.; Sodt, A.; Stein, T.; Stück, D.; Su, Y.-C.; Thom, A. J. W.; Tsushima, T.; Vanovschi, V.; Vogt, L.; Vydrov, O.; Wang, T.; Watson, M. A.; Wenzel, J.; White, A.; Williams, C. F.; Yang, J.; Yeganeh, S.; Yost, S. R.; You, Z.-Q.; Zhang, I. Y.; Zhang, X.; Zhao, Y.; Brooks, B. R.; Chan, G. K. L.; Chipman, D. M.; Cramer, C. J.; Goddard, W. A., III; Gordon, M. S.; Hehre, W. J.; Klamt, A.; Schaefer, H. F., III; Schmidt, M. W.; Sherrill, C. D.; Truhlar, D. G.; Warshel, A.; Xu, X.; Aspuru-Guzik, A.; Baer, R.; Bell, A. T.; Besley, N. A.; Chai, J.-D.; Dreuw, A.; Dunietz, B. D.; Furlani, T. R.; Gwaltney, S. R.; Hsu, C.-P.; Jung, Y.; Kong, J.; Lambrecht, D. S.; Liang, W.; Ochsenfeld, C.; Rassolov, V. A.; Slipchenko, L. V.; Subotnik, J. E.; Van Voorhis, T.; Herbert, J. M.; Krylov, A. I.; Gill, P. M. W.; Head-Gordon, M. Advances in Molecular Quantum Chemistry Contained in the Q-Chem 4 Program Package. *Mol. Phys.* **2015**, *113* (2), 184–215.
- (38) Baker, N. APBS-PDB2PQR Documentation. <https://buildmedia.readthedocs.org/media/pdf/apbs-pdb2pqr/latest/apbs-pdb2pqr.pdf> (accessed 2018).
- (39) Connolly, M. L. Solvent-Accessible Surfaces of Proteins and Nucleic Acids. *Science* **1983**, *221* (4612), 709–713.
- (40) Rose, A. S.; Hildebrand, P. W. NGL Viewer: A Web Application for Molecular Visualization. *Nucleic Acids Res.* **2015**, *43* (W1), W576–W579.
- (41) Hartshorn, M. J. AstexViewer: A Visualisation Aid for Structure-Based Drug Design. *J. Comput.-Aided Mol. Des.* **2002**, *16* (12), 871–881.
- (42) Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. The Amber Biomolecular Simulation Programs. *J. Comput. Chem.* **2005**, *26* (16), 1668–1688.
- (43) Chollet, F.; et al.. *Keras*; GitHub, 2015.
- (44) Abadi, M.; Agarwal, A.; Barham, P.; Brevdo, E.; Chen, Z.; Citro, C.; Corrado, G. S.; Davis, A.; Dean, J.; Devin, M.; Ghemawat, S.; Goodfellow, I.; Harp, A.; Irving, G.; Isard, M.; Jia, Y.; Jozefowicz, R.; Kaiser, L.; Kudlur, M.; Levenberg, J.; Mané, D.; Monga, R.; Moore, S.; Murray, D.; Olah, C.; Schuster, M.; Shlens, J.; Steiner, B.; Sutskever, I.; Talwar, K.; Tucker, P.; Vanhoucke, V.; Vasudevan, V.; Viégas, F.; Vinyals, O.; Warden, P.; Wattenberg, M.; Wicke, M.; Yu, Y.; Zheng, X. TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems. *arXiv:1603.044*, **2015**.
- (45) Glorot, X.; Bordes, A.; Bengio, Y. Deep Sparse Rectifier Neural Networks. *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*; JMLR, 2011; pp 315–323.
- (46) Kingma, D. P.; Ba, J. Adam: A Method for Stochastic Optimization. *ArXiv14126980* **2014**.