



# HHS Public Access

Author manuscript

*Phys Chem Chem Phys.* Author manuscript; available in PMC 2021 April 06.

Published in final edited form as:

*Phys Chem Chem Phys.* 2020 April 06; 22(13): 6848–6860. doi:10.1039/d0cp00088d.

## Impact of Electronic Polarizability on Protein-Functional Group Interactions

Himanshu Goel<sup>a</sup>, Wenbo Yu<sup>a</sup>, Vincent D. Ustach<sup>a</sup>, Asaminew H. Aytenfisu<sup>a</sup>, Delin Sun<sup>a</sup>, Alexander D. MacKerell Jr.<sup>\*,a</sup>

<sup>a</sup>Computer Aided Drug Design Center, Department of Pharmaceutical Sciences, University of Maryland School of Pharmacy, 20, Penn St. Baltimore, Maryland 21201, United States.

### Abstract

Interactions of proteins with functional groups are key to their biological functions, making it essential that they be accurately modeled. To investigate the impact of the inclusion of explicit treatment of electronic polarizability in force fields on protein-functional group interactions, the additive CHARMM and Drude polarizable force field are compared in the context of the Site-Identification by Ligand Competitive Saturation (SILCS) simulation methodology from which functional group interactions patterns with five proteins for which experimental binding affinities of multiple ligands are available, were obtained. The explicit treatment of polarizability produces significant differences in the functional group interactions in the ligand binding sites including overall enhanced binding of functional groups to the proteins. This is associated with variations of the dipole moments of solutes representative of functional groups in the binding sites relative to aqueous solution with higher dipole moments systematically occurring in the latter, though exceptions occur with positively charged methylammonium. Such variation indicates the complex, heterogeneous nature of the electronic environments of ligand binding sites and emphasizes the inherent limitation of fixed charged, additive force fields for modeling ligand-protein interactions. These effects yield more defined orientation of the functional groups in the binding pockets and a small, but systematic improvement in the ability of the SILCS method to predict the binding orientation and relative affinities of ligands to their target proteins. Overall, these results indicate that the physical model associated with the explicit treatment of polarizability along with the presence of lone pairs in a force field leads to changes in the nature of the interactions of functional groups with proteins versus that occurring with additive force fields, suggesting the utility of polarizable force fields in obtaining a more realistic understanding of protein-ligand interactions.

---

\* alex@outerbanks.umaryland.edu, Tel: 410-706-7442, Fax: 410-706-5017.

Electronic Supplementary Information (ESI) available: [details of any supplementary information available should be included here].  
See DOI: 10.1039/x0xx00000x

Conflicts of interest

ADM Jr. is co-founder and Chief Scientific Officer of SilcsBio LLC.

## Introduction

Computer simulations have become an important tool to model physical phenomena on the molecular scale in the biological, chemical, and material sciences. Such simulations have benefited from advancements in computational resources, methods and algorithms. Ideally, quantum chemistry methods can be employed for molecular simulations. However, these methods are limited to relatively small system sizes and moderately sized molecules.<sup>1-3</sup> Accordingly, for most condensed phase applications, the molecular simulations are performed using molecular mechanical (MM) techniques allowing access to larger system sizes and longer time scales.<sup>4</sup> These techniques, particularly Monte Carlo (MC) or molecular dynamics (MD) simulations, are based on intra/inter molecular potential energies or forces of the given coordinates obtained from an additive or polarizable force field.<sup>5-7</sup>

Force fields use an analytical potential energy function to study the behaviour of any given system and predict a range of properties. These functions are parametric in nature. The accuracy of the results obtained from the MM methods relies on the sampling of the conformations, which may be facilitated using enhanced sampling methods,<sup>8</sup> as driven by the accuracy of the force field. Currently, there are a number of highly specialized molecular force fields based on an additive, fixed or non-polarizable treatment of the partial atomic charges, such as CHARMM,<sup>9, 10</sup> AMBER,<sup>11-13</sup> GROMOS,<sup>14, 15</sup> OPLS,<sup>16, 17</sup> and others which are well established and extensively used.<sup>18</sup> The greatest advantage with additive force fields is their ability to handle large system sizes even for millions of atoms. However, keeping the partial atomic charges on the atomic nucleus fixed represents a significant approximation as in reality the charge distribution in a molecule varies in response to their environment. To overcome this, polarizable force fields for molecular simulations have been developed and are becoming more widely used with examples including ion-protein binding, ion permeations, ionic liquids, and hydrogen bond dynamics.<sup>19-25</sup> Examples of polarizable force fields under various levels of development include the CHARMM Drude polarizable force field,<sup>26</sup> AMBER ff02,<sup>27</sup> AMOEBA,<sup>28-30</sup> and CHEQ.<sup>31, 32</sup> While polarizable force fields offer multiple advantages in modeling environmental responses, using these increases the computational time as compared to simulations using the additive force fields.<sup>33-35</sup>

The Drude polarizable force field is based on the classical Drude oscillator model<sup>36</sup> and has been undergoing active development.<sup>26, 33, 37-44</sup> The Drude oscillators (or particles) are connected to the polarizable non-hydrogen atom nuclei via a harmonic spring. In the Drude force field, the hydrogen atoms are considered to be non-polarizable. The Drude particle carries some negative charge with a small mass of 0.4 amu. Most importantly, the Drude particle can move around the non-hydrogen atom in response to the electric field, making the way for the induced dipole moment. Moreover, the model also includes lone pairs modelled as virtual particles on the hydrogen bond acceptor atoms for improving hydrogen bonding in polar compounds.<sup>45</sup> While the Drude force field is still under development, it has been used to predict thermodynamic, transport and structural properties in selected chemical and biomolecular systems, though the true potential of the force field is largely unexplored.<sup>46-53</sup>

In this study, our aim is to demonstrate the impact of electronic polarizability on the interactions of functional groups with proteins in the context of the Site Identification by

Ligand Competitive Saturation (SILCS)<sup>54, 55</sup> framework. The SILCS methodology is based on a precomputed ensemble of water and small solutes representing a range of different types of functional groups around a target macromolecule typically a protein. This ensemble is then converted to free energy maps for the different functional groups and water, termed FragMaps, that may be used to discover and/or design new molecules as well as for protein-protein interactions.<sup>56–60</sup> Similar to SILCS, other co-solvents based simulation methods are increasingly applied in the context of a variety of applications.<sup>61–70</sup> To directly understand the impact of polarizability, the SILCS MD-based simulations were performed with both the CHARMM36 additive and Drude polarizable force fields with the FragMaps extracted for the two models for five proteins. The FragMaps are then analyzed to determine the similarities of the maps based on overlap coefficients, prediction of the relative affinities of ligand binding, dipole moment distribution plots of solutes used in the simulations, and vector maps for those solutes to monitor their orientation in selected binding sites. The SILCS methodology has previously been used on a diverse set of proteins using a combined Grand canonical Monte Carlo (GCMC)-MD sampling approach with the additive CHARMM36 and CHARMM General Force Field.<sup>58, 71, 72</sup> However, as the GCMC methodology has yet to be extended to the Drude polarizable force field the goal of the present study is achieved through the use of MD sampling alone in the SILCS simulations.

The next section explains the computational methodology for the SILCS system setup, SILCS MD simulation, and SILCS-MC ligand sampling details along with other terms. In the results and discussion section, we discuss and compare various outcomes obtained from additive and Drude force field-based simulations, and the last section presents concluding remarks.

## Computational Methodology

### A SILCS System Setup

Five diverse proteins for which experimental data in relation to the ligand binding affinities are available were selected for the present study. These include (1) Farnesoid X acceptor (Factor Xa, 1FJS),<sup>73</sup> (2) p38 $\alpha$  Mitogen-Activated protein kinase (P38, 3FLN),<sup>74</sup> (3) tyrosine kinase 2 (TYK2, 4GIH),<sup>75</sup> (4) tRNA m1G37 methyltransferase enzyme TrmD (TRMD, 4YPW),<sup>76</sup> and (5) mouse double minute 2 homolog (HDM2, 4JV7).<sup>77</sup> For each protein, the ligands were accessed from our previous works yielding a total of 175 ligands. Additional details regarding ligand preparation can be found in Ustach et al.<sup>58</sup> (for P38, HDM2, TYK2, and TRMD) and Raman et al.(Factor Xa).<sup>78</sup> The crystal water, ligands, and cofactors were removed from each protein. The simulation system for the SILCS methodology involves a protein, multiple solutes, and water. In order to have efficient sampling of the system, 10 independent simulations are performed on each protein. The initial structure of the proteins was altered by setting the  $\chi^1$  dihedral in ten individual orientations in 36 $^\circ$  increments for solvent accessible residues based on a solvent accessibility of 0.005 nm<sup>2</sup> determined based on solvent accessible surface area (SASA)<sup>79</sup> from GRONINGEN MACHINE FOR CHEMICAL SIMULATION (GROMACS).<sup>80</sup> Each of those 10 structures were then solvated by water at 55 M along with the eight solutes, each at a concentration of 0.25 M, keeping a minimum distance of 10 Å from the edge of the simulation box. The solutes include benzene, propane,

methanol, formamide, acetaldehyde, methylammonium, imidazole and acetate, as previously performed.<sup>58</sup> An in-house script comprising several GROMACS utilities is used to perform the system setup with different initial positions and orientations of protein, solutes and water. Each system is energy minimized using the steepest descent algorithm<sup>81</sup> for 5000 steps followed by equilibration and production. The additive system set up for each protein were converted to the Drude polarizable model using the CHARMM<sup>82</sup> program. During this conversion, the Drude particles were added to all the non-hydrogen atoms of the system along with the lone pairs to hydrogen bond acceptors.

## B SILCS MD Details

Protein force field parameters were from the additive CHARMM36<sup>83</sup> model or Drude-2013 polarizable force field<sup>37</sup>. Additive force field parameters for the solutes and ligands were derived from the CHARMM general force field (CGenFF)<sup>84</sup> program. Drude parameters for the solutes were from the original publications<sup>26, 85, 86</sup> and modified parameters for selected solutes are included in the final section of the Supporting Information. For water, the additive simulation uses CHARMM TIP3P water model,<sup>87, 88</sup> which was converted to the polarizable SWM4-NDP<sup>89</sup> model for Drude simulations. For all the additive simulations, GROMACS software version 5.1.0 was used for the energy minimization, equilibration and production runs. The additive simulation employed the Nosé-Hoover thermostat<sup>90, 91</sup> and Parrinello-Rahman barostat<sup>92</sup> for maintaining the system temperature and pressure at 298 K and 1 atm, respectively, and the LINCS algorithm<sup>93</sup> was used to constrain all bonds with hydrogen atoms including water. The backbone C $\alpha$  atoms were restrained with a force constant of 0.12 kcal mol<sup>-1</sup>Å<sup>-2</sup> for the production simulations. The nonbonded interaction was computed with a cutoff of 8.0 Å by using the Verlet cutoff scheme. The Lennard-Jones (LJ) interactions were switched off smoothly in the range of 5.0 to 8.0 Å along with the long-range isotropic dispersion correction to the energy and pressure for LJ interaction beyond 8.0 Å. The particle mesh Ewald (PME)<sup>94</sup> method is used to handle the long-range electrostatics with a real space cutoff of 8.0 Å, maximum grid spacing 1.2 Å, and fourth order B-spline interpolation. The time constant for temperature and pressure coupling was set to 1 ps. The equilibration was carried out for 1 ns followed by the production run for 200 ns for each of the 10 systems for each protein with a timestep of 2 fs.

The Drude simulations were performed with the OpenMM<sup>95</sup> toolkit which was recently customized to handle Drude polarizable force field based simulations.<sup>33</sup> Simulation details were identical to those used for the additive force field with the following differences. The Drude particles were optimized using steepest descent<sup>81</sup> (1000 steps) and adopted-basis Newton-Raphson (2000 steps) energy minimization methods followed by MD equilibration in the *NPT* ensemble for 10,000 steps with a timestep of 1 fs at the temperature and pressure of 298 K and 1 bar, respectively. A hard wall restraint of 0.2 Å between the parent atom and the Drude was applied to prevent instability and large displacements of Drude particles.<sup>96</sup> The hard wall is designed to avoid polarization catastrophe that may occur due to low frequency close interactions between atoms that may occur during the MD simulation leading to over polarization. In the present study, the hard wall was only encountered in the C and O atoms of the FORM solute in 5.7E-5 % of the time steps. The Drude polarizable simulations used the extended Lagrangian dynamics scheme<sup>97</sup> where the real atoms and the

Drude particles are coupled to two thermostat responsible for uniting their dynamics. The “physical” and Drude thermostats were maintained at temperatures of 298 K and 1 K with the friction coefficients of 5 ps<sup>-1</sup> and 20 ps<sup>-1</sup>, respectively. Periodic boundary conditions (PBC)<sup>98</sup> were used in all the simulations with the nonbonded interactions and Ca atom restraints treated as described above for the additive model. The pressure was regulated using a Monte Carlo barostat in OpenMM. The production simulations were carried out for 200 ns for each of the 10 systems for each protein with a timestep of 1 fs. The total sampling time for additive and Drude simulation is 2 μs. Simulation snapshots were saved every 10 ps during the production runs.

### C FragMaps, GFE, and Overlap Coefficients

FragMaps were generated from the simulation trajectories by binning selected solute and water non-hydrogen atoms into  $1 \times 1 \times 1 \text{ \AA}^3$  cubic volume elements (voxels) and the occupancy of the selected atoms determined. Selected atoms of the solutes are categorized into the FragMaps as described below. The probability distributions are normalized for the concentration of the solutes in solution and converted into the grid free energies (GFE) via Boltzmann transformation<sup>99</sup> yielding FragMaps defined based on free energies of the selected solute or water atoms relative to being in bulk solution. GFE is an atom-based metric which allocates a free energy score to selected ligand atoms in correspondence to the overlap of those atoms with the FragMaps. In the present study selected solute atom types were combined into generic FragMaps as follows by merging voxel occupancies of the different solutes and converting the normalized distributions into GFE.<sup>58</sup> They are generic nonpolar or apolar (GENN, benzene and propane), generic acceptor (GENA, formamide O, acetaldehyde O, imidazole acceptor N, and methanol O), generic donor (GEND, formamide N and imidazole donor N, and methanol O). In addition, specific maps for methylammonium nitrogen (DPOS, MAMN), acetate oxygen (ANEG, ACEO) and methanol (MEOO) were used, with the latter used for calculating the GFE scores of alcohols. All visualization of the FragMaps was presented using visual molecular dynamics (VMD)<sup>100</sup> program.

Analysis of the similarity of the FragMaps was performed based on the overlap coefficient. The overlap coefficient (OC) provides a quantitative evaluation of the similarity of two sets of FragMaps. OCs values greater than 0.8 indicate a high degree of similarity and values of 0.6–0.8 depicts reasonable level of similarity<sup>78</sup> with lower values indicating significant differences. The overlap coefficient (OC)<sup>99</sup> is defined as

$$\text{Overlap coefficient} = \sum_{i=1}^N \min \left( \frac{Q_i^1}{\sum_{j=1}^N Q_j^1}; \frac{Q_i^2}{\sum_{j=1}^N Q_j^2} \right)$$

where  $N$  is the number of voxels in the FragMaps and  $Q_j^1$  and  $Q_j^2$  are their occupancies of voxel  $i$ . The OCs were calculated individually for each FragMap and vary between 0 to 1, with 1 signifying equivalent FragMaps.

## D LGFE and SILCS-MC Sampling

SILCS-MC sampling method is utilized to predict the binding conformation of the ligands in the presence of FragMaps.<sup>78</sup> The SILCS-MC is a Monte Carlo sampling method where ligand moves are accepted/rejected based on the Metropolis<sup>101</sup> criteria. The Metropolis criteria is based on the ligand grid free energy (LGFE) scores combined with intramolecular conformational energies associated with the CGenFF energy function. LGFE scores are obtained by classifying selected ligand atoms into different FragMaps types, as mentioned above, based on their chemical similarity to the different type of FragMaps. Based on the overlap of the ligand atoms with the FragMaps each atom is assigned a GFE score, with the LGFE score being the summation of the atomic GFE scores. Details of current implementation of the GFE and LGFE scoring is detailed in Ustach et al.<sup>58</sup> The LGFE scores are representative of binding free energies; however, as they are a sum of the GFE scores of the classified atoms in each ligand and, therefore, are missing contributions such as the configurational entropy associated with connecting the atoms to make the full ligand, they are not formal free energies of binding. For the SILCS-MC the temperature was initially set to 300 K for the normal MC run and gradually lowered it to 0 K. The CGenFF contribution to the Metropolis criteria, in addition to the LGFE scores, includes van der Waals (vdW), electrostatic and dihedral terms. All the force field parameters for the ligands were derived from the CGenFF program.<sup>84</sup> The SILCS-MC method does not contain any protein and solvent environment explicitly, therefore, a distance dependent dielectric ( $=4|r|$ ) was employed to compute intramolecular electrostatics.<sup>102–104</sup> Before starting the SILCS-MC run, each ligand structure was optimized in the context of CGenFF by employing energy minimization for 10,000 steps. During the SILCS-MC sampling involved translations, rotations and rotations of dihedrals about rotatable bonds. The CGenFF intramolecular energy used in the SILCS-MC Metropolis criteria is the additive form for both the additive and Drude FragMap calculations. Moreover, we performed five independent runs of SILCS-MC cycles for each ligand to thoroughly identify a minimum free energy binding conformation.

Two types of SILCS-MC protocols were used, namely Local and Exhaustive. The protocols differ in the number of MC cycles, types of attempted moves and initial placement of the ligand. The starting conformation and orientation for the Local protocol involves a user defined position of the ligand, typically based on the crystallographic location, whereas in the Exhaustive protocol the ligand is positioned randomly within a radius defined by the user. The Exhaustive protocol explores a larger region of conformational space as compared to the Local due to initial placement of the ligand in a random conformation and orientation in the binding pocket. For the Exhaustive protocol, a 10 Å radius was used. Each ligand was then subjected to 100 MC steps and 1000 annealing steps for the Local protocol and 10,000 MC steps and 40,000 annealing steps for the Exhaustive protocol with the final ligand orientation being selected based on the most favorable LGFE score. Final scoring of each ligand is based solely on the LGFE score. Additional details about the SILCS MC protocols can be accessed from the recent work by Ustach et al.<sup>58</sup> We note that while the SILCS simulations are performed using the Drude polarizable force field and, therefore, are computationally more demanding than the additive SILCS simulations by ~4 fold, once the SILCS FragMaps are available the ligand SILCS-MC docking calculations have the identical

computational requirements for the two models as the SILCS-MC sampling is performed directly in the field of the FragMaps.

## Results and discussion

SILCS FragMaps are 3D probability distributions of selected solute atoms that define the functional group affinity pattern of the target protein. The FragMaps quantify regions of both favorable and unfavorable affinities of the functional groups relative to being in aqueous solution. The utility of the SILCS methodology is largely based on the accuracy of the FragMaps with respect to the true functional group distributions that would be observed experimentally (if such experiments were available). Accordingly, it may be assumed that improvements in the underlying potential energy function may lead to more accurate distributions of solutes and water around the target protein and, therefore, improved FragMaps. To test this, in the present study SILCS simulations are applied to 5 proteins using both the CHARMM36 additive and Drude-2013 polarizable force fields. As the Drude force field explicitly treats electronic polarization as well as lone pairs in the case of hydrogen bond acceptors, it is anticipated that the force field will more accurately model the distribution of solute functional groups around the protein. In addition, the proposed calculations allow for details of variations in dipole moments and the orientation of functional groups that occur in ligand binding sites associated with the explicit treatment of polarizability and lone pairs to be investigated.

Analysis of the overlap coefficient (OC) values of the FragMaps from simulations 1–5 and 6–10 from each of the systems for all 5 proteins was performed to determine the extent of convergence of the SILCS MD simulations (Table S1 of the supporting information). The OC values are all 0.8 or greater, indicating high convergence. With respect to treatment of the ligands, the SILCS method offers the advantage that once the FragMaps are available the SILCS-MC ligand docking is primarily based on the FragMaps themselves rather than the underlying potential energy function. Thus, for SILCS-MC of the ligands the combined CGenFF/MolCal<sup>78</sup> engine used to convert mol2 files into the SILCS atom classification used for the additive-based FragMaps may also be used for the Drude-based FragMaps. Essentially, once the polarization contributions are embedded in the FragMaps, the same tools for ligand sampling, and the same computational demand, may be used for the Drude SILCS-MC calculations.

### A FragMap Analyses

Initial comparison of the FragMaps from the additive and Drude force fields involved OC analysis. Results, shown in Table S2 of the Supporting Information, shows the values of OC for different solutes for all the protein targets for the entire simulation box. The majority of values are > 0.8 with the lowest value being 0.71 indicating the overall maps to be similar. The overall high OC values between the additive and Drude FragMaps are expected given that majority of 3D space defining the FragMaps is in the solution around the proteins. To determine if the FragMaps are similar in the regions important for ligand binding the OC values in the binding pockets were evaluated.

Table 1 represents the OC values for regions centered on the binding pocket of size  $20 \times 20 \times 20 \text{ \AA}^3$  for P38 and Factor Xa and  $16 \times 16 \times 16 \text{ \AA}^3$  for the other proteins; the sizes were selected based on the spatial extent of the known ligands. Interestingly, there are major differences observed between the additive and Drude models for the majority of FragMaps and proteins. Some of the protein show OC values close to 0.3 in particular for charged solutes. The OC values for HDM2 show a reasonable level of similarity for all FragMaps and the agreement for the water tipo maps is also high in all cases. Thus, the OC values in the majority of binding pockets indicates that there are significant differences present in the FragMap distributions between the two force fields. Additional analysis focused on the details of these differences.

Qualitative evaluation of the FragMap distributions obtained from the Drude and additive simulations in the bindings sites was based on the data shown in Figures 1 and S2. Included are ligands from the crystal structures used to initiate the SILCS simulations, allowing for determination of the ability of the FragMaps to qualitatively recapitulate the location of different functional groups of the ligands. Discussion of results for Factor Xa, P38 and TRMD are below with those for HDM2 and TYK2 in the Supporting Information. The present discussion is based on visualization of the maps at a GFE contour level of  $-1.2$  kcal/mol unless noted.

Factor Xa: The FragMaps in the active site are shown in Figure 1A and 1B for additive and Drude simulations, respectively. Beginning with the apolar FragMaps, the Drude FragMaps clearly replicate both aromatic phenyl rings of the ligand as shown by the arrow 1 in the S1 pocket and arrow 2 in the S4 pocket. In contrast, the additive apolar FragMaps entirely miss the aromatic ring in the S1 pocket (Figure 1A). The additive FragMaps indicate the presence of the benzene ring in pocket S4 (arrow 3); however, their extent is lower than with the Drude FragMap. Hydrogen bond acceptor or donor (GENA/GEND) FragMaps are evident in the S1 and S4 pockets, with the extent of those maps larger with the Drude FragMaps. With the S4 pocket, the ligand's imidazole group is captured by the Drude GENA FragMaps while these are not present with the additive model at the contour level of  $-1.2$  kcal/mol. The charged amidine group in the S1 pocket is correctly mapped by positive MAMN FragMaps for both force fields. In addition, the Drude MAMN FragMaps are present in the lower region of the S4 pocket which is consistent with the presence of an imidazole or ammonium group on ligands in the 1EZQ, 1MQ5, and 1Z6E crystal structures.<sup>78</sup> The additive simulation also shows certain FragMaps for negative (ACEO) and apolar groups in the upper left corner of Figure 1A. The apolar maps (green) are overlaid on the crystallographic structure whereas the negative maps (orange) are adjacent to the apolar maps. The ligand does contain a carboxylate group in that region, consistent with the presence of the ACEO FragMaps. With the Drude force field ACEO FragMaps are present in that region when visualized at a contour of  $-0.4$  kcal/mol.

P38 MAP Kinase: The P38 FragMaps are shown in Figure 1C and 1D for additive and Drude simulations, respectively. The apolar FragMaps for both models are similar as shown by arrows 1 and 2 in Figure 1C recapitulating the locations of the aromatic rings. However, the Drude apolar FragMaps form a continuous distribution across the binding site, better representing the central bicyclic heterocycle of the ligand. The positive (MAMN) FragMaps



are similar for the Drude and additive models as shown by arrow 3 in Figure 1C. Finally, significant differences are observed in the hydrogen bond donor and acceptor (GENA and GEND) FragMaps. Arrow 4 in Figure 1D denotes the hydrogen bond acceptor FragMaps from the Drude model which overlap with the fluorine atoms of ligand; these FragMaps are not evident in the additive results. Additional GENA/GEND FragMaps, shown by arrow 5 in Figure 1C, are present for both force fields. However, the additive FragMaps are quite limited as compared to the Drude results, where the FragMaps recapitulate more of the donor and acceptor atoms in the ligands as indicated by arrow 6 in the Figure 1D.

TYK2: The TYK2 FragMaps are shown in Figure 1E and 1F for additive and Drude models, respectively. The apolar FragMaps for both models capture the central region of the ligand overlapping with the aromatic ring of the ligand as shown by the arrow 1 in Figure 1F. These apolar maps extend beyond the pyridine ring, indicating that larger ring systems or hydrophobic functional groups could be accommodated in this region, consistent with the presence of an additional pyridine ring in this region in known inhibitors.<sup>75</sup> Both the force fields largely fail to have apolar maps overlapping with the dichloro aromatic ring of the ligand (left side of Figures 1E and 1F). A small region of Drude apolar FragMaps is present at the lower portion of that region, which correspond to one of the chlorine atoms on the ring. However, significant hydrogen-bond donor and acceptor maps are present in the vicinity of one of the chlorine atoms (arrow 2) and a small acceptor region is present adjacent to the second chlorine with the Drude FragMaps. The presence of these maps may indicate that the chlorines are interacting favorably with the protein via halogen bond or halogen-hydrogen bond donor interactions.<sup>43, 105</sup> Better modelling of these types of interactions may be performed using SILCS simulations that include halogenated solutes, as recently performed.<sup>58</sup> Additional hydrogen bond donor or acceptor FragMaps capture other heteroatoms on the ligand (arrow 3). In addition, the Drude FragMaps show the possibility of having polar groups in the vicinity of the isopropyl group (arrow 4 on Figure 1F). The acceptor FragMap is adjacent to the oxygen in the peptide bond and both the donor and acceptor maps indicate the potential for adding polarity to the isopropyl ring.

## B SILCS-MC Ligand Binding Predictions

Quantitative predictions centered upon the ability of both force fields to rank the order of ligand affinities based on LGFE scores to further evaluate the impact of the inclusion of polarizability on the SILCS method. The LGFE scores approximate experimental binding affinities though they are not true binding free energies, as discussed above. Two SILCS-MC sampling protocols, Local and Exhaustive, were applied for both the Drude and additive FragMaps of the five proteins. Local sampling is based on a known initial ligand conformation while Exhaustive is based on randomization of the ligand orientation in the binding site followed by more extensive MC conformational sampling. In the present study, comparisons were initially performed based on the predictive index (PI), root mean square error (RMSE), and the percent correct (PC).<sup>58</sup> The PI correlation for ligand scoring ranges between 1 for 100% true relative predictions and -1 for 100% false predictions, and 0 for the random predictions.<sup>106</sup> It is designed to be a metric of the ability of the method to rank order the ligands based on their affinities. The weighting term in the PI increases with the increment in the magnitude of the differences in affinities. The sum of true positive and true

negative comparisons is referred to as the PC for each series of ligands. The PC values should be maximal for lead optimization purposes and is simply indicating the percentage of occurrences when correct predictions in the direction of the binding affinity are made. The reported PC values are average values over individual PC values with each ligand in the series taken as the reference ligand. RMSE is based on the root-mean-square difference between the LGFE and the experimental binding free energies over all the ligands of each protein.

The PI, RMSE and PC values obtained from SILCS-MC simulation for both sets of FragMaps with Exhaustive and Local protocols are listed in Table 2. For the Drude FragMaps, the averaged values of PI, RMSE and PC for Exhaustive protocol are  $0.44 \pm 0.09$ ,  $2.18 \pm 0.28$  and  $0.65 \pm 0.03$ , respectively, where the errors are the standard errors over the 5 proteins. In comparison, the Exhaustive protocol for additive FragMaps results in  $0.36 \pm 0.10$ ,  $2.36 \pm 0.53$ , and  $0.61 \pm 0.03$  for PI, RMSE and PC, respectively. For the Local protocol the Drude FragMap averaged values for PI, RMSE and PC are  $0.46 \pm 0.13$ ,  $3.50 \pm 0.51$  and  $0.64 \pm 0.04$ , respectively, while for the additive FragMaps the results are  $0.47 \pm 0.05$ ,  $4.33 \pm 1.21$ , and  $0.64 \pm 0.02$  for PI, RMSE and PC, respectively. Thus, the Exhaustive and Local protocol results show overall improvements with the Drude FragMaps versus the additive, although the PI and PC values are close to each other with the Local sampling. However, as is evident from Table 2 in certain cases the additive FragMaps (e.g. TRMD with both the Exhaustive and Local protocols) yield better agreement with experiment, consistent with the system specific nature of CADD methods.

The final orientations of the ligands obtained from the SILCS-MC are an additional yardstick to quantify the Drude or additive force fields. For this purpose, we calculated the root mean squared distance (RMSD) and center of mass difference (COMD) of the final orientations of the ligands relative to the initial crystal structure-based orientations. The RMSD and COMD values for each protein with both protocols are listed in Table 2. For the Exhaustive SILCS-MC docking with the Drude set, the averaged values of RMSD and COMD are  $5.89 \pm 0.19$  and  $3.05 \pm 0.46$  Å, respectively while the additive values are  $6.27 \pm 1.29$  and  $2.95 \pm 1.14$  Å, respectively. The lower value of RMSD for the Drude set shows a small improvement in the prediction capability of the ligand orientation when compared to the additive values while the averaged COMD values are similar. Thus, the Drude model leads to a small improvement in the orientation of the ligands in the binding sites though their overall locations are similar with the two model. With the Local protocol, the RMSD and COMD averaged values for Drude are  $1.84 \pm 0.20$  and  $1.20 \pm 0.15$  Å, respectively while the additive values are  $1.88 \pm 0.24$  and  $1.09 \pm 0.12$  Å, respectively, indicating similar behaviour for the two FFs.

Shown in Figure 2 are the minimum LGFE conformations for Factor Xa, P38, and TYK2 based on the Exhaustive protocol. The analogous results for HDM2 and TRMD are shown in Figure S3. As is evident in the large majority of cases, despite the radius of 10 Å of the sphere used for initial ligand placement, the docked ligands are within the binding pockets. However, in all the systems variations in the orientations among all the ligands are observed to varying degrees and, with HDM2 and TRMD (Figure S2), a couple of the ligands occupy regions outside the binding pocket. In such cases use of a smaller radius may be applied in

the Exhaustive protocol to avoid departure of the ligand from the binding pocket. This was not performed in the present study as we simply applied the two default SILCS-MC approaches as previously presented.<sup>58</sup>

The final metric subjected to quantitative analysis was the magnitude of the LGFE scores. The LGFE scores do not directly correspond to the binding energy due to exclusion of energy terms associated with the covalent connectivity of the ligands such as the configurational entropy as well as additional terms. However, the LGFE should approximate the experimental binding affinity as they are based on the contributions of all individual functional group GFE that include contributions from desolvation, loss of rotational and translational degrees of freedom as well as functional group-protein interactions. The averaged differences between the binding affinity and the LGFE values are provided in the last column of the Table 2. For Exhaustive SILCS-MC the Drude and additive average differences are  $1.53\pm 0.53$  and  $1.88\pm 0.62$  kcal/mol, respectively, while with the Local sampling the values are  $3.20\pm 0.52$  and  $4.07\pm 1.25$  kcal/mol. In addition, the Drude and additive LGFE values obtained from Exhaustive protocol are plotted with respect to the experimental binding energy for all the ligands of all the protein as shown in Figure 3. All the data points related to Drude LGFE are closer to the linear line as compare to additive data sets except a few points. The smaller difference with the Drude model indicates that the free energies of the interactions of the solutes with the protein are more accurately modelled by the polarizable force field over the additive model. Notable, the Drude values are systematically less unfavorable than the additive values with respect to the experimental affinities, indicating that the polarizable force field is leading to overall enhanced binding of the solutes to the protein. Such improvements are associated with the electronic response of the solutes upon going from the aqueous solution to protein bound environments.

### C Dipole Moment Distributions

The binding of a particular solute to the protein corresponds with a change in its electrostatic environment from aqueous solution to the variety of environments in and around the protein. Such a change leads to alteration of the electron distribution of the solute, an effect that is accessible to the Drude polarizable force field. In order to quantify the polarization effects occurring upon the interaction of the solutes with the protein, we computed the dipole moment distributions of all the solutes from the SILCS simulations. The dipole moments were computed in the ligand binding pockets for all solutes within 3 Å of the selected residue and in the bulk phase as defined as 6 Å away from the protein surface based on the non-hydrogen atoms. The dipole moments for each solute were computed by orienting the solute to the same axis and center in the bulk and binding pocket for all different proteins, respectively, with the x, y and z vectors aligned on each solute shown in Figure S4. Figure 4 shows the dipole moment distribution for all the solutes in the binding pocket and the bulk phase with the pockets defined by glycine 106 of P38 Kinase, serine 184 of Factor Xa, and lysine 40 of TYK2 from the Drude SILCS simulations. For comparison purposes, Figure S5 shows the dipole moment distributions for all the solutes from the Drude and additive simulations. Moreover, the x, y, and z components of the dipole moment distribution for all solutes from the Drude simulations are presented in Figure S6, S7 and S8, respectively.

For most of the solutes, the dipole distributions obtained from the Drude simulations show noticeable differences between the bulk and the binding pocket as shown in Figure 4. For the solutes the extent of variations between the bulk and protein environments differ. Minimal changes occur with propane with a similar effect seen in MAMY for bulk and two of the proteins while a significant difference is present in the binding site of TYK2, where an increase in the dipole moments occur. With the other solutes there is generally a decrease in the dipole moment distributions in the binding sites of the proteins. The difference, which often ranges from 0.1 to 0.5 D, shows the extent by which the aqueous and protein environment perturb the electron distribution of the solutes. These results differ significantly from those with the additive force field, as expected, where no evident differences in the distributions are observed in the different environments and those distributions are more narrow than with the Drude force field (Figure S5).

Table 3 presents the average and standard errors for the dipole moments of all solute in the binding pockets versus in bulk solution for Drude model of P38, Factor Xa and TYK2 protein. Consistent with the distributions, there is a systematic decrease in the dipole moments in the binding sites as compared to bulk. The only exception occurs with MAMY where the differences are negligible or larger in the binding pocket with TYK2. Similar effects have been observed in studies of base flipping from DNA<sup>107</sup> and in proteins<sup>37, 39</sup> using the Drude force field where increases in bases or amino acid side chains, respectively, upon increased exposure to aqueous solution have been shown to occur. Such overall trends are interesting as they further emphasize the limitation of the approximation associated with the use of fixed charges in additive force fields and how it will tend to underestimate the equilibrium of the solutes in solution versus in the bound state. In addition, given the variability in the percent differences, even between individual solutes in the different proteins, it's evident that selection of a common scaling factor for the charges to account for the condensed phase environment is not possible.

More detailed analyses of the impact of interactions with proteins on the dipoles of the solutes are presented in Figures S6, S7 and S8 showing the x, y and z component distributions of the dipole moments. Results are shown for both the additive and Drude simulations. The trends are similar to that with the total dipole moments with respect to the degree of variation of the solute types in the bulk versus protein environments. However, interesting specific variations are observed. With FORM there is a peak in the all three distributions in the binding site of Tyk2 that leads to an overall downshift in the total dipole moment distribution (Figure 4). Large peaks also occur with MAMY in Factor Xa and, to a smaller extent, with P38. These effects are associated with specific interactions between the solutes and the unique environments of the individual binding sites. Together, these results indicate complex electrostatic effects upon moving from the bulk to the binding pockets with shifts in the dipole distributions ranging from decreases through minimal changes to increases depending on the solute and binding pocket, with the effects occurring with both polar and apolar solutes. In addition, in certain cases specific interactions occur that lead to anisotropic perturbations of the dipole moments. These results, along with the wide distributions of the Drude versus additive dipoles indicates the physical forces dictating the equilibrium between the bulk phase and binding sites differ significantly between the additive and Drude models.

## D Vector Maps

Additional analysis of the impact of the inclusion of polarizability and lone pairs focused on the orientation of the solutes in the ligand binding pockets. This was performed by calculating vectors of selected solutes that were within 3 Å of a residue in the binding site. Vectors were based on the ND1 to HD1 bond for IMIA, OG to HG1 bond for MEOH, and C to O bond for AALD, with analysis performed on Factor Xa for IMIA and MEOH, P38 kinase for IMIA and MEOH, and TYK2 for AALD and MEOH. Figure 5 shows the comparison of the Drude and additive vector maps for these solutes.

With Factor Xa (Figure 5A and B), the Drude has a large number of IMIA N-H vectors towards the serine 184 hydroxyl oxygen, associated with an N-H...O hydrogen bond, whereas with the additive a smaller number of vectors are in this orientation. This scenario was also observed for IMIA with glycine 106 hydroxyl oxygen of P38 kinase (Figure 5E and F). Similar behaviour occurred for the Drude MEOH solute in Factor Xa (Figure 5C and D) and TYK2 (Figure 5K and L) where the MEOH O-H vector orients towards the hydroxyl in Factor Xa and away from the amino group in TYK2 thereby participating in hydrogen bonds. In both cases the orientation effect is less pronounced with the additive force field. The Drude AALD C=O vectors are dominated by interactions with the backbone amino group of glycine 106 for P38 kinase (Figure 5G and H) whereas no specific orientations are observed with the additive model. For TYK2 (Figure 5I and J) with the Drude, the majority of AALD C=O vectors point towards the amino group of lysine 40 consistent with a C=O...H-N hydrogen bond. In contrast, the additive vectors are not near the amino group, showing a clear distinction between the two force fields. These examples show the impact of lone pairs and electronic polarizability on the location and orientation of the solutes in the ligand binding sites, effects that contribute to the differences in the FragMaps in the ligand binding pockets between the two force fields (Table 1).

## Conclusions

The present results indicate a large impact associated with the inclusion of electronic polarization and lone pairs on the nature of the interactions of functional groups with proteins. This points to the importance of the improved treatment electrostatic interactions in force fields in more accurately modelling intermolecular interactions. Such improvements, which we anticipate will be achieved with other polarizable force fields, are related to variations in dipole moments upon moving between bulk and protein bound states that, along with lone pairs (or atomic multipoles in the case of AMOEBA<sup>29</sup>) lead to more defined orientations of the functional groups in the binding sites and produce increased occupancy of the functional groups in the bound state. Notably, these effects lead to a small, but systematic improvement in the prediction of the affinity of ligands to the protein in the context of the SILCS methodology. This observation is consistent with improvements in calculated binding affinities previously reported in free-energy perturbation studies using both the Drude and AMOEBA polarizable force fields,<sup>108–111</sup> and speaks to the utility of polarizable force fields in more accurately model molecular interactions.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgements

This work was supported by NIH grant R44GM109635 and R01GM131710 and the Samuel Waxman Cancer Research Foundation. The authors acknowledge computer time and resources from the Computer-Aided Drug Design (CADD) Center at the University of Maryland, Baltimore.

## References

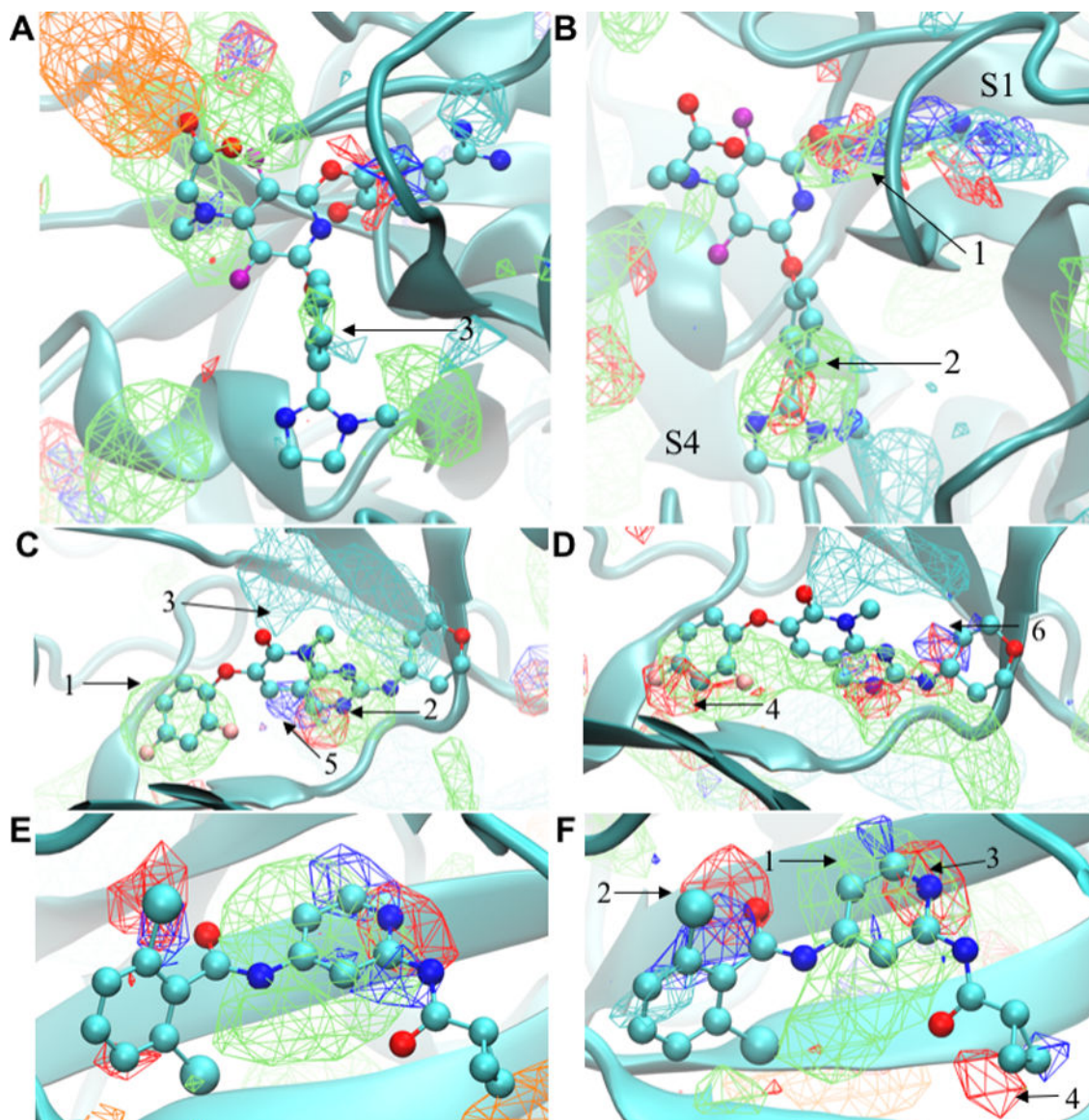
1. Goel H, Ling S, Ellis BN, Taconi A, Slater B and Rai N, *J Chem. Phys.*, 2018, 148, 224501. [PubMed: 29907054]
2. Chen M, Ko H-Y, Remsing RC, Andrade MFC, Santra B, Sun Z, Selloni A, Car R, Klein ML and Perdew JP, *Proc. Natl. Acad. Sci.*, 2017, 114, 10846–10851. [PubMed: 28973868]
3. Del Ben M, Hutter J and VandeVondele J, *J. Chem. Phys.*, 2015, 143, 054506. [PubMed: 26254660]
4. Vanommeslaeghe K, Guvench O and MacKerell AD Jr., *Curr. Pharm. Des.*, 2014, 20, 3281–3292. [PubMed: 23947650]
5. Durrant JD and McCammon JA, *BMC Biol.*, 2011, 9, 71. [PubMed: 22035460]
6. Hu J, Ma A and Dinner AR, *J. Comp. Chem.*, 2006, 27, 203–216. [PubMed: 16323162]
7. Deng Y and Roux B, *J. Chem. Phys.*, 2008, 128, 115103. [PubMed: 18361618]
8. Bernardi RC, Melo MC and Schulten K, *Biochim. Biophys. Acta-General Subjects*, 2015, 1850, 872–877.
9. MacKerell AD Jr., Brooks B, Brooks CL III, Nilsson L, Roux B, Won Y and Karplus M, in *Encyclopedia of Computational Chemistry*, eds. Schleyer P. v. R., Allinger NL, Clark T, Gasteiger J, Kollman PA, Schaefer HF III and Schreiner PR, John Wiley & Sons, Chichester, 1998, vol. 1, pp. 271–277.
10. Zhu X, Lopes PE and MacKerell AD Jr., *Wiley interdiscip. Rev. Comput. Mol. Sci.*, 2012, 2, 167–185. [PubMed: 23066428]
11. Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW and Kollman PA, *J. Am. Chem. Soc.*, 1995, 117, 5179–5197.
12. Wang J, Wolf RM, Caldwell JW, Kollman PA and Case DA, *J. Comp. Chem.*, 2004, 25, 1157–1174. [PubMed: 15116359]
13. Case DA, Cheatham TE, Darden T, Gohlke H, Luo R, Merz KM, Onufriev A, Simmerling C, Wang B and Woods RJ, *J. Comp. Chem.*, 2005, 26, 1668–1688. [PubMed: 16200636]
14. Oostenbrink C, Villa A, Mark AE and van Gunsteren WF, *J. Comp. Chem.*, 2004, 25, 1656–1676. [PubMed: 15264259]
15. Horta BA, Merz PT, Fuchs PF, Dolenc J, Riniker S and Hunenberger PH, *J. Chem. Theory Comput.*, 2016, 12, 3825–3850. [PubMed: 27248705]
16. Jorgensen WL, Maxwell DS and Tirado-Rives J, *J. Am. Chem. Soc.*, 1996, 118, 11225–11236.
17. Jorgensen WL and Tirado-Rives J, *J. Am. Chem. Soc.*, 1988, 110, 1657–1666. [PubMed: 27557051]
18. Nerenberg PS and Head-Gordon T, *Curr. Opin. Struct. Biol.*, 2018, 49, 129–138.
19. Halgren TA and Damm W, *Curr. Opin. Struct. Biol.*, 2001, 11, 236–242. [PubMed: 11297934]
20. Antila HS and Salonen E, *Methods Mol. Biol.*, 2013, 924, 215–241. [PubMed: 23034751]
21. Lagardere L, Jolly L-H, Lipparini F, Aviat F, Stamm B, Jing ZF, Harger M, Torabifard H, Cisneros GA, Schnieders MJ, Gresh N, Maday Y, Ren PY, Ponder JW and Piquemal J-P, *Chem. Sci.*, 2018, 9, 956–972. [PubMed: 29732110]
22. Cieplak P, Dupradeau F-Y, Duan Y and Wang J, *J. Phys.: Condens. Matter* 2009, 21, 333102. [PubMed: 21828594]
23. Rick SW and Stuart SJ, *Rev. Comp. Chem.*, 2002, 18, 89–146.

24. Baker CM, Wiley Interdiscip. Rev. Comput. Mol. Sci, 2015, 5, 241–254.
25. Bedrov D, Piquemal J-P, Borodin O, MacKerell AD, Roux B and Schröder C, Chem. Rev, 2019, 119, 7940–7995. [PubMed: 31141351]
26. Lemkul JA, Huang J, Roux B and MacKerell AD Jr., Chem. Rev, 2016, 116, 4983–5013. [PubMed: 26815602]
27. Cieplak P, Caldwell J and Kollman P, J. Comp. Chem, 2001, 22, 1048–1057.
28. Ponder JW, Wu C, Ren P, Pande VS, Chodera JD, Schnieders MJ, Haque I, Mobley DL, Lambrecht DS, DiStasio RA Jr., Head-Gordon M, Clark GN, Johnson ME and Head-Gordon T, J. Phys. Chem. B, 2010, 114, 2549–2564. [PubMed: 20136072]
29. Shi Y, Xia Z, Zhang J, Best R, Wu C, Ponder JW and Ren P, J. Chem. Theory Comput, 2013, 9, 4046–4064. [PubMed: 24163642]
30. Zhang C, Lu C, Jing Z, Wu C, Piquemal J-P, Ponder JW and Ren P, J. Chem. Theory Comput, 2018, 14, 2084–2108. [PubMed: 29438622]
31. Bauer BA and Patel S, Theor. Chem. Acc, 2012, 131, 1153.
32. Patel S and Brooks CL III, J. Comp. Chem, 2004, 25, 1–15. [PubMed: 14634989]
33. Huang J, Lemkul JA, Eastman PK and MacKerell AD Jr., J. Comp. Chem, 2018, 39, 1682–1689. [PubMed: 29727037]
34. Inakollu VS, Geerke DP, Rowley CN and Yu H, Curr. Opin. Struc. Biol, 2020, 61, 182–190.
35. Jing Z, Liu C, Cheng SY, Qi R, Walker BD, Piquemal J-P and Ren P, Annu. Rev. Biophys, 2019, 48, 371–394. [PubMed: 30916997]
36. Drude P, Millikan RA and Mann RC, The Theory of Optics, Longmans, Green, and Co., New York, 1902.
37. Lopes PEM, Huang J, Shim J, Luo Y, Li H, Roux B and MacKerell AD Jr., J. Chem. Theory Comp, 2013, 9, 5430–5449.
38. Yu W, Lopes PE, Roux B and MacKerell AD Jr., J. Chem. Phys, 2013, 138, 034508. [PubMed: 23343286]
39. Huang J, Lopes PE, Roux B and MacKerell AD Jr., J. Phys. Chem. Lett, 2014, 5, 3144–3150. [PubMed: 25247054]
40. Lemkul JA and MacKerell AD Jr., J. Chem. Theory Comput, 2017, 13, 2072–2085. [PubMed: 28398748]
41. Aytenfisu AH, Yang M and MacKerell AD Jr., J. Chem. Theory Comput, 2018, 14, 3132–3143. [PubMed: 29694037]
42. Lemkul JA and MacKerell AD Jr, J. Comp. Chem, 2018, 39, 2624–2646. [PubMed: 30515902]
43. Lin F-Y and MacKerell AD, J. Chem. Theory and Comp, 2018, 14, 1083–1098.
44. Lin F-Y, Doctoral Dissertation, 2019.
45. Harder E, Anisimov VM, Vorobyov IV, Lopes PEM, Noskov S, MacKerell AD Jr. and Roux B, J. Chem. Theory Comp, 2006, 2, 1587–1597.
46. Ding Y, Xu Y, Qian C, Chen J, Zhu J, Huang H, Shi Y and Huang J, J. Comput. Aid. Mol. Des, 2020, 1–15.
47. Ngo VA, Fanning JK and Noskov SY, Advanced Theory and Simulations, 2019, 2, 1800106.
48. Ngo V, da Silva MC, Kubillus M, Li H, Roux B, Elstner M, Cui Q, Salahub DR and Noskov SY, J. Chem. Theory Comput, 2015, 11, 4992–5001. [PubMed: 26574284]
49. Li H, Ngo V, Da Silva MC, Salahub DR, Callahan K, Roux B. t. and Noskov SY, J. Phys. Chem. B, 2015, 119, 9401–9416. [PubMed: 25578354]
50. Noskov SY, Lamoureux G and Roux B, J. Phys. Chem. B, 2005, 109, 6705–6713. [PubMed: 16851754]
51. Lemkul JA, Nucleic Acids Res, 2020, 48, 561–575.
52. Salisbury AM and Lemkul JA, J. Phys. Chem. B, 2018, 123, 148–159. [PubMed: 30525627]
53. Davidson DS, Brown AM and Lemkul JA, J. Mol. Biol, 2018, 430, 3819–3834. [PubMed: 29782833]
54. Raman EP, Yu W, Guvench O and Mackerell AD, J. Chem. Inf. Model, 2011, 51, 877–896. [PubMed: 21456594]

55. Guvench O and MacKerell AD Jr, PLOS Comput. Biol, 2009, 5, e1000435. [PubMed: 19593374]
56. Aloy P and Russell RB, Nat. Rev. Mol. Cell Biol, 2006, 7, 188. [PubMed: 16496021]
57. Yu W, Jo S, Lakkaraju SK, Weber DJ and MacKerell AD Jr, Proteins: Struct., Funct., Bioinf, 2019, 87, 289–301.
58. Ustach VD, Lakkaraju SK, Jo S, Yu W, Jiang W and MacKerell AD Jr., J. Chem. Inf. Model, 2019, 59, 3018–3035. [PubMed: 31034213]
59. Zhang H, Jiang W, Chatterjee P and Luo Y, J. Chem. Inf. Model, 2019, 59, 2093–2102. [PubMed: 30763080]
60. Hu X, Provasi D, Ramsey S and Filizola M, Biophys. J, 2019.
61. Ghanakota P, DasGupta D and Carlson HA, J. Chem. Inf. Model, 2019, 59, 2035–2045. [PubMed: 31017411]
62. Graham SE, Leja N and Carlson HA, J. Chem. Inf. Model, 2018, 58, 1426–1433. [PubMed: 29905479]
63. Bakan A, Nevins N, Lakdawala AS and Bahar I, J. Chem. Theory Comput, 2012, 8, 2435–2447. [PubMed: 22798729]
64. Sayyed-Ahmad A, Curr. Top. Med. Chem, 2018, 18, 2278–2283. [PubMed: 30499401]
65. Uehara S and Tanaka S, J. Chem. Inf. Model, 2017, 57, 742–756. [PubMed: 28388074]
66. Zariquiey FS, de Souza JV and Bronowska AK, Sci. Rep, 2019, 9, 1–14. [PubMed: 30626917]
67. Graham SE, Smith RD and Carlson HA, J. Chem. Inf. Model, 2018, 58, 305–314. [PubMed: 29286658]
68. Arcon JP, Defelipe LA, Modenutti CP, López ED, Alvarez-Garcia D, Barril X, Turjanski A. n. G. and Martí MA, J. Chem. Inf. Model, 2017, 57, 846–863. [PubMed: 28318252]
69. Kimura SR, Hu HP, Ruvinsky AM, Sherman W and Favia AD, J. Chem. Inf. Model, 2017, 57, 1388–1401. [PubMed: 28537745]
70. Sayyed-Ahmad A and Gorfe AA, J. Chem. Theory Comput, 2017, 13, 1851–1861. [PubMed: 28252958]
71. Lakkaraju SK, Yu W, Raman EP, Hershfeld AV, Fang L, Deshpande DA and MacKerell AD Jr., J. Chem. Inf. Model, 2015, 55, 700–708. [PubMed: 25692383]
72. Lakkaraju SK, Raman EP, Yu W and MacKerell AD Jr., J. Chem. Theory Comput, 2014, 10, 2281–2290. [PubMed: 24932136]
73. Adler M, Davey DD, Phillips GB, Kim S-H, Jancarik J, Rumennik G, Light DR and Whitlow M, Biochemistry, 2000, 39, 12534–12542. [PubMed: 11027132]
74. Goldstein DM, Soth M, Gabriel T, Dewdney N, Kuglstatter A, Arzeno H, Chen J, Bingenheimer W, Dalrymple SA, Dunn J, Farrell R, Frauchiger S, La Fargue J, Ghate M, Graves B, Hill RJ, Li F, Litman R, Loe B, McIntosh J, McWeeny D, Papp E, Park J, Reese HF, Roberts RT, Rotstein D, San Pablo B, Sarma K, Stahl M, Sung ML, Suttman RT, Sjogren EB, Tan Y, Trejo A, Welch M, Weller P, Wong BR and Zecic H, J. Med. Chem, 2011, 54, 2255–2265. [PubMed: 21375264]
75. Liang J, Tsui V, Van Abbema A, Bao L, Barrett K, Beresini M, Berezhkovskiy L, Blair WS, Chang C and Driscoll J, Eur. J. Med. Chem, 2013, 67, 175–187. [PubMed: 23867602]
76. GSK Trmd Dataset, <https://drugdesigndata.org/about/datasets/226>.
77. Gonzalez-Lopez de Turiso F, Sun D, Rew Y, Bartberger MD, Beck HP, Canon J, Chen A, Chow D, Correll TL and Huang X, J. Med. Chem, 2013, 56, 4053–4070. [PubMed: 23597064]
78. Raman EP, Yu W, Lakkaraju SK and Mackerell AD Jr., J. Chem. Inf. Model, 2013, 53, 3384–3398. [PubMed: 24245913]
79. Eisenhaber F, Lijnzaad P, Argos P, Sander C and Scharf M, J. Comp. Chem, 1995, 16, 273–284.
80. Van der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE and Berendsen HJC, J. Comput. Chem, 2005, 26, 1701–1718. [PubMed: 16211538]
81. Levitt M and Lifson S, J. Mol. Biol, 1969, 46, 269–279. [PubMed: 5360040]
82. Brooks BR, Brooks CL 3rd, MacKerell AD Jr., Nilsson L, Petrella RJ, Roux B, Won Y, Archontis G, Bartels C, Boresch S, Caflisch A, Caves L, Cui Q, Dinner AR, Feig M, Fischer S, Gao J, Hodoscek M, Im W, Kuczera K, Lazaridis T, Ma J, Ovchinnikov V, Paci E, Pastor RW, Post CB,

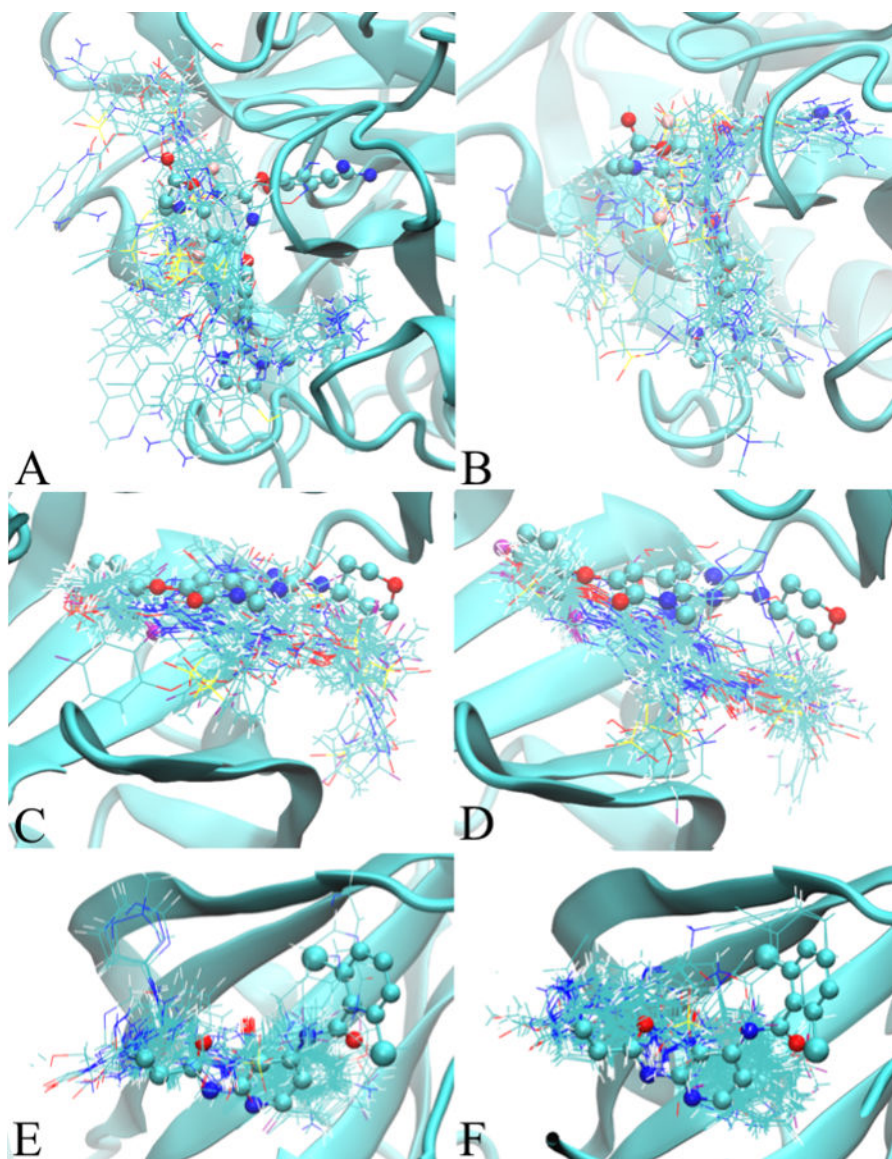


- Pu JZ, Schaefer M, Tidor B, Venable RM, Woodcock HL, Wu X, Yang W, York DM and Karplus M, *J. Comput. Chem.*, 2009, 30, 1545–1614. [PubMed: 19444816]
83. Huang J, Rauscher S, Nawrocki G, Ran T, Feig M, de Groot BL, Grubmuller H and MacKerell AD Jr., *Nat. Methods*, 2017, 14, 71–73. [PubMed: 27819658]
84. Vanommeslaeghe K, Hatcher E, Acharya C, Kundu S, Zhong S, Shim J, Darian E, Guvench O, Lopes P, Vorobyov I and Mackerell AD Jr., *J. Comp. Chem.*, 2010, 31, 671–690. [PubMed: 19575467]
85. Lin F-Y, Lopes PE, Harder ED, Roux B and MacKerell AD, *J. Chem. Inf. Model.*, 2018, 58, 993–1004. [PubMed: 29624370]
86. Small MC, Aytenfisu AH, Lin F-Y, He X and MacKerell AD Jr., *J. Comput. Aided Mol. Des.*, 2017, 31, 349–363. [PubMed: 28190218]
87. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW and Klein ML, *J. Chem. Phys.*, 1983, 79, 926–935.
88. Reiher WE, Harvard University, 1985.
89. Lamoureux G, Harder E, Vorobyov IV, Roux B and MacKerell AD Jr, *Chem. Phys. Lett.*, 2006, 418, 245–249.
90. Nosé S, *J. Chem. Phys.*, 1984, 81, 511–519.
91. Hoover WG, *Phys. Rev. A*, 1985, 31, 1695–1697.
92. Parrinello M and Rahman A, *J. Appl. Phys.*, 1981, 52, 7182–7190.
93. Hess B, Bekker H, Berendsen HJ and Fraaije JG, *J. Comp. Chem.*, 1997, 18, 1463–1472.
94. Essmann U, Perera L, Berkowitz ML, Darden TA, Lee H and Pedersen LG, *J. Chem. Phys.*, 1995, 103, 8577–8593.
95. Eastman P, Friedrichs MS, Chodera JD, Radmer RJ, Brunts CM, Ku JP, Beauchamp KA, Lane TJ, Wang L-P and Shukla D, *J. Chem. Theory Comput.*, 2012, 9, 461–469. [PubMed: 23316124]
96. Chowdhary J, Harder E, Lopes PE, Huang L, MacKerell AD Jr. and Roux B, *J. Phys. Chem. B*, 2013, 117, 9142–9160. [PubMed: 23841725]
97. Lamoureux G and Roux B, *J. Chem. Phys.*, 2003, 119, 5185–5197.
98. Allen MP and Tildesley DJ, *Computer Simulation of Liquids*, Oxford University Press, New York, 1989.
99. Faller CE, Raman EP, MacKerell AD and Guvench O, in *Fragment-Based Methods in Drug Discovery*, Springer, 2015, pp. 75–87.
100. Humphrey W, Dalke A and Schulten K, *J. Molec. Graph.*, 1996, 14, 33–38. [PubMed: 8744570]
101. Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH and Teller E, *J. Chem. Phys.*, 1953, 21, 1087–1092.
102. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S and Karplus M, *J. Comp. Chem.*, 1983, 4, 187–217.
103. Friesner RA, Prigogine I and Rice SA, *Computational methods for protein folding*, Wiley New York, 2002.
104. Warshel A, Sharma PK, Kato M and Parson WW, *Biochim. Biophys. Acta, Proteins Proteomics*, 2006, 1764, 1647–1676.
105. Lin FY and MacKerell AD Jr., *J. Phys. Chem. B*, 2017, 121, 6813–6821. [PubMed: 28657759]
106. Pearlman DA and Charifson PS, *J. Med. Chem.*, 2001, 44, 3417–3423. [PubMed: 11585447]
107. Lemkul JA, Savelyev A and MacKerell AD Jr, *J. Chem. Phys. Lett.*, 2014, 5, 2077–2083.
108. Jiao D, Golubkov PA, Darden TA and Ren P, *Proc. Natl. Acad. Sci.*, 2008, 105, 6290–6295. [PubMed: 18427113]
109. Shi Y, Wu C, Ponder JW and Ren P, *J. Comp. Chem.*, 2011, 32, 967–977. [PubMed: 20925089]
110. Villa F, MacKerell AD Jr, Roux B and Simonson T, *J. Phys. Chem. A*, 2018, 122, 6147–6155. [PubMed: 29966419]
111. Harger M, Li D, Wang Z, Dalby K, Lagardère L, Piquemal JP, Ponder J and Ren P, *J. Comp. Chem.*, 2017, 38, 2047–2055. [PubMed: 28600826]

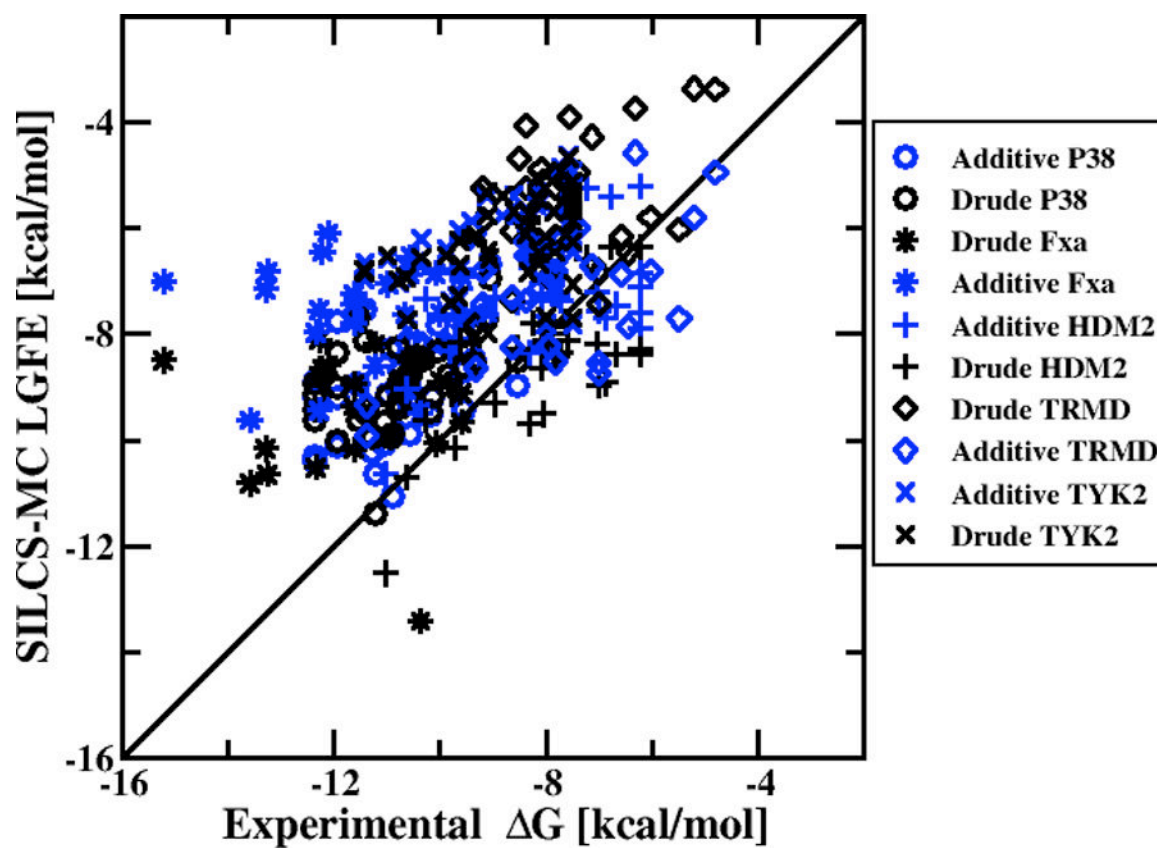


**Figure 1:**

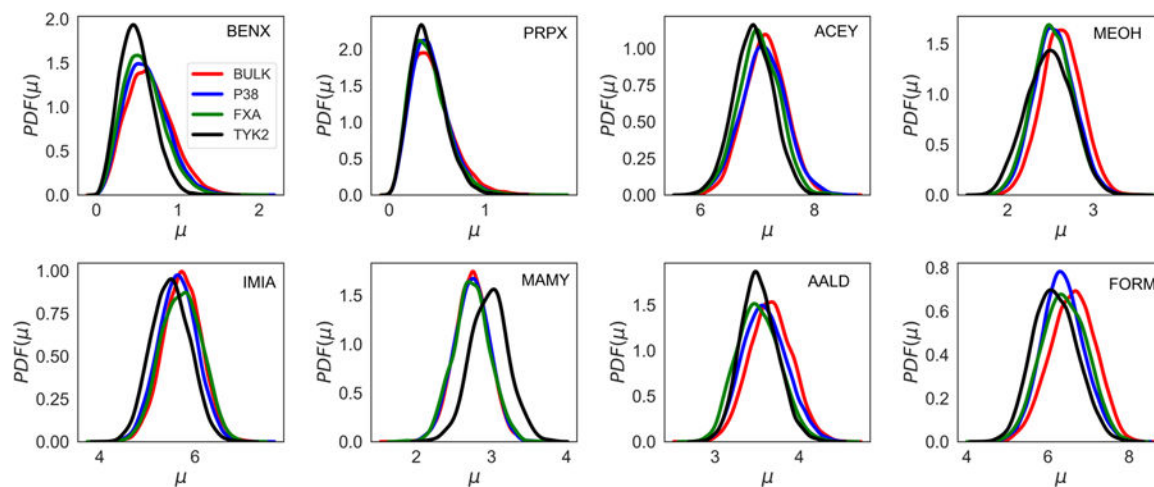
FragMaps overlaid on the proteins Factor Xa, P38 Map kinase and TYK2. Cartoon representations are shown based on the crystal conformations (PDB 1FJS, PDB 3FLN, and PDB 4GIH, respectively) with portions of the protein occluding the binding pocket view omitted. The ligands from the respective crystal structures are shown in CPK representation with atom type color. Panels A, C and E show the FragMaps for Factor Xa, P38 Map kinase and TYK2 obtained from the additive simulations while panels B, D, and F show the FragMaps for Factor Xa, P38 Map kinase and TYK2, respectively, obtained from the Drude simulations. The FragMaps color are GENN or APOLAR (green), GENA (red), GEND (blue), MAMN (cyan), and ACEO (orange). All FragMaps isocontour surfaces are displayed at a cutoff of  $-1.2$  kcal/mol. 2D images of the ligands are shown in Figure S1 of the Supporting Information.



**Figure 2:** Ligand minimum LGFE conformations from the Exhaustive SILCS-MC protocol for targeted proteins. The left (A, C, and E) and right (B, D, and F) panels shows the ligand conformations obtained from the additive and Drude SILCS-MC docking, respectively. Figures A/B, C/D, and E/F represent Factor Xa, P38, and TYK2, respectively. The crystallographic position of the ligand from the crystal structures are displayed in CPK model to show the location of the binding pocket. 2D images of the ligands are shown in Figure S1 of the supporting information.

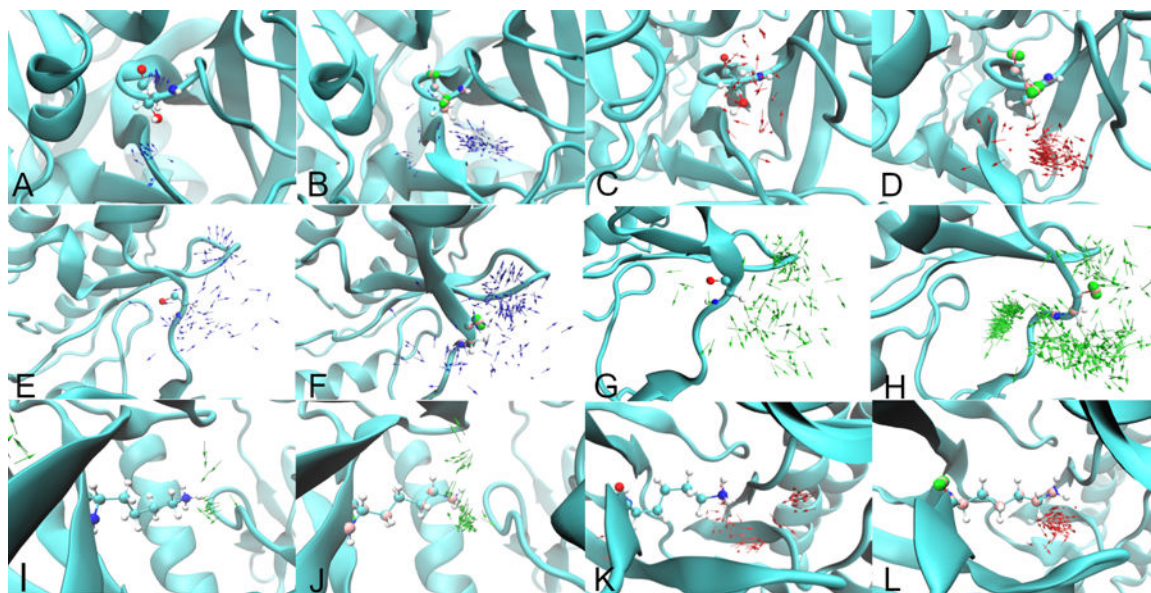


**Figure 3:**  
Correlation between the experimental binding data and the predicted LGFE scores for Drude and additive set with Exhaustive protocol.



**Figure 4:**

Total Dipole moment distribution obtained from the Drude simulations for all the eight solutes from the P38, Factor Xa, and TYK2 protein simulations. The red color shows the dipole moment for the solute in the bulk phase and the blue, green, and black color represents dipole moment in the ligand binding pocket of P38, Factor Xa, and TYK2 protein, respectively. The BENX, PRPX, ACEY, MEOH, IMIA, MAMY, AALD and FORM represents solutes for benzene, propane, acetate, methanol, imidazole, methylammonium, acetaldehyde, and formamide, respectively.



**Figure 5:** Vector maps for IMIA (Figure A and B) and MEOH (Figure C and D) adjacent to serine 184 of Factor Xa, IMIA (Figure E and F) and AALD (Figure G and H) adjacent to glycine 106 of P38 MAP kinase, AALD (Figure I and J) and MEOH (Figure K and L) adjacent to lysine 40 of TYK2. Panels A, C, E, G, I and K show the vector maps obtained from the additive simulations and Panels B, D, F, H, J, and L show the vector maps from the Drude simulations. With the Drude the visualization of the targeted binding-site residues includes the Drude particles (pink) and the lone pairs (green). The blue, red, and green color arrow represents the IMIA, MEOH, and AALD fragments.

**Table 1:**

Overlap coefficients for the different solutes in a cubic box encompassing the binding pockets in the Drude versus additive simulation for each protein target.

	<b>P38</b>	<b>Factor Xa</b>	<b>HDM2</b>	<b>TRMD</b>	<b>TYK2</b>
FragMap					
meoo	0.58	0.74	0.79	0.63	0.59
imin	0.49	0.62	0.75	0.48	0.53
iminh	0.50	0.61	0.75	0.52	0.54
foro	0.55	0.63	0.79	0.54	0.47
forn	0.55	0.62	0.80	0.61	0.50
mamn	0.52	0.47	0.77	0.57	0.34
aceo	0.54	0.53	0.79	0.34	0.60
benc	0.49	0.55	0.63	0.53	0.57
prpc	0.46	0.54	0.62	0.58	0.66
aalo	0.61	0.65	0.73	0.65	0.61
tipa	0.71	0.81	0.89	0.78	0.75
aalc	0.62	0.66	0.75	0.65	0.67
gehc	0.54	0.64	0.78	0.51	0.60
acec	0.48	0.49	0.75	0.32	0.56
Average	0.55	0.61	0.76	0.55	0.57

meoo: methanol O, imin: imidazole acceptor N, imih, imidazole donor N, foro: formamide O, forn, formamide N, mamn: methylammonium N, aceo, acetate O, benc, benzene Cs, prpc, propane Cs, aalo, acetaldehyde O, tipa, Tip3P O, aalc, acetaldehyde carbonyl C, gehc: imidazole Cs and acec: acetate carbonyl C.

Comparison of SILCS-MC LGFE scores and docked orientations with experimental data for the Exhaustive and Local SILCS\_MC docking protocols in Drude and additive FragMaps for different protein targets. The PI, RMSE, PC, RMSD and COMD signifies the predictive index, root mean square error, percent correct, root mean squared distance (Å), and center of mass difference (Å), respectively, along with the average LGFE- G<sub>bind</sub> (kcal/mol) difference.

**Table 2:**

System	Protocol	PI	RMSE	PC	RMSD	COMD	Average LGFE- G <sub>bind</sub>
P38	Exhaustive (Drude)	0.44	2.04	0.67	6.11	2.10	1.80
	Exhaustive (Additive)	0.36	2.09	0.63	7.28	1.69	1.77
	Local (Drude)	0.59	3.97	0.71	1.84	1.29	3.89
	Local (Additive)	0.55	3.93	0.69	1.26	0.88	3.85
Factor Xa	Exhaustive (Drude)	0.09	2.73	0.55	5.35	2.70	2.07
	Exhaustive (Additive)	-0.01	4.28	0.51	10.50	7.40	3.95
	Local (Drude)	-0.02	4.13	0.51	1.69	0.98	3.56
	Local (Additive)	0.27	9.06	0.58	2.68	1.56	8.83
HDM2	Exhaustive (Drude)	0.58	1.15	0.69	5.87	2.82	-0.55
	Exhaustive (Additive)	0.36	1.33	0.60	5.70	2.59	0.64
	Local (Drude)	0.75	1.47	0.75	1.84	1.16	1.14
	Local (Additive)	0.57	2.88	0.70	2.10	1.09	2.64
TYK2	Exhaustive (Drude)	0.61	2.61	0.69	5.66	2.80	2.41
	Exhaustive (Additive)	0.55	2.61	0.67	5.16	2.02	2.40
	Local (Drude)	0.56	3.80	0.63	2.53	1.74	3.63
	Local (Additive)	0.50	3.50	0.63	1.81	1.04	3.33
TRMD	Exhaustive (Drude)	0.46	2.35	0.64	6.45	4.82	1.91
	Exhaustive (Additive)	0.52	1.48	0.66	2.69	1.07	0.64
	Local (Drude)	0.43	4.11	0.60	1.28	0.85	3.79
	Local (Additive)	0.45	2.29	0.61	1.53	0.89	1.68



**Table 3:**

Comparison of the averaged dipole moment values for all the solutes in the binding pockets versus in bulk solution for Drude model of P38, Factor Xa, and TYK2 protein.

Protein	Solute	Pocket		Bulk		Difference	% Difference
		Average	Std. Error	Average	Std. Error		
P38 Kinase	BENX	0.616	0.005	0.656	0.003	-0.041	-6.60
	PRPX	0.431	0.003	0.447	0.002	-0.016	-3.74
	MAMY	2.741	0.005	2.740	0.001	0.001	0.05
	IMIA	5.634	0.029	5.731	0.003	-0.097	-1.72
	MEOH	2.546	0.003	2.630	0.002	-0.083	-3.27
	AALD	3.595	0.010	3.664	0.002	-0.068	-1.90
Factor Xa	ACEY	7.108	0.033	7.137	0.003	-0.029	-0.40
	FORM	6.368	0.014	6.606	0.005	-0.238	-3.74
	BENX	0.582	0.016	0.654	0.003	-0.072	-12.37
	PRPX	0.418	0.012	0.443	0.002	-0.025	-5.98
	MAMY	2.737	0.018	2.739	0.002	-0.002	-0.07
	IMIA	5.715	0.071	5.727	0.004	-0.012	-0.21
TYK2	MEOH	2.526	0.008	2.630	0.003	-0.104	-4.12
	AALD	3.515	0.027	3.666	0.002	-0.151	-4.30
	ACEY	7.023	0.030	7.132	0.003	-0.109	-1.55
	FORM	6.402	0.038	6.600	0.006	-0.198	-3.09
	BENX	0.495	0.009	0.656	0.003	-0.161	-32.55
	PRPX	0.407	0.008	0.449	0.001	-0.042	-10.29
TYK2	MAMY	2.970	0.026	2.738	0.002	0.232	7.82
	IMIA	5.488	0.042	5.732	0.003	-0.244	-4.45
	MEOH	2.489	0.032	2.627	0.002	-0.139	-5.57
	AALD	3.523	0.015	3.669	0.003	-0.146	-4.15
	ACEY	6.931	0.012	7.134	0.003	-0.204	-2.94

Protein	Solute	Pocket		Bulk		Difference	% Difference
		Average	Std. Error	Average	Std. Error		
	FORM	6.180	0.032	6.599	0.004	-0.419	-6.78

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript