Electrostatic Similarities between Protein and Small Molecule Ligands Facilitate the Design of Protein-Protein Interaction Inhibitors

Arnout Voet[•], **Francois Berenger**[•], **Kam Y. J. Zhang***

Zhang Initiative Research Unit, Institute Laboratories, RIKEN, Wako, Saitama, Japan

Abstract

One of the underlying principles in drug discovery is that a biologically active compound is complimentary in shape and molecular recognition features to its receptor. This principle infers that molecules binding to the same receptor may share some common features. Here, we have investigated whether the electrostatic similarity can be used for the discovery of small molecule protein-protein interaction inhibitors (SMPPIIs). We have developed a method that can be used to evaluate the similarity of electrostatic potentials between small molecules and known protein ligands. This method was implemented in a software called EleKit. Analyses of all available (at the time of research) SMPPII structures indicate that SMPPIIs bear some similarities of electrostatic potential with the ligand proteins of the same receptor. This is especially true for the more polar SMPPIIs. Retrospective analysis of several successful SMPPIIs has shown the applicability of EleKit in the design of new SMPPIIs.

Citation: Voet A, Berenger F, Zhang KYJ (2013) Electrostatic Similarities between Protein and Small Molecule Ligands Facilitate the Design of Protein-Protein Interaction Inhibitors. PLoS ONE 8(10): e75762. doi:10.1371/journal.pone.0075762

Editor: Andreas Hofmann, Griffith University, Australia

Received February 20, 2013; **Accepted** August 19, 2013; **Published** October 10, 2013

Copyright: © 2013 Voet et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was supported by the JSPS and the Initiative Research Unit program from RIKEN, Japan. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: kamzhang@riken.jp

• These authors contributed equally to this work.

Introduction

With the advent of the 'omics' era, it has become clear that most proteins do not act in solitude but depend on Protein-Protein Interactions (PPIs) to exert their biological function. It has been estimated that the number of PPIs in humans ranges from ~130,000 [1] to ~650,000 [2] and these PPIs are crucial for the regulation of many biological processes. PPIs are often involved in processes associated with diseases, therefore targeting PPIs with small molecule PPI inhibitors (SMPPIIs) opens a pipeline for the development of novel drug classes against a variety of diseases.

While many small molecule drugs targeting enzymes, nuclear receptors, ion channels and G-protein coupled receptors have been developed, the number of reported successes in the discovery of SMPPIIs remains fairly low. As a matter of fact, PPIs were once thought to be high hanging fruits for drug discovery [3]. PPIs were even considered to be undruggable, mostly because of their relative flat but extensive interfaces [4]. Though initially thought to be undruggable, an increasing number of SMPPIIs have been reported in recent years [5]. However, the number of deposited 3D SMPPII receptor complex structures remain far more limited than the number of reported successful cases. This hinders the understanding of their mechanism of action and chemical space properties [6]. Commonly used methods for screening are computational docking [7] and pharmacophore-based screening [8]. It was observed that the crucial interactions between a protein ligand and its protein receptor are often similar to those between the SMPPII and the protein receptor [9,10]. Thus, the PPI interface can be used to create a pharmacophore query to screen for small molecule ligands [11,12].

Another approach is to exploit the principle of electrostatic complementarity in molecular recognition. Next to steric complementarity, electrostatics are one of the main driving forces involved in molecular recognition [13]. Despite the complex biophysical nature of the electrostatic potential, calculations for macromolecular systems are nowadays tractable [14,15].

Electrostatics are known to play a key role in protein-DNA [16], protein-protein [17] and protein-substrate [13] recognitions. Given the importance of electrostatics for the molecular recognition event, electrostatics have been used to study protein similarity [18–20] and the nature of protein-protein interactions [17,21–24]. More specifically, the electrostatic complementarity between protein-protein interfaces has long been a subject of investigation [22,23]. Using the correlation of electrostatic potentials as a quantitative measure, the electrostatic complementarity between PPI interfaces has been demonstrated [17,24]. Other studies focused on the conservation of the electrostatic potentials through evolution [25] and its role in molecular association kinetics [26].

It is generally accepted that there is a high degree of complementarity in shape and electrostatics between a ligand and its receptor. This implies that molecules with similar shape and electrostatic properties may bind to the same receptor. This principle has been used to identify small molecule inhibitors similar to natural substrates or known inhibitors by screening for compounds with similar shape, volume and electrostatics [27–30].

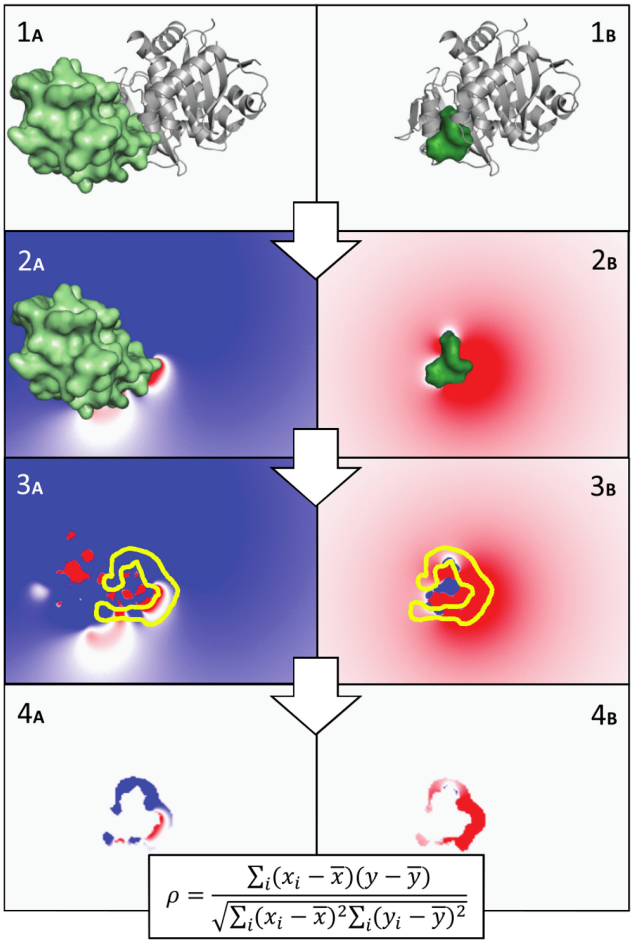

Figure 1. Overview of EleKit applied to PDB codes 2B4J (1_A) and 3LPU (1_B). The ligand protein (L_P) is shown as a green surface in 1_A and 2_A. The ligand small molecule (L_{SM}) is shown as a smaller green surface in 1_B and 2_B. The receptor protein (R_P) is shown as a gray cartoon in 1_A and 1_B. and are placed on (1_A and 1_B). The electrostatic potentials of and are calculated and stored in distinct grids (2_A and 2_B). Then, a mask is created to select the solvent region near the interface (3_A and 3_B). Finally, the similarity between electrostatic potentials of and over this region is calculated using the Spearman rank correlation coefficient (4_A and 4_B).
doi:10.1371/journal.pone.0075762.g001

An SMPPII cannot occupy the same shape and volume as its much bigger protein-ligand counterpart. However, it can still be assumed that there is some local electrostatic potential similarity between an SMPPII and a ligand protein, since they recognize the same binding site on the receptor. A recent example of the usefulness of taking electrostatic potential similarity into account while designing an SMPPII can be found in the work of Cavalluzo *et al.* [31], where an SMPPII was designed *de novo* by including electrostatic similarity. This success has motivated our effort to systematically investigate the complementarity in electrostatic potential between small molecules and protein ligands binding to the same protein receptor, and its potential use to assist in the rational design of SMPPIIs. For this purpose, a tool named EleKit was developed.

Table 1. APBS commands used for a protein.

read	
	mol pqr ligandprotein.pqr
end	
elec	
	mg-auto
	dime 161 193 161
	cglen 105.5210 127.4643 91.1458
	fglen 82.0712 94.9790 73.6152
	cgcent -6.603267 -6.904766 -18.393622
	fgcent -6.603267 -6.904766 -18.393622
	mol 1
	lpbe
	bcfl sdh
	pdie 2.0000
	sdie 78.5400
	srfm smol
	chgm spl2
	sdens 10.00
	srad 1.40
	swin 0.30
	temp 298.15
	calcenergy no
	calcforce no
	write smol dx ligandprotein.ms
	write pot dx ligandprotein
end	
quit	

For a small molecule, the input and output file names would differ. Parameters were determined with pdb2pqr.py. Grid-related parameters vary upon each case (dime, cglen, fglen, cgcent and fgcent).
doi:10.1371/journal.pone.0075762.t001

Methods

To compute the partial charges and electrostatic potentials, EleKit builds upon PDB2PQR [32] and APBS [15]. EleKit requires two sets of complex structures in order to calculate the electrostatic similarity between a protein ligand and a small molecule ligand: (i) the PPI complex of the protein-ligand (L_P) with the protein-receptor (R_P) and (ii) a small molecule ligand (L_{SM}) in its predicted or experimentally determined conformation on the protein-receptor (R_P).

The EleKit method is shown schematically in figure 1. First, the electrostatic potentials around and are computed using APBS (parameters listed in table 1) and stored in 3D grids. Since only the area where and intersect is most likely to be relevant for molecular recognition, a bit mask is created on the electrostatic potential grids (figure 2). The goal of this mask is to take into account only those points in space that are not only in the solvent region around and but also near the interface atoms of R_P. To create this mask, a distance cutoff is needed. This distance is used when dilating (a morphological mathematical operation) the molecular surface. Based on the hydrogen bond length (~2.5Å) and the facts that enough points are needed for correlation and that the local similarity is our focus, a cutoff value ranging from 1.4 Å to 3.5 Å seems reasonable. All experiments reported in this study were

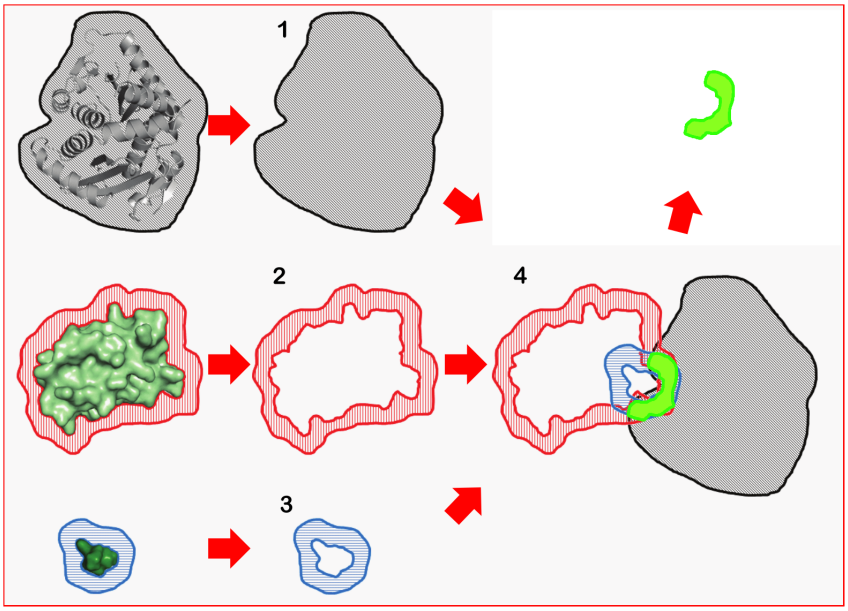

Figure 2. Overview of the bit-mask construction in EleKit. (1) a near-or-inside mask of R_P is created, (2) a near-but-not-inside mask of L_P is created, (3) a near-but-not-inside mask of L_{SM} is created, (4) the logical conjunction of the three masks is used to select points to correlate from the electrostatic potentials of L_P and L_{SM}.
doi:10.1371/journal.pone.0075762.g002

performed with an intermediate cutoff value of 2.0 Å. Using 3.0 Å or 4.0 Å would have very little impact on the results (data not shown). Finally, the similarity between electrostatic potentials of and is assessed by correlating values at the grid points within the mask using the Spearman rank-order correlation coefficient (ρ). Additional similarity scores (Carbo index [33], Hodgkin index [34], Pearson's r and a Tanimoto score) are also calculated.

EleKit is written in OCaml [35] (http://biocaml.org) and computations are parallelized by the Parmap library [36]. Experiments were run on Linux computing nodes with 2.4GHz Intel Xeon processors. EleKit takes between ten seconds to two minutes per ligand molecule and can parallelize the computation of several ligands when run on a multi-core machine. EleKit is released as open source and available from the authors' website http://www.riken.jp/zhangiru/software.html.

Results and Discussion

The EleKit method was applied to analyze previously reported cases of SMPPIIs, for which accurate structures of the PPI as well as the SMPPII receptor complex are available in the PDB (table 2). Additionally, the SMPPIIs are required to bind in the PPI interface, allowing for a substantial overlap between the protein ligand and the SMPPII and thus excluding allosteric inhibition mechanisms.

The approach used in EleKit to perform comparison of electrostatic potentials resembles what has been done previously on proteins [18–21]. Analysis of Electrostatic Similarities of Proteins (AESOP) [20], the method of Dlugosz *et al.* [19] and Protein Interaction Property Similarity Analysis (PIPSA) [18,37] also use APBS as their electrostatic computation engine. PIPSA can also use University of Houston Brownian Dynamics [38] (UHBD). While EleKit relies on the Spearman rank-order correlation coefficient (as McCoy *et al.* [17]), PIPSA uses the Hodgkin index [34] to numerically assess the similarity of electrostatic potentials. AESOP uses the Average Normalized Difference [39]. The method of Dlugosz *et al.* [19] approximates the electrostatic potential with spherical harmonics and uses a similarity index specifically designed to compare the obtained rotation-invariant descriptors. EleKit, similarly to several other methods [18,19,37], uses boolean masks to select a region over which electrostatic potentials are compared. All methods vary in the way masks are constructed.

Analysis with EleKit

Electrostatic similarity analysis for these different SMPPII-related structures indicate that several exhibit correlation. In general, correlation between electrostatic potentials of SMPPIIs and electrostatic potentials of the respective ligand proteins are observed (table 2). This is especially true for the SMPPIIs targeting the HDM2:p53, HIV-1 Integrase:LEDGF/p75, Integrin:Fibrinogen, IL2:IL2R and XIAP:smac interactions. The highest similarity between a protein ligand and a small molecule ligand can be observed in the HIV-1 Integrase:LEDGF/p75 and the Integrin:Fibrinogen interactions and their respective inhibitors. In these cases, ρ is on average ~0.52 and ~0.73 respectively (table 2). The origin of these classes of SMPPIIs can be traced back to pharmacophore based discovery of lead compounds designed to mimic the interactions observed at the PPI interface [40,41].

For the inhibitors of the HDM2:p53 interaction, the majority of the inhibitors exhibit electrostatic potential similarity. However, a few show low correlations ($\rho < 0.2$) and in one case even some anti-correlation ($\rho \approx -0.15$). Interestingly, the Tanimoto score shows similarity in all HDM2:p53 cases. The electrostatic potentials between inhibitors and protein ligands in ZipA:FtsZ and VHL:HIF1 still correlate although less strongly than in other cases. These inhibitors are observed to be less active when tested. For inhibitors targeting the XIAP:smac interaction, which originated from peptidomimetic design, some compounds exhibit lower similarity than expected. This can be explained by the divergence of conformations of the receptor protein, since the XIAP:smac complex was solved by NMR while the structures of

Table 2. Analysis of SMPPII electrostatic mimicry.

PPI target	SMPPI complex (observations)	#points	ρ	r	t_2
HDM2:p53	1rv1	5515	0.275	0.263	0.816
(1ycr)	1t4e (g)	5281	0.433	0.355	0.757
	1ttv	4862	-0.152	-0.123	0.588
	3jkz	4384	0.200	0.018	0.566
	3lbj (g)	5159	0.353	0.190	0.611
	3lbk (g)	4681	0.316	0.209	0.883
	3lbl	4129	0.131	-0.015	0.672
	3tu1 (g)	4492	0.589	0.329	0.764
	4dij	4758	0.239	0.111	0.619
IL2:IL2R	1m48 (g)	3186	0.315	0.276	0.658
(1z92)	1m49 (g)	3523	0.758	0.654	0.737
	1py2	3231	-0.011	0.095	0.583
	1pw6 (g)	4119	0.653	0.574	0.678
	1qvn	3819	0.093	0.098	0.626
Integrase:LEDGF/p75	3lpt (g)	3653	0.468	0.437	0.205
(2b4j)	3lpu (g)	3489	0.460	0.464	0.220
	4dmn (g)	3880	0.471	0.534	0.229
	4e1m (g)	3612	0.609	0.601	0.228
	4e1n (g)	3486	0.615	0.591	0.267
Integrin:Fibrinogen	2vdm (g)	6471	0.731	0.751	0.617
(2vdo)	2vc2 (g)	6469	0.736	0.720	0.614
VHL:HIF1 (1lqb)	3zrc (g)	4135	0.123	0.127	0.062
XIAP:smac	2i3i (g)	5979	0.535	0.501	0.650
(1g3f)	1tft (g)	5853	0.377	0.409	0.525
	3eyl (g)	4377	0.321	0.375	0.506
	2jk7 (g)	5446	0.464	0.447	0.631
	3clx	5570	0.055	0.127	0.431
	3cm7	6005	0.227	0.229	0.534
	3hl5	6044	0.263	0.302	0.484
	3mup (g)	6028	0.480	0.404	0.671
	3g76 (g)	5008	0.396	0.391	0.371
	3oz1	5340	0.191	0.184	0.580
ZipA:FtsZ	1s1s (w_i)	4320	0.150	-0.061	0.704
(1f47)	1y2f (w_i, g)	3878	0.365	0.173	0.604
	1y2g (w_i)	3809	-0.318	-0.147	0.335
	1s1j (w_i, g)	2691	0.331	0.088	0.465

ρ: Spearman rank correlation coefficient; r: Pearson linear correlation coefficient; t2: a Tanimoto score (positive or negative electrostatic potential); observations: **g** = good ($\rho > 0.3$), $\mathbf{w_i}$ = weak inhibitor (potency <100 μM in the literature).

doi:10.1371/journal.pone.0075762.t002

XIAP bound to inhibitors were solved by X-ray crystallography. The PPI complex solved by NMR spectroscopy are more difficult to superpose onto the crystal structure conformation obtained for the SMPPII complex. The inhibitors of the IL2:IL2R interaction are well known for binding to the IL2R interface by causing a rotameric change of a phenyl alanine creating a binding pocket [42]. In this case, the PPI interface is only partially covered in a hydrophobic area caused by the induced fit. However, the observed similarity between the ligand protein and the inhibitor mainly originates from the mimicry of the arginine guanidinium group, which is not influenced by conformational changes or induced fit.

There are no significant electrostatic correlations found in the cases of the inhibitors of the Bcl2 family of proteins, the TNFα trimerization and the HPV polymerase. A careful analysis of the structures of these molecules revealed that the SMPPII in these cases is bound after a major reorganization of the receptor protein surface at the PPI interface. For the SMPPIIs bound to the Bcl2

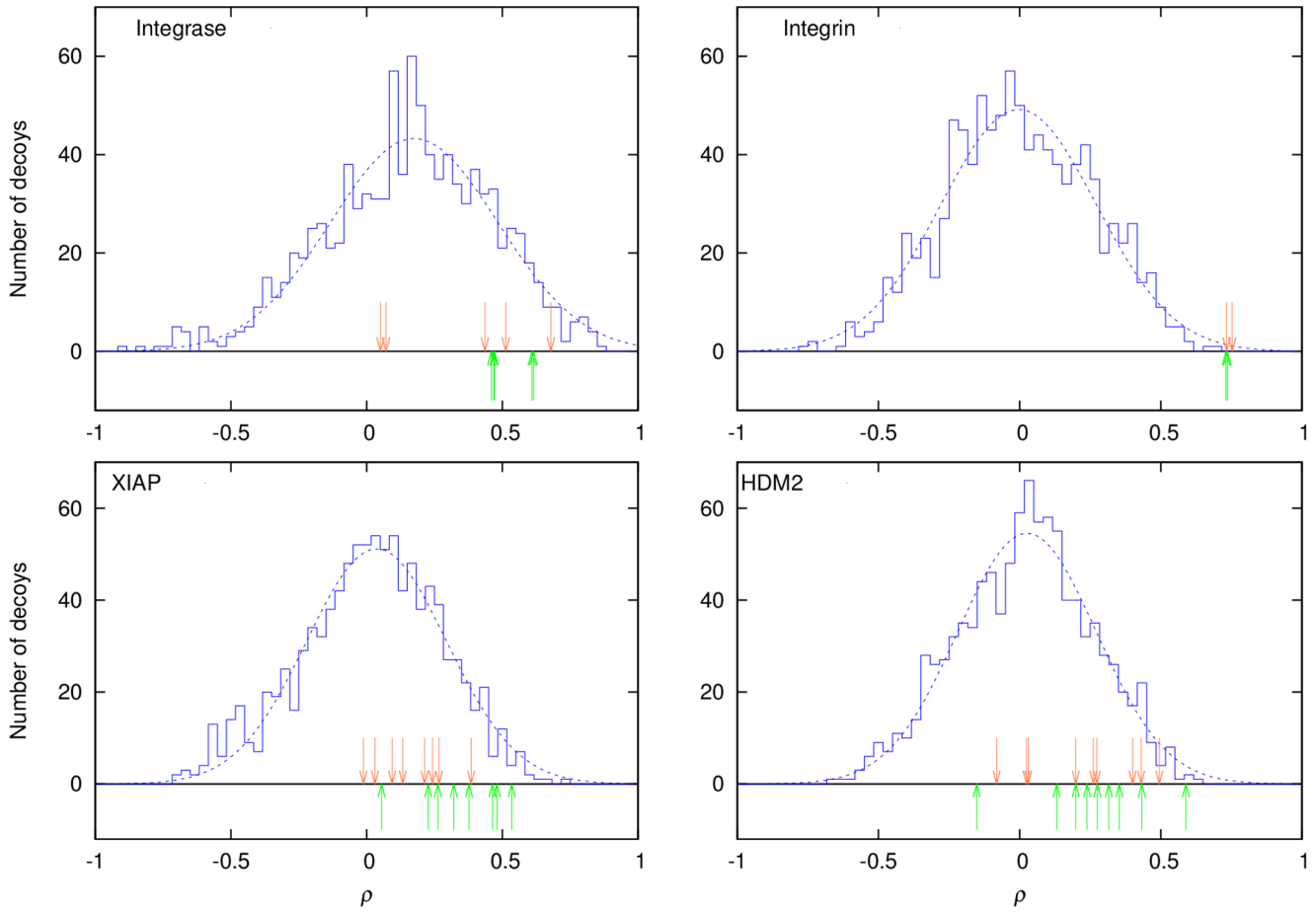

Figure 3. Distribution of Spearman scores (ρ) for active ligands among decoys. EleKit Spearman scores for a population of ligand decoys are shown as a blue histogram. The active ligands at their best docked pose in terms of RMSD to the crystallized active ligand are shown as orange arrows and the crystallized active ligands as green arrows. The dotted line is a Gaussian fitted to the decoys' histogram.
doi:10.1371/journal.pone.0075762.g003

family proteins, there is a major induced fit not only involving side chain atoms, but also including a rearrangement of a single helix, in order to comfortably fit the SMPPII inside the same cleft that was originally occupied by a full and more bulky α-helical protein ligand. The inhibitors of the TNFα and HPV polymerase bind in a pocket at the PPI interface created by the assumption of different side chain orientations with more open conformations. Furthermore, the SMPPIIs that break the E1:E2 interaction of the HPV polymerase act as a dimer. In these cases, the SMPPIIs do not act by mimicking and competing with the ligand protein and no similarity of electrostatic potentials is observed.

EleKit is able to assess electrostatic potential similarity by a variety of measures including ρ, r and a Tanimoto score (table 2). Overall, relying on ρ over r is preferred as it is more robust and does not suffer from uncertainties in interpreting the significance of the observed correlations [43]. The Tanimoto score, which is based on binning the electrostatic potential values as positive or negative does not seem to work in cases where the overall charge of the protein ligand is significantly different from that of the small molecule. An example is the case of the inhibitors of the Integrase:LEDGF/p75 interaction, where the protein ligand has a high positive net charge while the SMPPII has one negatively charged group. The Tanimoto scores are significantly affected and tend to be low. This problem is not observed when using correlation scores. The local shape similarity between the protein ligand and an SMPPII in EleKit is reflected by the number of electrostatic potential grid points being correlated. The more locally similar in shape a ligand protein is to a given SMPPI, the more grid-points will remain in the final mask.

Application of EleKit

To test the utility of EleKit for the post-filtering of results from virtual screening such as docking, four different cases where strong electrostatic similarities between SMPIIs and protein ligands have been observed were selected for further analysis. In each case, a set of 100 diverse decoy compounds that bear similar chemical properties as the active SMPPII were extracted from the purchasable subset of the ZINC database [44] using the procedure described by Huang *et al.* [45,46] to create the Directory of Useful Decoys (DUD).

To position these decoy compounds inside the receptor protein, as if it were a virtual screening experiment, molecular docking was performed using the Molecular Operating Environment (MOE) with the London dG score [47,48]. The receptor protein structure was extracted from the coordinates of the respective PPI complex structure and optimized using the Protonate3D functionality from MOE [47]. Similarly, the active compounds were also docked inside their respective receptor structures. For each compound, the top scoring docked pose was retained and the electrostatic potential similarity with the ligand protein was calculated.

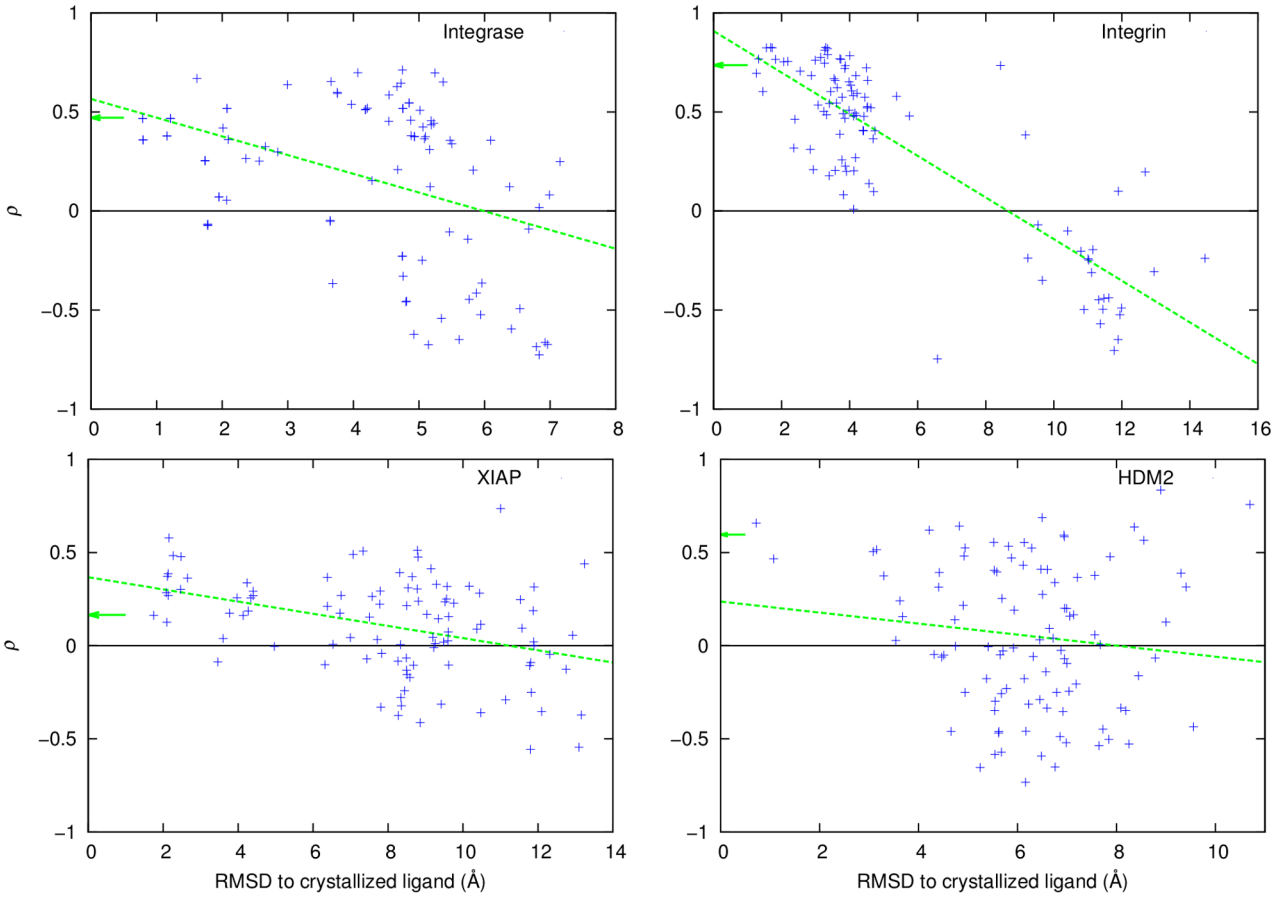

Figure 4. EleKit Spearman scores (ρ) versus RMSD to crystal structure for computationally docked poses of known active ligands. Each green arrow indicates the Spearman score for the known active ligand as seen in the crystal structure. The green dotted line was obtained using gnuplot's fit command (χ^2_{red}(Integrase)$\simeq$0.15, χ^2_{red}(Integrin)$\simeq$0.06, χ^2_{red}(XIAP)$\simeq$0.06, χ^2_{red}(HDM2)$\simeq$0.16).
doi:10.1371/journal.pone.0075762.g004

As expected, EleKit scores (based on a ligand protein and a ligand small molecule pair) are uncorrelated to docking scores (based on a ligand small molecule protein complex). The Spearman correlation between docking scores and EleKit scores fall in the range [−0.1;0.1] in every case (data not shown). In all four cases, the Spearman rank correlation coefficient for the decoys follows a Gaussian distribution (figure 3). The mean of the distribution is situated at $\rho \approx 0.0$ with an exception for HIV-1 Integrase:LEDGF/p75 ($\rho \approx 0.17$). In this case, the active SMPPIIs are all mainly hydrophobic in nature except for one acid functional group, thus all decoys also bear a similar group. During the placement of those decoys into the hydrophobic pocket, this functional group has a high likelihood of adopting a similar arrangement to the active compounds, creating a higher correlating tendency and shifting the mean ρ. This distribution is independent from the scoring function as revealed by the low R^2 values for the correlation between London dG and ρ. The highest observed value was $R^2 \approx 0.006$ for the decoys of the Integrin:Fibrinogen interaction.

When SMPPIIs in their experimentally determined binding modes and with their best docked conformations are considered, a positive ρ is always observed (figure 3). Only in the case of HDM2:p53, one of the nine active molecules has a negative ρ. Furthermore, in the cases of the Integrin:Fibrinogen and Integrase:LEDGF/p75, their ρ values reside on the even higher end of the distribution. In general, a cut-off value of at least $\rho = 0$ can be suggested as it would rightfully discard approximately half of the decoys. Such filtering using ρ would only remove one inhibitor of the HDM2:p53 interaction, based on experimentally determined poses. However, the electrostatic similarity for docking results is influenced by the correctness of the predicted binding mode. In the Integrase case, the docking algorithm was unable to identify the correct binding mode within it's top solution for two compounds (the two orange arrows nearest to $\rho = 0$ in the Integrase plot of figure 3). For one of these compounds, this can be explained by the small size of the molecule (PDB: 3lpt) and its lack of an important hydrophobic group, which leaves a huge hydrophobic cavity in the protein receptor unoccupied. However, the docking algorithm forced this small inhibitor to fill the unoccupied hydrophobic cavity resulting in a wrong predicted binding mode for this inhibitor. The second compound (PDB: 4e1n) has a significantly larger substituent group and would require a minor induced fit to bind correctly. The conformational difference of the receptor protein between its ligand protein and ligand small molecule bound forms can be problematic. In the case of the XIAP:smac inhibitors, this conformational difference exists since the structure of the PPI complex was determined using NMR spectroscopy and the structures of the SMPII complexes were determined by X-ray crystallography. The hydrophobic nature of the receptor protein can be a challenge. In the HDM2:p53 interaction, only a limited number of polar interactions that might

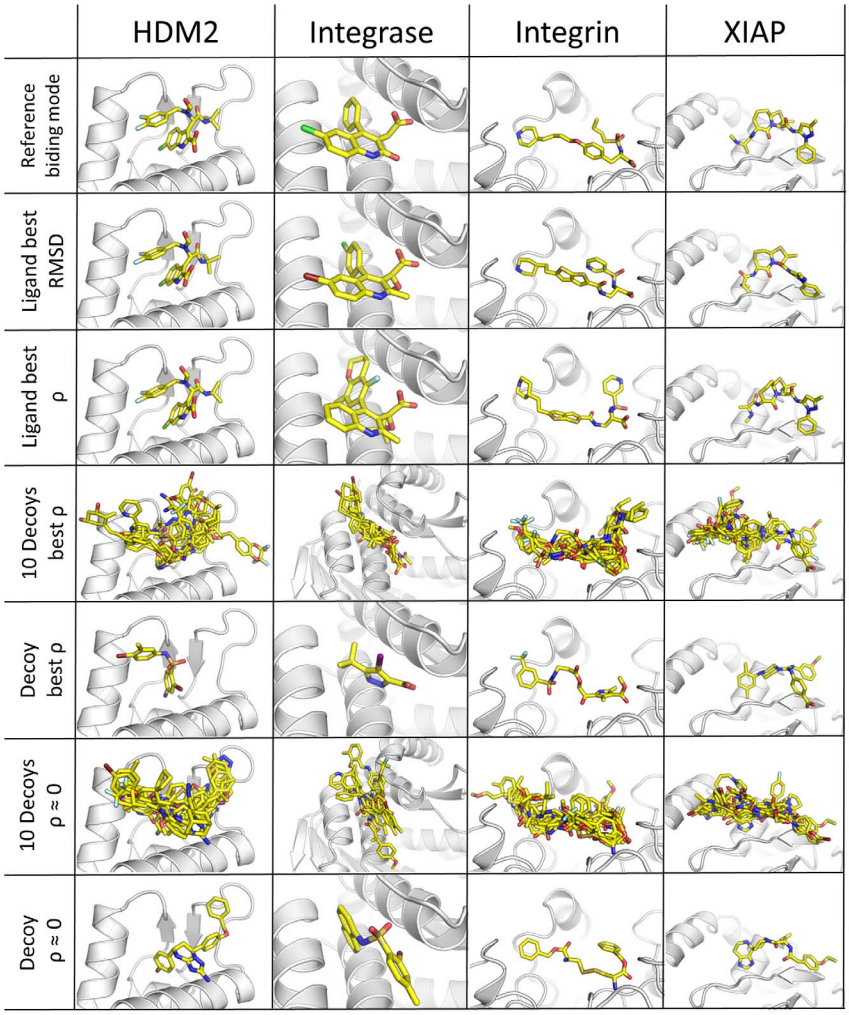

Figure 5. Analysis of docking poses. Comparison of binding modes predicted by docking for known active ligands and decoys to the reference ligands binding mode as experimentally determined. The best poses based on RMSD to the crystal structure, the pose with best ρ (ligand and decoys) and poses with ρ near zero (decoys) are shown. These binding modes indicate that the binding mode of the docked compounds are similar to the binding observed in X-ray or NMR structures. Despite the decoy compounds have binding modes in which their (a)polar contacts are similar to those of the active ligands, their electrostatic similarity with the ligand protein is different.
doi:10.1371/journal.pone.0075762.g005

help orienting the molecules in the right binding mode are present in the pocket.

An overall analysis of the docked conformation revealed that in every case the docking algorithm was able to reproduce binding modes of the active compounds in agreement with the crystallographically determined binding modes. In the four receptors examined in details (Integrase, Integrin, XIAP, HDM2), computational docking was able to place the active ligands in binding modes almost identical to those determined crystallographically (RMSD less than 1 or 2 Å, figure 4). The higher ρ even corresponds to binding modes that are closer in RMSD (Fconv [49] was used to compute RMSD for small molecules) to the experimentally determined poses. Similarly, the decoy compounds were docked within the right binding pocket making similar contacts with the receptor protein as the active compounds (figure 5), therefore validating the suitability of the docking simulations. Despite the decoy compounds made similar contacts compared to the binding modes of the active ligands, it is clear that the electrostatic similarity of the decoy compounds with the ligand protein has a normal distribution, with its mean ρ around 0. The ligands presented similar chemical groups in similar places driven by the complementarity of polar interactions in the pocket in a majority of the cases. The sole exception is found in the case of the HDM2/p53 that is hallmarked by a mainly apolar interface. Nevertheless, the apolar functions of the decoys and active ligands overlap in the binding mode.

The further away from the crystallographic pose the docked ligand is, the lower the Spearman rank correlation becomes. As a remark, the Receiver Operating Characteristic (ROC) analysis is typically used to assess the predictive and enrichment power of a method. But due to the lack of a significant number of active SMPPIIs for which structural information is available for a single target, this type of analysis could not be performed.

The development of EleKit was inspired by the computational work on electrostatic complementarity at protein-protein interfaces by McCoy *et al.* [17]. But EleKit bears salient differences with this former study. Whereas McCoy *et al.* studied the complementarity of protein-protein interfaces, EleKit measures the local similarity between one ligand protein and small molecules targeting the same receptor interface. McCoy *et al.* measured the

correlation of electrostatic potentials at molecular surface points while EleKit works on a 3D volume in the solvent region near the binding interface. There are some significant prior works that compare electrostatic potentials and other molecular interaction fields for proteins only [18–21,37,50].

EleKit was also inspired by the commercial tools EON [27–29] and ElectroShape [30]. In effect, these tools can assess similarity of electrostatic potentials. However, they work exclusively on small molecules. Thus, they can identify compounds similar to a known small molecule inhibitor, but this is not applicable when searching for a first-in-class SMPPII [31].

Conclusions

We have developed a method (EleKit) to investigate the similarity between protein ligands and their respective SMPPIIs. Our analysis of available SMPPII structures indicates that in cases where SMPPIIs bind without induced fit, there is similarity of electrostatic potentials at the interacting interface between a protein ligand and a small molecule inhibitor. This insight can be applied to post-filter virtual screening results to remove unpromising compounds and thus has some potential for the rational design of novel SMPPIIs.

Acknowledgments

We thank RIKEN, Japan, for an allocation of computing resources on the RIKEN Integrated Cluster of Clusters (RICC). Dr. Eleanor Banwell is acknowledged for proofreading the manuscript.

Author Contributions

Conceived and designed the experiments: AV FB KYJZ. Performed the experiments: AV FB. Analyzed the data: AV FB KYJZ. Wrote the paper: AV FB KYJZ.

References

1. Stumpf MPH, Thorne T, de Silva E, Stewart R, An HJ, et al. (2008) Estimating the size of the human interactome. Proc Natl Acad Sci U S A 105: 6959–64.
2. Venkatesan K, Rual JF, Vazquez A, Stelzl U, Lemmens I, et al. (2009) An empirical framework for binary interactome mapping. Nat Meth 6: 83 – 90.
3. Wells JA, McClendon CL (2007) Reaching for high-hanging fruit in drug discovery at protein-protein interfaces. Nature 450: 1001–1009.
4. Verdine GL, Walensky LD (2007) The Challenge of Drugging Undruggable Targets in Cancer: Lessons Learned from Targeting BCL-2 Family Members. Clin Cancer Res 13: 7264–7270.
5. Villoutreix BO, Labbe CM, Lagorce D, Laconde G, Sperandio O (2012) A Leap into the Chemical Space of Protein-Protein Interaction Inhibitors. Curr Pharm Des 18: 4648–67.
6. Morelli X, Bourgeas R, Roche P (2011) Chemical and structural lessons from recent successes in protein-protein interaction inhibition (2P2I). Curr Opin Chem Biol 15: 475 – 81.
7. Dias R, de Azevedo WFJ (2008) Molecular Docking Algorithms. Curr Drug Targets 9: 1040–7.
8. Horvath D (2011) Pharmacophore-Based Virtual Screening, volume 672 of *Methods Mol Biol*. Humana Press, 261–98 pp.
9. Fry DC (2008) Drug-like inhibitors of protein-protein interactions: A structural examination of effective protein mimicry. Curr Protein Pept Sci 9: 240–7.
10. Fry DC (2012) Small-Molecule Inhibitors of Protein-Protein Interactions: How to Mimic a Protein Partner. Curr Pharm Des 18: 4679–84.
11. Voet A, Zhang KYJ (2012) Pharmacophore Modelling as a Virtual Screening Tool for the Discovery of Small Molecule Protein-protein Interaction Inhibitors. Curr Pharm Des 18: 4586–98.
12. Voet A, Banwell E, Sahu K, Heddle J, Zhang KYJ (2013) Protein interface pharmacophore mapping tools for small molecule protein: Protein interaction inhibitor discovery. Curr Top Med Chem 13: 989–1001.
13. Nray-Szab G (1993) Analysis of molecular recognition: steric electrostatic and hydrophobic complementarity. J Mol Recognit 6: 205–10.
14. Nicholls A, Honig B (1991) A rapid finite difference algorithm, utilizing successive over-relaxation to solve the PoissonBoltzmann equation. Journal of Computational Chemistry 12: 435–445.
15. Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA (2001) Electrostatics of nanosystems: application to microtubules and the ribosome. Proc Natl Acad Sci U S A 98: 10037–41.
16. Rohs R, West SM, Sosinsky A, Liu P, Mann RS, et al. (2009) The role of DNA shape in protein-DNA recognition. Nature 461: 1248–1253.
17. McCoy AJ, Chandana Epa V, Colman PM (1997) Electrostatic complementarity at protein/protein interfaces. J Mol Biol 268: 570–84.
18. Blomberg N, Gabdoulline RR, Nilges M, Wade RC (1999) Classification of protein sequences by homology modeling and quantitative analysis of electrostatic similarity. Proteins: Structure, Function, and Bioinformatics 37: 379–387.
19. Dlugosz M, Trylska J (2008) Electrostatic similarity of proteins: Application of three dimensional spherical harmonic decomposition. The Journal of Chemical Physics 129: 015103.
20. Kieslich CA, Morikis D, Yang J, Gunopulos D (2011) Automated computational framework for the analysis of electrostatic similarities of proteins. Biotechnology Progress 27: 316–325.
21. Gorham R, Kieslich C, Morikis D (2011) Electrostatic Clustering and Free Energy Calculations Provide a Foundation for Protein Design and Optimization. Annals of Biomedical Engineering 39: 1252–1263.
22. Janin J, Chothia C (1990) The structure of protein-protein recognition sites. Biol Chem 265: 16027–30.
23. Jones S, Thornton JM (1995) Protein-protein interactions – a review of protein dimer structures. Progr Biophy molec Biol 63: 31–65.
24. Chau PL, Dean PM (1994) Electrostatic complementarity between proteins and ligands. 1. Charge disposition, dielectric and interface effects. J Comput-Aided Mol Des 8: 51325.
25. Kieslich CA, Morikis D (2012) The Two Sides of Complement C3d: Evolution of Electrostatics in a Link between Innate and Adaptive Immunity. PLoS Comput Biol 8: e1002840.
26. Dlugosz M, Antosiewicz JM, Zielinski P, Trylska J (2012) Contributions of Far-Field Hydrodynamic Interactions to the Kinetics of Electrostatically Driven Molecular Association. J Phys Chem B 116: 5437–5447.
27. Muchmore SW, Souers AJ, Akritopoulou-Zanze I (2006) The Use of Three-Dimensional Shape and Electrostatic Similarity Searching in the Identification of a Melanin-Concentrating Hormone Receptor 1 Antagonist. Chem Biol Drug Des 67: 174–176.
28. Naylor E, Arredouani A, Vasudevan SR, Lewis AM, Parkesh R, et al. (2009) Identification of a chemical probe for NAADP by virtual screening. Nat Chem Biol 5: 220–226.
29. OpenEye Scientific Software, Inc (2010) OEChem 1.7.4. Santa Fe, NM, USA: OpenEye Scientific Software, Inc.
30. Armstrong M, Morris G, Finn P, Sharma R, Moretti L, et al. (2010) ElectroShape: fast molecular similarity calculations incorporating shape, chirality and electrostatics. Journal of Computer-Aided Molecular Design 24: 789–801.
31. Cavalluzzo C, Voet A, Christ F, Singh BK, Sharma A, et al. (2012) De novo design of small molecule inhibitors targeting the LEDGF/p75-HIV integrase interaction. RSC Advances 2: 974–84.
32. Dolinsky TJ, Nielsen JE, McCammon JA, Baker NA (2004) PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. Nucleic Acids Res 32: W665–7.
33. Carbó R, Leyda L, Arnau M (1980) How similar is a molecule to another? An electron density measure of similarity between two molecular structures. International Journal of Quantum Chemistry 17: 1185–1189.
34. Hodgkin EE, Richards WG (1987) Molecular similarity based on electrostatic potential and electric field. International Journal of Quantum Chemistry 32: 105–110.
35. Leroy X, Doligez D, Frisch A, Garrigue J, Rémy D, et al. (2011) The OCaml system release 3.12 Documentation and user's manual. INRIA, France.
36. Danelutto M, Di Cosmo R (2012) A ''Minimal Disruption'' Skeleton Experiment: Seamless Map and Reduce Embedding in OCaml. Procedia Computer Science 9: 1837 – 1846.
37. Wade RC, Gabdoulline RR, De Rienzo F (2001) Protein interaction property similarity analysis. International Journal of Quantum Chemistry 83: 122–127.
38. Davis ME, Madura JD, Luty BA, McCammon J (1991) Electrostatics and diffusion of molecules in solution: simulations with the University of Houston Brownian dynamics program. Computer Physics Communications 62: 187 – 197.
39. Petke J (1993) Cumulative and discrete similarity analysis of electrostatic potentials and fields. Journal of Computational Chemistry 14: 928–933.
40. Hartman G, Egbertson M, Halczenko W, Laswell WL, Duggan ME, et al. (1992) Non-peptide fibrinogen receptor antagonists. 1. Discovery and design of exosite inhibitors. J Med Chem 35: 46402.
41. Christ F, Voet A, Marchand A, Nicolet S, Desimmie B, et al. (2010) Rational design of small-molecule inhibitors of the LEDGF/p75-integrase interaction and HIV replication. Nat Chem Biol 6: 442–8.
42. Arkin MR, Randal M, DeLano WL, Hyde J, Luong TN, et al. (2003) Binding of small molecules to an adaptive proteinprotein interface. Proc Natl Acad Sci U S A 100: 1603–8.
43. Press WH, Teukolsky SA, Vetterling WT, Flannery BP (2007) Numerical Recipes 3rd Edition: The Art of Scientific Computing. New York, NY, USA: Cambridge University Press.

44. Irwin JJ, Shoichet BK (2005) ZINC A Free Database of Commercially Available Compounds for Virtual Screening. J Chem Inf Model 45: 177–82.
45. Huang N, Shoichet BK, Irwin JJ (2006) Benchmarking Sets for Molecular Docking. J Med Chem 49: 6789–801.
46. Mysinger MM, Carchia M, Irwin JJ, Shoichet BK (2012) Directory of Useful Decoys, Enhanced (DUD-E): Better Ligands and Decoys for Better Benchmarking. J Med Chem 55: 6582–94.
47. Chemical Computing Group Inc (2011) Molecular Operating Environment (MOE) 10th edition. 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7: Chemical Computing Group Inc.
48. Naim M, Bhat S, Rankin KN, Dennis S, Chowdhury SF, et al. (2007) Solvated interaction energy (SIE) for scoring protein-ligand binding affinities. 1. Exploring the parameter space. J Chem Inf Model 47: 122–33.
49. Neudert G, Klebe G (2011) fconv: Format conversion, manipulation and feature computation of molecular data. Bioinformatics 27: 1021–2.
50. Ullmann GM, Hauswald M, Jensen A, Kostić NM, Knapp EW (1997) Comparison of the Physiologically Equivalent Proteins Cytochrome c6 and Plastocyanin on the Basis of Their Electrostatic Potentials. Tryptophan 63 in Cytochrome c6 May Be Isofunctional with Tyrosine 83 in Plastocyanin. Biochemistry 36: 16187–16196.