PEER REVIEWED

Accurate Protein Electrostatics Add a New Dimension to Structure-based Drug Design

Electrostatic interactions are at the heart of drug discovery. The combination of advanced molecular force fields and improved visualisation capabilities have enabled chemists to expand their ligand-only view of the electrostatics of their system to include proteins, water and ligands. Important insights into the causes of ligand binding are now available on desktop computers and laptops, giving chemists unprecedented control over drug design.

Protein Fields have Great Potential

The most important factors affecting molecular recognition are electrostatics, 3D shape and hydrophobicity. Being able to create accurate computational models for all three of these factors across both ligand and protein active sites holds great potential for successful drug design.

Ligand electrostatics have become increasingly accurate over recent years and are now widely used in drug design[1]. Computational and scientific advances have meant that we have moved from thinking about electrostatics in terms of specific features of the ligand – such as the potential hydrogen bond network – to a detailed view of electron-rich and electron-poor regions around the ligand. The ability to see this detail has enabled chemists to gain a far better understanding of how activity is affected by chemical changes, leading to great improvements in new molecule design and increased efficiency in lead optimisation.

Until recently, the corresponding view of protein electrostatics had remained limited. Protein-ligand contacts tended to be viewed in the context of a limited number of interaction types, focusing primarily on hydrogen bonds and hydrophobic contacts. Now, with new methods and force fields, it is possible to model the electrostatics of proteins in far more detail, leading to practical insights into how and why particular ligands are effective inhibitors.

Protein Interaction Potentials

Protein fields, or protein interaction potentials (PIPs), work by probing a virtual protein with a charged atom to detect regions of positive and negative within the protein active site. The results are visualised as a surface with contours that reflect the change in potential. These detailed maps of the electrostatic character of the protein active site help researchers to understand more about the fundamental processes that underlie ligand-protein binding.

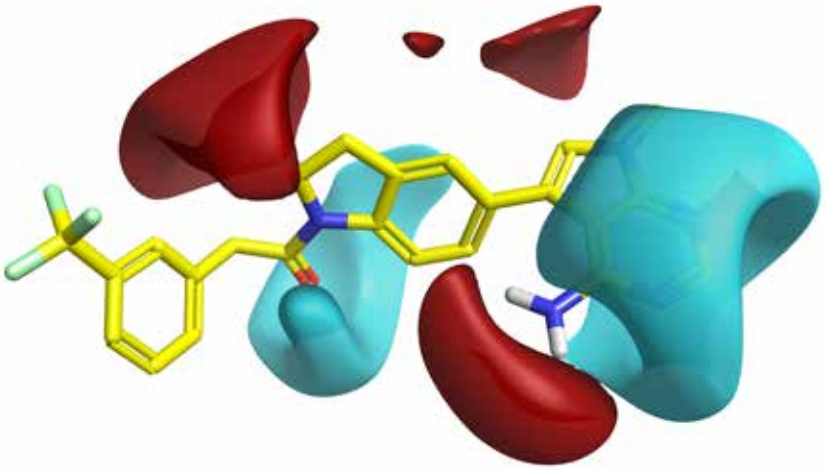

Figure 1: Ligand electrostatics (contoured at 3.5 Kcal/mole) for a 0.4nM PERK kinase inhibitor showing negative (blue) and positive (red) electrostatic interaction potentials.

PIPs can be calculated in seconds, making it possible to compare calculations for different conformations of the same protein. For example, chemists can see how loop movements can change the binding of fragments that are not interacting directly with the moving residues. Different proteins in the same family can also be compared, to fully understand where and how selectivity can be achieved and where this might prove elusive.

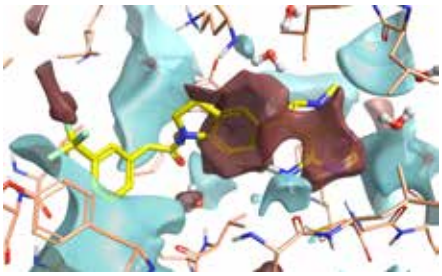

Figure 2: Protein fields (contoured at 3.5 Kcal/mole) on the dry protein structure from PDB 4G31, PERK Kinase showing electron negative (blue) and positive (red) regions of the active site. Ligand (yellow) shown for orientation. The protein electrostatics predict electron-rich aromatics are required in the centre of the pocket (ligand has an indazole) and for the hinge-binding motif (ligand has pyrrolopyrimidine). Protein fields predict trifluoro aromatic and associated benzylic CH2 to be favourable in the DFG out pocket.

However, the accurate calculation of protein electrostatics requires a thorough electrostatic treatment of all amino acids within the protein active site, not just those with a polar side chain. In particular, the treatment of aromatic residues is critical to understanding where pi-stacking would be preferred with an electron-rich or an electron-poor ligand. This is achieved in more advanced force fields such as AMOEBA[2] or Cresset's XED[3], which have a more complex electrostatic model allowing for the properties of aromatic residues to be more correctly computed.

The Need to Understand Water

In addition to a more complex description of atoms, the correct handling of water is critical to success in generating accurate electrostatics inside protein active

sites. No calculation of electrostatic potentials in the active site can be complete without also considering whether water molecules are tightly bound and contributing to the potentials and ligand binding.

Here there are two possible solutions: modelling water as explicit molecules, or including a mathematical model to represent the water that is present.

Modelling water using a mathematical construct to simulate the electrostatic effect of water molecules on the ligand or protein electrostatics is a popular way to solve many problems within physical chemistry. However, in the consideration of the electrostatics of the protein active site, this approach is limited. The problem is that the protein active site is not bulk solvent, nor is it a vacuum, nor pure protein. In general, it exists at the intersection of solvent and protein phases. As such, many of the assumptions of continuum water models break down.

To calculate protein interaction potentials requires an approach that can both simulate the electrostatic shielding effects of bulk water when considering amino acid residues that are distal to the point of interest, while also in the same calculation give full weighting to residues that are only a few Ångstroms away. This is achieved using a dielectric function that has a distance dependence but is also sensitive to the local electrostatic environment. Using this more complex electrostatic modification, close residues have more influence than those further away from the active site and yet cooperation between residues is maintained. This approach works well for many protein-ligand systems with good complementarity observed between the protein electrostatics and those of the ligand.

However, in many proteins, localised water molecules are central to the functioning of the protein or critical to maintaining the correct protein fold. These waters cannot sensibly be modelled by a continuum solvation model, and it is better to think of them as effectively being part of the protein system rather than part of the solvent.

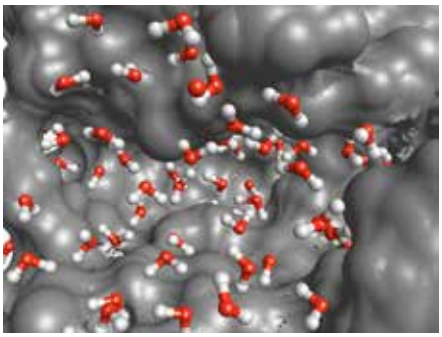

Figure 3: Where does bulk water stop and bound water begin? Many computational models deal with water as a solvent continuum which breaks down as water becomes explicitly bound.

Ligand, Protein and Water Molecules Must Be Treated Equally Well

The Reference Interaction Site Model (RISM) is a modern approach to solvation based on the integral equation theory of liquids [3D RISM 4]. It has seen increasing use as a method to analyse the structure of water in and around protein active sites.

The accuracy of 3D-RISM depends on the closure used and the functions used to compute the intermolecular potential between the protein and water molecules. In practice, the potential function consists of van der Waals interactions and electrostatic interactions. Using 3D-RISM in combination with an electrostatically accurate force field should lead to more accurate 3D-RISM water placements and energies.

The calculation of water positions and energetics on empty proteins using 3D-RISM is not new. However, as we understand more about ligand binding and water's role in mediating interactions with the protein, there is a need to understand how the stability of water changes with ligand binding. To do this requires force fields that treat the ligand, protein and water molecules equally and equally well. Some more advanced protein force fields such as AMOEBA require extensive parameterisation before they can be applied to ligands, while traditional ligand force fields such as MMFF lack the required complexity in their electrostatics. The XED force field bridges this gap: it is a universal ligand and protein force field that includes more detailed electrostatics, enabling RISM calculations to produce accurate results both for *apo* proteins and protein-ligand complexes.

Combining Protein Fields and Water Informs Molecular Design

To demonstrate the effectiveness of calculating protein and ligand electrostatics from the same force field, together with accurate water prediction, several ligand-protein complexes were downloaded from the Protein Data Bank (PDB) into the structure-based design application Flare[5].

Method

The ligand and protein structures were carefully prepared using Build Model[6]. Residues lining the active site were minimised with the XED force field.

Positive and negative protein interaction potentials were calculated and displayed as iso-surfaces, both for the 'dry' protein structure and the 'hydrated' structure, including crystallographic water molecules predicted to be strongly bound by 3D-RISM analysis as implemented in Flare.

The ligand electrostatic potentials were also calculated using the XED force field and compared with the protein interaction potentials of the dry and hydrated protein active site.

Application to Bruton's Tyrosine Kinase

Bruton's tyrosine kinase (BTK) is a target important to proliferation of cancers that depend heavily on B-cells such as chronic lymphocytic leukemia and mantle cell lymphoma but autoimmune diseases such as rheumatoid arthritis. Several crystal structures of BTK bound to inhibitors are available in the Protein Data Bank. Examining structures 4ZLZ and 4Z3V[7], the ligands interact with the active site by making H-bonds both with the hinge region and P-loop backbone residues. For the ligand from 4ZLZ, the interaction with the P-loop is mediated by a water molecule. The protein interaction potentials of these structures were

calculated and studied for insights into ligand design.

Each protein was downloaded, prepared and protein interaction potentials added in the region of the active site. The ligand and all crystallographic water were ignored in the calculation.

The electrostatics of the dry active site of 4Z3V match the ligand fields in a very precise manner. As would be expected, the hinge residues give rise to regions of negative (blue below) associated with H-bond acceptor carbonyls of Glu^{475} and Met^{477} while the H-bond donor NH of Met^{477} gives rise to a region of positive in the protein interaction potential. More surprisingly, this positive region stretches away from the hinge to encompass much of the region occupied by the ligand cinnoline ring. This strongly suggests that ligands with electron-rich heterocycles are preferred in this area. The edge of the cinnoline ring also shows favourable interactions, particularly the 5 and 7-position hydrogens which are close to or enter regions of negative potential on the protein.

Examining the protein interaction potential in the region of the ligand's pendant indazole ring shows further favourable interactions from the NH to a region of negativity created by the carbonyl of Gly^{414} and from the 4-H and 5-methyl into a region of negative formed by the residues 525–527. Interestingly, this negative patch comes from multiple sources – the carbonyl of Arg^{525}, the side chain carbonyl of Asn^{526} and the carbonyl of Cys^{525}. Additionally, the NH of Phe^{413} on the P-loop creates a region of positive that is matched by the nitrogen lone pair on the indazole ring. However, this region of positive mimics that found in the hinge region by stretching into the region of space occupied by the electron-rich heterocycle.

In contrast, the protein interaction potential of the dry active site of 4ZLZ shows a good, but imperfect, match of the ligand fields (compare Figure 4 and Figure 5). While the match of the protein fields in the region of the cinnoline ring is similar,

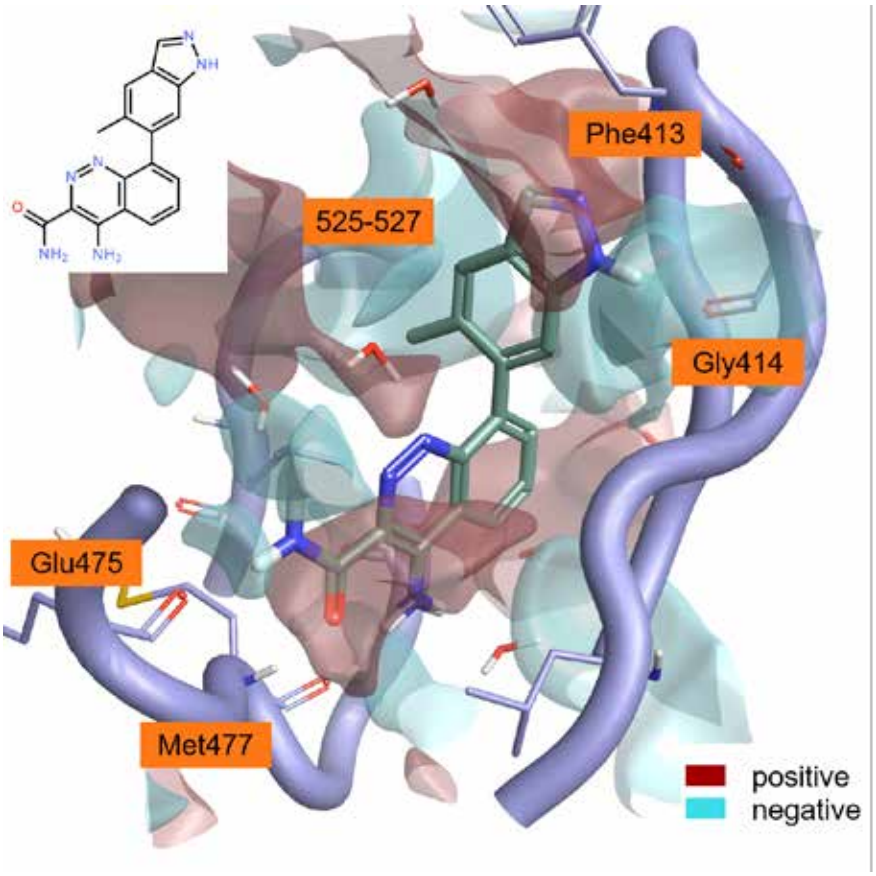

Figure 4: protein interaction potentials for 4Z3V.

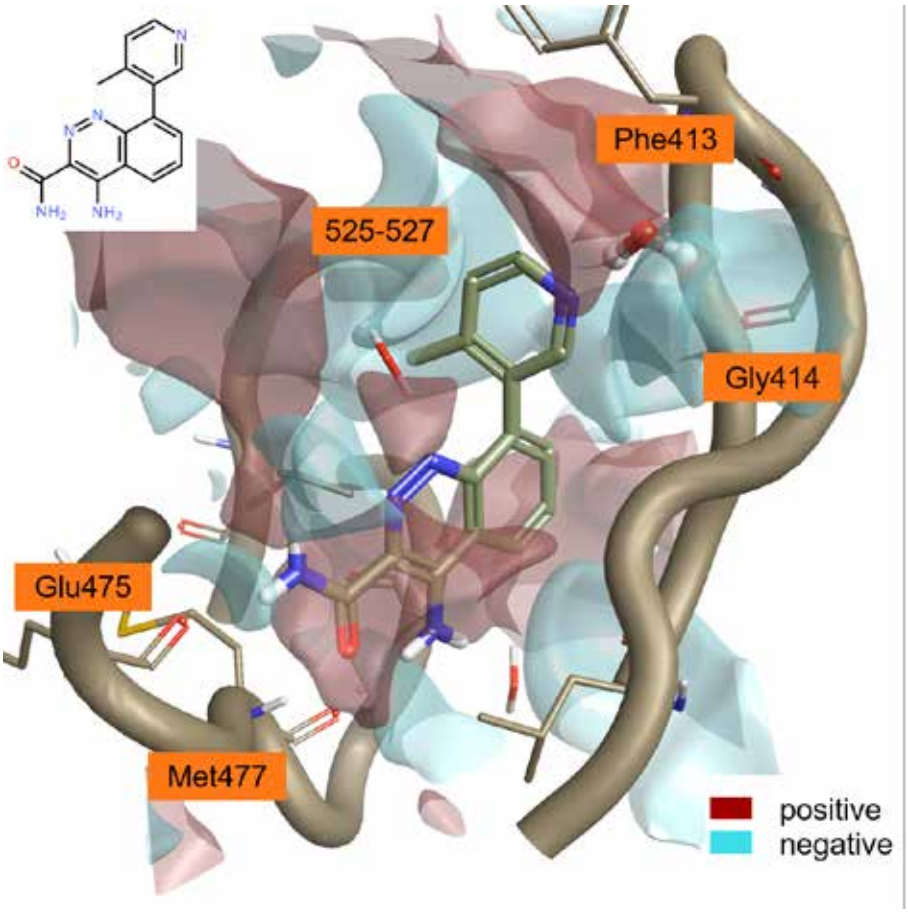

Figure 5: protein interaction potentials for 4ZLZ showing (in ball and stick rendering) a water molecule that fits the protein potentials and mediates an H-bonding interaction between the protein and ligand.

the field derived from the P-loop residues does not stretch to the ligand's pyridyl group. Instead the positive and negative fields encode for a water molecule that mediates an interaction between the protein and ligand. In this case, inclusion of this water molecule in the definition of the protein active site restores the complementarity between protein interaction potential and observed ligand.

Interestingly, the complementarity between protein interaction potential and ligand when including the water molecule in the calculation is mimicked when looking at ligand electrostatics. Including the water molecule as a part of the ligand gives an electrostatic pattern that more closely matches the pattern of the 4Z4V indazole ligand than is obtained from the native 4LZL ligand alone.

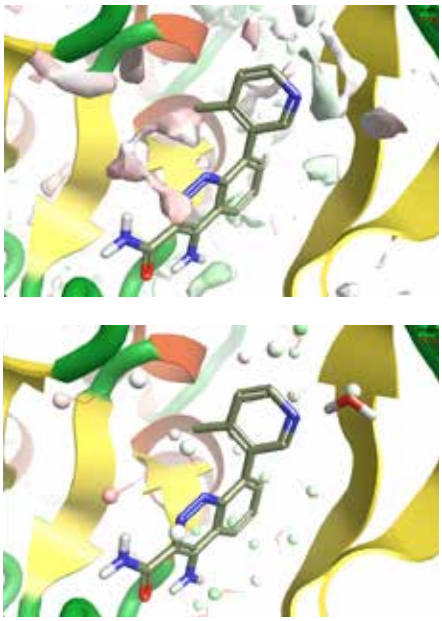

Figure 7: RISM results on 4ZLZ. The top picture shows the oxygen isodensity surface at ρ=5. The bottom shows the localised RISM waters, coloured by ΔG.

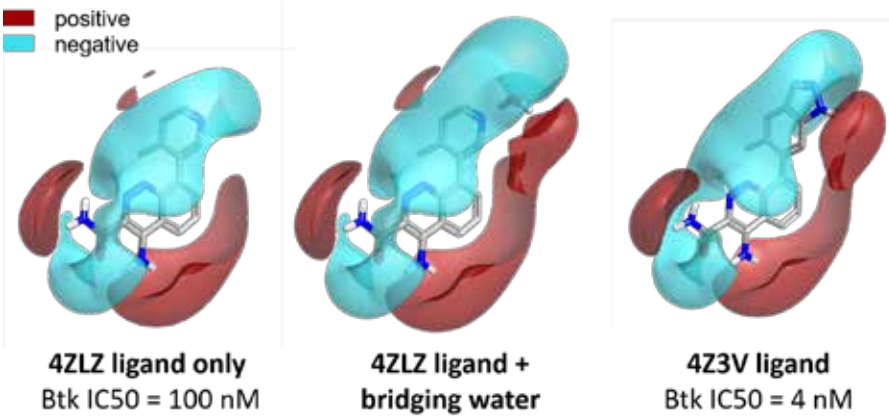

Figure 6: electrostatic interaction potentials calculated for ligands from 4Z3V, 4Z3V including a single water molecule and from 4ZLZ showing greater complementarity when the water molecule is included.

In the absence of 4Z3V structure, it would be natural to ask how displaceable is the bridging water from 4ZLZ? The 3D-RISM water analysis method using the XED force field can answer this question.

Figure 7 shows the results of the 3D-RISM calculation on 4ZLZ. The oxygen density surface clearly shows a region of localised water near the nitrogen of the pyridine, and the 3D-RISM localisation algorithm suggests a water molecule should exist in exactly the spot where it is seen in the crystal structure. The 3D-RISM thermodynamic analysis indicates that this water molecule is neither particularly stable nor particularly unstable. This is consistent with the fact that this water molecule is displaceable (as proven by the 4Z3V ligand and structure), but also indicates that the displacing group needs to have the correct electrostatics and shape to avoid losing affinity.

Protein Electrostatics and Water Prediction Improve Ligand Design

Protein interaction potentials and ligand fields are a powerful way of understanding the electrostatics of ligand-protein interactions. The inclusion of stable water molecules following a 3D-RISM analysis dramatically improves the precision of the method for the characterisation of protein active sites. The information gained from protein interaction potentials can be used to inform ligand design, compare related proteins to identify selectivity opportunities, and understand SAR trends and ligand binding from the protein's perspective.

REFERENCES

1. Comprehensive Medicinal Chemistry III, 2017, Vol. 3, Chp. 11, 253-296. See also citations at http://cresset-group.com/citations/ and https://docs.eyesopen.com/eon/pub.html.
2. J.W. Ponder and C. Wu. "Current Status of the AMOEBA Polarizable Force Field", J Phys. Chem. B. 2010, 114 (8), 2549-2564.
3. J. G. Vinter. "Extended electron distributions applied to the molecular mechanics of some intermolecular interactions", J. Comp. Aided. Mol. Des. 1994, 8, 653-668
4. D. Chandler, J.D. McCoy and S.J. Singer. J. Chem. Phys., 1986, 85, 5971. A. Kovalenko and F. Hirata, J. Chem. Phys., 1999, 110, 10095.
5. http://www.cresset-group.com/flare
6. O.V. Stroganov, F.N. Novikov, A.A. Zeifman, V.S. Stroylov and G.G. Chilov. Proteins 2011, 79(9), 2693-2710.
7. C.R. Smith, D.R. Dougan, M. Komandla, T. Kanouni, B. Knight, J.D. Lawson, M. Sabat, E.R. Taylor, P. Vu and C. Wyrick. J. Med. Chem. 2015, 58, 5437-5444

Tim Cheeseright

Tim has a DPhil in Chemistry from the University of Oxford. He joined Cresset immediately after its founding in 2002. As Director of Products Tim is responsible for Cresset's software. He has been instrumental in developing the easy to use scientific software that characterizes Cresset products today.

Katriona Scoffin

is a science writer working for Cresset.