# Binding Energies of Tyrosine Kinase Inhibitors: error assessment of computational methods for imatinib and nilotinib binding

Clifford W Fong

**Computational Biology and Chemistry, 18 May 2015, doi:10.1016/j.compbiolchem.2015.05.002**
**Computational Biology and Chemistry 58 (2015) 40–54**

# Binding Energies of Tyrosine Kinase Inhibitors: error assessment of computational methods for imatinib and nilotinib binding

**Clifford W. Fong***

Eigenenergy, Adelaide, South Australia, Australia

* Author to whom correspondence should be addressed; E-Mail: cwfong@internode.on.net

---

The binding energies of imatinib and nilotinib to tyrosine kinase have been determined by quantum mechanical (QM) computations, and compared with literature binding energy studies using molecular mechanics (MM). The potential errors in the computational methods include these critical factors:

- Errors in X-ray structures such as structural distortions and steric clashes give unrealistically high van der Waals energies, and erroneous binding energies.
- MM optimisation gives a very different configuration to the QM optimisation for nilotinib, whereas the imatinib ion gives similar configurations
- Solvation energies are a major component of the overall binding energy. The QM based solvent model (PCM/SMD) gives different values from those used in the implicit PBSA solvent MM models. A major error in inhibitor – kinase binding lies in the non-polar solvation terms.
- Solvent transfer free energies and the required empirical solvent accessible surface area factors for nilotinib and imatinib ion to give the transfer free energies have been reverse calculated. These values differ from those used in the MM PBSA studies.
- An intertwined desolvation – conformational binding selectivity process is a balance of thermodynamic desolvation and intramolecular conformational kinetic control.
- The configurational entropies ($T\Delta S$**)** are minor error sources.

**Keywords:** Tyrosine kinase inhibitors; nilotinib; imatinib; quantum mechanics; binding energies; error assessment

---

## 1. Introduction

Protein kinases have been the focus for many drug-based cancer treatments. Tyrosine kinases are enzymes responsible for the activation of many proteins by signal transduction cascades. The proteins are activated by adding a phosphate group to the protein (phosphorylation). Tyrosine kinase inhibitors can compete with adenosine triphosphate (ATP), the phosphorylating entity, the substrate, or both, or can act in an allosteric fashion, by binding to a site outside the active site, affecting its activity by a conformational change. Targeting the tyrosine kinases that regulate cell growth and proliferation has been very productive, as witnessed by the success of inhibitors such as imatinib mesylate, or Gleevec, and nilotinib, or Tasigna, for the treatment of both chronic myeloid leukemia (CML) and gastrointestinal stromal tumors (GIST). BCR-ABL is the oncogenic protein-tyrosine kinase responsible for the pathogenesis of chronic myelogenous leukemia [1, 2]. Development of new tyrosine kinase inhibitors to overcome resistance, or improved efficacy, requires an understanding of the binding efficiency and effectiveness of the inhibitor with the tyrosine kinase. Nilotinib is a second generation inhibitor which is similar to imatinib, but 30 times more effective in treating CML.

The computation of binding free energies between small molecule ligands and proteins is a difficult task since the binding energy usually involves small differences amongst large enthalpies and entropies related to the free and bound states of protein and the ligand. Inter-atomic forces are strong and short ranged, resulting in steep energy functions that are strongly dependent on molecular conformation. Proteins and ligands are usually very flexible, having many degrees of freedom, making conformational energy profiles very dominant contributing factors. Solvation effects can be very large for both free and bound protein and ligand, including protonation and salt effects. Dispersion or van der Waals and hydrophobic effects between ligands and proteins can also be large. It is also conceivable that small configurational changes (including molecular strain) in bond lengths and angles might occur during binding, which might have significant energy impacts [3,4,5,6]. These effects will have large first order effects in any calculation. Second order effects include translational, rotational, vibrational, and repulsive effects, which are known to be smaller.

Enthalpic contributions to binding free energy are driven by the strength and directed specificity of ionic, polar, hydrogen bond, electrostatic (coulombic), van der Waals and polarization interactions. These interactions usually have small entropy contributions. Changes in binding entropy include small configurational translational and rotational processes, somewhat larger conformational processes, or much larger solvent effects (which include desolvation or rearrangement during binding). The enthalpy changes (in vacuo or gas phase) between the protein and ligand before and after binding closely approximates the free energy of binding interaction, less any configurational entropy change, which may be significant. Gilson has suggested that configurational entropy ($T\Delta S$) between the free and bound ligand can be a loss of ca. 25 kcal/mol (vibrational entropy and conformational entropy differences of 24.6 and 1.8 kcal/mol respectively) on amprenavir ligand binding to HIV protease, using a M2 mining minima method [3, equation 7]. The M2 method is based on calculating binding free energies, potential energy wells and solvation free energies, assuming molecules are rigid rotator/harmonic oscillators, and using force field energies. It is not known what errors are involved in all these assumptions, and the use of empirical force field energies which have no electrons, so cannot evaluate excited electronic or delocalised states. Often force field parameters have to be generated for specific inhibitor molecules during an investigation, when using molecular mechanics methods

calibrated for protein molecules. In addition, various conformational states of the studied inhibitor would need specific force field parameters to be generated for the different states for each calculation which might be used in a molecular dynamics simulation [7]. Real large molecules are far from rigid rotators/harmonic oscillators, and it has been shown [8] that anharmonic, not-separable, ro-vibrational states must be considered. The contribution of electronically excited states, which have their own ro-vibrational states, have important roles.

The free energies involved in solvation processes before and after binding can be calculated using the well established PCM/SMD quantum mechanical solvation model to compare solvation energies in vacuo and water [9]. This approach would be more rigorous than the solvent models typically used in molecular mechanics based methods (PBSA, GBSA) used to investigate protein-inhibitor binding, which comprise an electrostatic component and a non-polar component based on an arbitrary solvent accessible surface area (SASA) factor usually somewhere between 5-45 cal/mol/$Å^2$ . Explicit solvent models require empirical interaction potentials between the solvent and solute, and between solvent molecules.

The computational methodologies used [3,4] vary considerably, from docking and scoring techniques to those using advanced force fields, incorporating electronic polarization, ie iteratively adjusting partial atomic charges from quantum mechanical calculations during docking. Molecular mechanics force field calculations can handle large molecular systems, like proteins, but due to the empirical nature of force fields, the neglect of electrons, electron polarisation and charge transfer are not accounted for explicitly. This can limit the accuracy with which interactions are calculated and consequently the free energies obtained. Ideally ab initio quantum chemistry approaches should be used as these explicitly include electrons, however quantum mechanics is not practicable for large proteins.

Calculating the conformational energies vary from docking methods using one conformation to more sophisticated molecular mechanics – molecular dynamics techniques where the free ligand, free protein, and the complexed ligand-protein conformations are simulated using an explicit solvent model. The difference between energy minimised snapshots of free and bound molecules can be found in molecular dynamics calculations. This approach has to account for many possible low energy conformations by seeding the calculations with best guesses. Molecular mechanics force fields may not give the lowest energy conformation, due to their empirical shortcomings and difficulties with parameter optimisation particularly for conformational states[7]. Force field energy minimisations can result in strained torsional angles, as torsion angles are the softest conformational parameters, and have the largest effects on molecular geometries [10]. Molecules which have significant ability to maximise $\pi$ electronic delocalisation over multiple aromatic rings, as found in nilotinib and imatinib, could show large conformational energy differences by using quantum mechanical energy minimisation compared to force field energies.

The conformational structure of the protein binding pocket is a crucial factor which defines inhibitor effectiveness [11,12,13]. Imatinib is known to bind 2400 times less tightly (experimentally measured 4.6 kcal/mol penalty) to the c-SRC form (DFG-in) of tyrosine kinase than to the closely related c-ABL form (DFG-out), even though the X-ray crystal structures of both complexes are very similar. It has been shown that c-SRC can adopt the inactive ABL conformation which gives strong binding, but the free energy between these conformational states for the c-SRC is the dominant factor. The 4.6 kcal/mol difference in binding energies between the c-SRC and c-ABL forms has been shown

to almost the same (4.4.kcal/mol) as the difference between the DGF-out and DFG-in conformations of the c-SRC kinase, suggesting that *conformational selection is the main source of imatinib binding selection* [12].

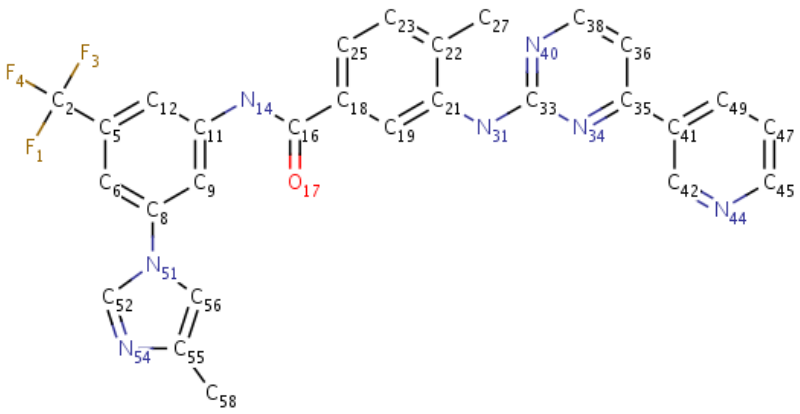
Electrostatic interactions in molecular mechanics force fields are commonly calculated from quantum mechanical Merz Kollman (MK) / RESP partial charges. Comparisons of MK, RESP and CHELPG methods show very similar performances [14]. However Sigfridsson [14] has shown that the MK method showed a significant rotational dependence on the orientation of the grid coordinate system (0.04-0.05 e), whereas CHELPG showed a lesser rotational dependence, as it was specifically designed to remove such dependence [15]. Hence use of MK/RESP charges may induce some errors when calculating electrostatic interactions as inputs into force field calculations.

The starting point for any calculation of binding energies is an X-ray structure of the complexed ligand- protein complex. However it has been shown [16] that published X-ray structures suffer from many problems, including poor fits, such that conformations and configurations (bond lengths, angles etc) can be unrealistic and highly distorted. Amides can be cis or non-planar, and even planar aromatic rings can be deformed. There can be severe steric clashes in the X-ray structure which are not easily apparent. It has been shown [17] that molecular strain can be induced by intermolecular interactions in single-component crystal structures of molecules with no intramolecular hydrogen bonding, resulting in some molecules being distorted by up to 5 kcal/mol by crystal packing forces. As inter-molecular hydrogen bonding is involved in inhibitor – protein binding, large distortions may occur due to crystal packing forces. Literature references often contain statements to the effect that anomalous bond lengths (orders) and angles were "corrected" prior to calculation of binding energies. The other major issue is the need to add hydrogen atoms to X-ray structures, which can also induce errors in molecular geometry. In particular the location of H atoms that may be involved in hydrogen bonding are critical. The geometry of hydrogen bonds in proteins has been analysed. Baker & Hubbard (1984) [18,19] found that 90% of analysed NH -- O hydrogen bonds in the PDB structure data bank have bond angles about $158^{o}$, and most analysed C=O – H hydrogen bonds have angles about $129^{o}$. Morozov (2004) [20] in a later examination of the high resolution X-ray PDB data bank found the average bond angle for single hydrogen bonds was around $110\text{-}120^{o}$. The extent of these errors in the X-ray structures and their effect on calculated binding energies is unknown, and may be large if configurational or molecular geometry distortions are also present.

This study will focus on using established quantum mechanical methods to examine how possible distortions of ligand and protein structure from published X-ray structures can affect binding energy calculations for the tyrosine kinase inhibitors, nilotinib and imatinib, and compares these results from recent molecular mechanics – molecular dynamics studies. An assessment of potential errors in the different computational methods will be made. Accurate calculations of drug – protein binding are essential tools in understanding and predicting how new drugs might behave in vivo. An understanding of the various component energies that contribute to the overall binding energy can allow drug designers to better target desirable features required in potential new drugs.

*2. Experimental Section*

**Figure (Experimental).** Atom labelling scheme for Nilotinib.




**Nilotinib**:4-methyl-*N*-[3-(4-methyl-1*H*-imidazol-1-yl)- 5-(trifluoromethyl)phenyl]-3- [(4-pyridin-3-ylpyrimidin-2-yl) amino]benzamide. **Imatinib**:4-[(4-methylpiperazin-1-yl)methyl]-*N*-(4-methyl-3-{[4-(pyridin-3-yl)pyrimidin-2-yl]amino}phenyl)benzamide. The ion is protonated at the 4-methyl piperazinyl N atom.

All calculations were carried out using the Gaussian 09 package. Electrostatic potential at nuclei for solutions were calculated using the CHELPG method in Gaussian 09. The atomic charges produced by CHELPG are not strongly dependant on basis set selection. Using the B3LYP level of theory, calculated atomic charges were almost invariant amongst the basis sets 6-31G(d), 6-311(d,p), 6-311+(2d,2p), 6-311G++(3df,3dp) [50,51]. Errors between calculated and experimental dipole moments were 3%. A potential weakness of CHELPG (and other methods to calculate electrostatic charges at nuclei from the molecular electrostatic potential, MEP, around the molecule) is the treatment of larger systems, in which some of the innermost atoms are located far away from the points at which the MEP is evaluated. However, this study is concerned with charges at the molecular surface, and how such charges interact with solvents, or other atomic charges on molecules near the surface of the molecules. Test cases to probe for anomalous charges on "buried" atoms were conducted by removing "outer" non-hydrogen bonding residues and rerunning calculations where just hydrogen bonding residues interacted with the ligand, such that the atoms involved directly in hydrogen bonding were more exposed to the MEP grid. One example was found where the Thr315**O**H – **H**$N_{31}$(Ar)(Pyrim) hydrogen bond energies are artificially low, as a result of the bond being "buried" away from the MEP grid used in the CHELPG algorithm. *Finally high absolute computational accuracy is not the objective of this study, comparative differences amongst physical and chemical interactions, particularly in solution, are the foci of the study.*

Merz et al [52] have shown that for small molecules commonly found in large biological molecules such as proteins, that atomic and molecular properties (such as bond lengths and angles, ionization potentials, electron affinities, heats of formation and hydrogen bond interaction energies) modelled by DFT using 6-31G* / 6-31G** and the correlation consistent basis sets aug-cc-pVxZ showed very similar results. Studies of the accuracy of DFT functionals on biological hydrogen bonded systems have shown that B3LYP compares favourably with the MP2 functional [53,54], and is comparable with other DFT functionals, though may slightly underestimate bond energies [55,56], but gives the

best performance in calculating *relative* hydrogen energies. In this study, hydrogen bond energies are calculated from the electrostatic atomic charges, not by DFT energies.

All solvent calculations were at the B3LYP/6-31G*(6d, 7f) level of theory, using optimised geometries, as this level has been shown to give accurate electrostatic atomic charges, and was used to optimize the IEFPCM/SMD solvent model. Where a solvent study was carried to compare different solvents, the same optimised solute geometry was used. With the 6-31G* basis set, the SMD model achieves mean unsigned errors of 0.6 - 1.0 kcal/mol in the solvation free energies of tested neutrals and mean unsigned errors of 4 kcal/mol on average for ions [9]. It has been found that the B3LYP / 6-31G+* combination gives reasonably accurate PCM and SMD solvation energies for some highly polar polyfunctional molecules, which are not further improved using higher level basis sets [57].

Adding diffuse functions to the 6-31G* basis set (ie 6-31$^{+}$*) had no significant effect on the solvation energies with a difference of ca 1% observed, which is within the literature error range for the IEFPCM/SMD solvent model.

Rizzo et al [58] have also used the 6-31G* basis set with CHELPG charges (compared with 7 other atomic charge models) to calculate absolute free energies of solvation and compare these data with experimental results for more than 500 neutral and charged compounds. The calculated values were in good agreement with experimental results across a wide range of compounds. Sigfriddson has also compared CHELPG and MK electrostatic charges using the 6-31G* basis set, and found similar results with both methods, but the *MK charges showed an appreciable rotational dependence on the orientation (within the co-ordinate system) of the molecule*, which was much smaller in the CHELPG method. Charges calculated using the B3LYP and MP2 functionals showed comparable results. Kubelka [51] and Martin [50] have shown that there is only a small dependence of electrostatic charges on the size of the basis sets (from 6-31G* to aug-cc-pV6Z) other than for very small basis sets (eg 3-21G, and to small extent 6-31G). The correlated methods B3LYP. MP2 and QCISD gave comparable performance.

Other recent studies [59-62] have shown that weak non-covalent interactions, including those involving anionic halogens, are only moderately affected by the addition of diffuse functions. The use of polarised correlation-consistent basis sets eg cc-pV*X*Z (*X* = D, T, etc.) can improve the accuracy of grid based electrostatic potential methods such as CHELPG and MK [50]. Roy [26] has shown that the B3LYP functional can account for dispersion and induction interactions in intermolecular complexes as well as functionals (eg MP2) designed to treat dispersion.

Generally for the calculation of hydrogen and polar bond binding energies in this study, there were only small differences between the CHELPG electrostatic atomic charges with the 6-31G*, 6-31G** and the correlation-consistent cc-pVDZ basis sets. An investigation of the addition of diffuse functions viz aug-cc-pVDZ on calculated hydrogen and polar bond interactions (using the CHELPG atomic charges) showed that the side by side comparisons of hydrogen bond and polar bond energies using the diffuse aug-cc-pVDZ and cc-pVDZ basis sets were very similar, despite the overall molecular energy being lower for the aug-cc-pVDZ case as expected. This was demonstrated by an ONIOM calculation with the low level layer using the cc-pVDZ basis set, and the high level layer for the directly interacting group atoms of the hydrogen bond (eg C=O --- HN, NH --- $^{-}$OC(O), NH --- N(pyridinyl), etc) using the aug-cc-pVDZ basis set. This finding is consistent with previous studies that diffuse functions are not always necessary for (anionic) molecules which have a high degree of resonance delocalization, nor for weak non-covalent interactions such as hydrogen and polar bonds. It has been

found [59] that diffuse functions had a negligible effect on energy, geometry and charges for anions where conjugation or delocalisation of the negative charge was occurring. The inhibitors in this study have high degrees of aromatic conjugation, and the carboxylate anion of the glutamic acid 286 residue which engages in hydrogen bonding with the amide $N_{14}\underline{H}$ of nilotinib has a delocalised carboxylate structure (with both C-O bonds having almost identical bond lengths in the complexed state in the X-ray structures).

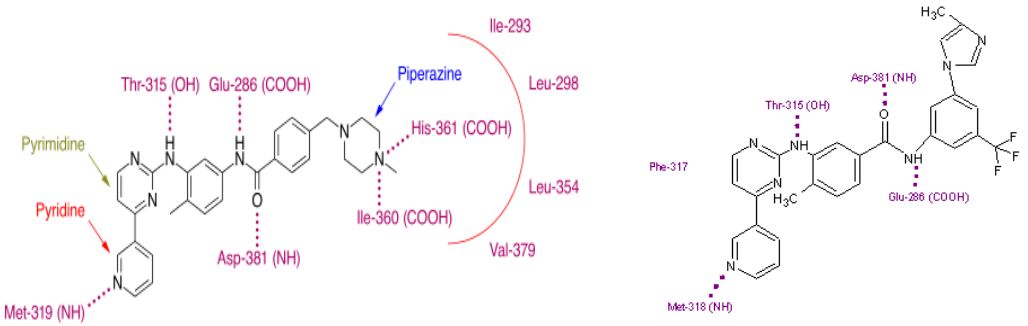
Dougherty [63] has shown that π interactions involving the π electron cloud of an aromatic ring with a interacting molecule involve coulombic interactions caused by the H—C dipoles of the aromatic ring which create the π electron cloud over the ring. The π interactions were estimated by averaging the pairs of attractive coulombic interactions between the carbon and hydrogen atoms of tyrosine 253 and the $C_{30}$ and its hydrogen atom of the pyrimidine ring. The pair interactions between C-H of phenylalanine 382 and the $N_{40}$ of the pyrimidine ring of nilotinib were averaged.

Several tests for any basis set superposition error involving nilotinib and a hydrogen bonded residues (tyrosine, methionine at the X-ray determined distances) revealed no significant counterpoise errors at the B3LYP/cc-pVDZ level of theory in the gas phase.

All inhibitors used in this study were downloaded from the PDB. The nilotinib-tyrosine kinase X-ray structure 3CS9 and the imatinib-tyrosine kinase X-ray structures (1IEP and 2HYY) were used in the binding studies, with all residues too far away (> 7-8 Å) from nilotinib to hydrogen/polar bond, or π stacked, being removed, along with all water molecules [21,64]. To test whether the binding interactions were subjected to any "buried" effects where inner atoms were shielded from the MEP grid used by CHELPG, a layered approach was used to test for consistency, using at first all amino acid residues surrounding to nilotinib or imatinib, (ie residues directly hydrogen bonded to the ligand and other residues which are within van der Waals distance of the ligand) then removing residues that were not hydrogen bonded to the ligand. The electrostatic hydrogen bonding energy calculations did not show major discrepancies with this approach. Molecular mechanics optimisations were carried out using the Gaussview force field with empirical parameters kept constant for all optimisations.

3. Results and Discussion

**Figure 1.** Nilotinib (right) and Imatinib (left) showing similarities in molecular structure and the common hydrogen or polar bonding interactions with the BCR ABL oncogene of tyrosine kinase (neglecting $CF_3$ polar interactions of nilotinib, common hydrophobic interactions, noting that imatinib exists as the cation at physiological pH levels: see text)




The X-ray structures of the nilotinib and imatinib complexes with the Abl protein show subtle differences in the mode of binding of nilotinib to Abl and a better topological fit to the Abl protein could account for the greater potency of the nilotinib. Like imatinib, nilotinib binds to the inactive conformation of the Abl tyrosine kinase, with P-loop folding over the ATP-binding site, and the activation-loop blocking the substrate binding site, to disrupt the ATP-phosphate-binding site and inhibit the catalytic activity of the enzyme [21].

Figure 1 shows the strong similarities [21] in molecular structure for both nilotinib and imatinib, with the trifluoromethyl-imidazole substituting for the piperazine ring, as well as the common hydrogen bonds to the Met318, Thr315, Glu286 and Asp381 of the *inactive* conformation of the Abl kinase domain. However it is shown in this study that the strongly polar $CF_3$ group on the imidazole ring of nilotinib can form strong polar bonds with Asp381, Ile293, Val379 and His361. Both inhibitors also engage in polar bonding to Met290 as well. Both inhibitors engage in π-π stacking interactions with Phe317, Phe382 and Tyr253 as well as other hydrophobic interactions which strongly influence the binding affinity between the inhibitors and the tyrosine kinase in addition to the directional specificity provided by the hydrogen and polar bonding. The end N of the piperazine ring of imatinib is protonated at physiological pH=7.4 levels in blood serum, such that the cation is the dominant species. It is clear that both inhibitors have significant degrees of freedom which would confer large variability in the energies of possible conformers, especially so for imatinib which has a flexible piperazine ring. The highly polar groups on both molecules would be expected to have large solvation energies in water.

*3.1. X-ray crystal structures*

The X-ray crystal structure of nilotinib (AMN-107) complexed with human ABL tyrosine kinase has been subjected to several interpretations, which vary according to the likelihood that hydrogen bonds and polar bonds can form between various sites on nilotinib and the appropriate amino acid residues on the enzyme. The interpretation is made more difficult by the uncertainty of the location of the H atoms in the structure, and by the fact that nilotinib can complex with 4 chains of the enzyme. A critical interpretative factor is the cut-off distances defining bonding interactions, with hydrogen bonds

typically range from 2.6 to 3.5 Å separating the non-H atoms, van der Waals interactions typically ranging from 2.5 to 4.6 Å, and π stacking interactions typically averaging 3.6 Å, but being less than 5.5 Å, and electrostatic or salt bridges (for example involving residues such as glutamic and aspartic acids which are usually negatively charged at blood serum pH 7.4) typically averaging around 3 Å.

The Protein Data Bank Europe X-ray structure [21] 3CS9 shows that 4 nilotinib molecules are complexed in protein chains A, B, C, D with SASA of 803, 809, 812, 805 $Å^2$ with 31, 29, 28, 26 interfacing amino acid residues, with enzyme interface areas of 576, 556, 572, 513 $Å^2$, with 6, 6, 5, 5 hydrogen bonds. The ligand nilotinib in chain A forms 6 hydrogen bonds, a SASA of 803 $Å^2$ and BSA (buried surface area) of 742 $Å^2$.

In a later X-ray study of the same complex [22], it was shown that the $CF_3$ group did engage in some polar bonding, forming one interaction orthogonal to an amide (Val379) besides two further contacts to the imidazole NH of a histidine residue (His361) as well as an isoleucine side chain (Ile293). This interpretation is different from the interactions identified by the PDB PISA analysis, which identified the interactions between Asp381 and Val379 and the trifluoromethyl group [6,23]. It is clear that the $CF_3$ group has 4 polar interactions with the amino acid side chains of the enzyme. It was noted that the $CF_3$ derivative is over 5-fold more active than the corresponding methyl derivative in an autophosphorylation assay, illustrating the importance of these polar interactions. An Atomistry analysis [Atomistry,64] of all residues within a 5 Å coordination sphere of the F atom showed A: ***Ile***293, A: ***Leu***298, A: ***Val***299, A: ***Val***379, A: ***Ala***380, A: ***Asp***381, A: ***Nil***600 fell within the sphere for chain A.

The Protein Data Bank Europe X-ray structure [21,24] 2HYY shows that 4 imatinib molecules (STI-571) are complexed in protein chains A, B, C, D with SASA of 793, 789, 795, 793 $Å^2$ with 27, 27, 29, 27 interfacing amino acid residues, with enzyme interface areas of 570, 567, 566, 563 $Å^2$, with 4, 4, 5, 4 hydrogen bonds. The ligand imatinib in chain A forms 4 hydrogen bonds, a SASA of 793 $Å^2$ and BSA (buried surface area) of 718 $Å^2$. These data are very similar to the X-ray data for nilotinib (vide supra). However, the imatinib is protonated in solution and in the crystal form, and clearly has polar interactions arising from the quaternary $N^+$ of the piperazine ring, which nilotinib does not possess.

It is apparent that even with high quality X-ray structure being available, identifying all possible polar and non-polar interactions (besides the hydrogen bonds, which are easily identified except for the exact position of the H atoms) is difficult and requires interpretation. The uncertain location of H atoms, missing residues, or parts of residues in the structure, and often several residues in close proximity to the ligand can mean that multiple non-covalent bonding interactions are possible based on arbitrary cut-off distances for bonding types such as hydrogen bonds, polar bonds, π stacking interactions etc. Determining the numerous non-polar interactions between the ligand and protein from just the X-ray structure is very difficult, particularly when the ligand has multiple aromatic rings which can engage in π stacking interactions with protein residues. Also hydrophobic interactions can be dominated by entropy effects rather than enthalpy effects. Generally, the tighter and more directed interactions (eg hydrogen and polar bonds) the less entropic and more enthalpic driven is the interaction [6,24,25].

Two X-ray structures of the imatinib – tyrosine kinase complex were downloaded from the PDB data base: 1IEP [38] and 2HYY [24]. Asaki [39] identified a hydrophobic pocket comprised of Ile293,

Leu298, Leu354, Val379 surrounding the protonated $NH^+$ of the piperazine moiety of imatinib. The nilotinib – tyrosine kinase structure 3CS9 was also downloaded from the PDB data base.

The *quality* of published X-ray structures in the PDB data base can be revealed from the *validation reports*. For example using the inhibitor ligand as a reference point: a bond length (or angle) with |Z| > 2 is considered an *outlier* worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles). For 1IEP, 2HYY and 3CS9 the #|Z| > 2 scores for bond lengths were 51, 19, 6% and for bond angles 12, 26, 12% respectively. The RMSZ values for bond lengths were 2.34, 1.40, 1.02 and for bond angles were 2.11, 2.08, 1.37 respectively. These data would indicate that the 1IEP structure is of a considerably lower quality than either the 2HYY or 3CS9 structures. The imatinib ligand in the 2HYY structure has more artificial bond length anomalies than does nilotinib in the 3CS9 structure, but that situation is reversed for bond angles. It would be expected that bond length distortions would have greater impacts on calculated quantum molecular energies than bond angles from first principles.

It is clear from a close examination of the free inhibitors (ie where the kinase residues and all water molecules have been removed from the PDB structure, see discussion 2. below) that the 1IEP imatinib structure is very distorted, with the pyrimidine and some aromatic ring being clearly distorted. The 2HYY and 3CS9 structures are far less distorted on inspection.

### *3.2. Calculation of binding energies, $\underline{\Delta G_{bind}}$*

A simple model to estimate inhibitor-kinase binding free energies is:

$$\mathbf{\Delta E_{int} = \Delta E_{complex} - \Delta E_{tyrkin} - \Delta E_{inhib} = \Delta E_{elect} + \Delta E_{vdW}} \quad \mathbf{(1)}$$

$$\mathbf{\Delta G_{bind} = \Delta E_{int} + \Delta G_{desolv\text{-}inhib} + \Delta G_{desolv\text{-}tyrkin} - \{T\Delta S_{free} - T\Delta S_{bound}\}_{inhib} - \{T\Delta S_{free} - T\Delta S_{bound}\}_{tk}} \quad \mathbf{(2)}$$

Quantum mechanical calculations of internal electronic energy (lowest point of the potential energy well) are corrected for the zero point energy to give the system energy at 0K. Vibrational, rotational, translational energies can also be calculated at 0K. The enthalpy can be calculated by applying the system temperature (usually 298K) and the gas constant to the energy at 0K, and free energy is calculated by applying the temperature and entropy to the enthalpy. For large molecules, the vibrational, rotational, translational energies, as well as the entropy, are usually small compared to the electronic energy and enthalpy. The $\mathbf{\Delta E_{int}}$ term is usually calculated in vacuo from the differences in internal potential energies of the complex, inhibitor and kinase. The $\mathbf{\Delta G_{desolv\text{-}inhib}}$ and $\mathbf{\Delta G_{desolv\text{-}tyrkin}}$ terms are calculated using the PCM-SMD solvation model, and are the differences between the energies in vacuo and water. The terms $\mathbf{T\Delta S_{free}}$ and $\mathbf{T\Delta S_{bound}}$ are the calculated configurational entropies for the free and bound states of the inhibitor and tyrosine kinase at 298K in water. All quantum mechanical calculations are only meaningful for differences between calculated energies using the same functional and basis sets.

Using molecular mechanics with arbitrary predefined atomic and dielectric parameters and charge optimisation can treat the whole ligand-protein complex, but the effect of empirical assumptions is unknown. Alternately the use of rigorous quantum mechanical methods is not feasible for such large systems. *The approach taken in this study to examine how possible distortions and other errors affect the calculation of binding interaction energies ($\Delta E_{int}$) between the inhibitors nilotinib and imatinib and tyrosine kinase. <u>First</u> the PDB X-ray structures is modified (after adding any missing H atoms) by*

*removing all protein residues (and water molecules) that are not immediately surrounding the bound inhibitor molecules, then calculating the energy of the complex (**$\Delta E_{complex}$**).* *Second, the inhibitor is then removed from the complex to leave only the kinase residues in exactly the same configuration as in the X-ray structure (**$\Delta G_{tyrkin}$**). Third all the kinase residues are removed to leave only the inhibitor in exactly the same configuration as in the X-ray structure (**$\Delta E_{inhib}$**).* This reduces the calculational load to a manageable level, but assumes that only short range residue interactions immediately around the inhibitor are the major influence on binding energies. *The binding interaction energy between nilotinib and imatinib is then the energy difference between the complexed inhibitor from the first step, minus the sum of the energies from steps two and three*. The level of theory chosen is the B3LYP/6-31G*, which has been shown to be the minimal level of theory to give good accuracy in biological systems. Importantly, as differences in calculated energy levels are involved, possible deficiencies in functional accuracy are substantially cancelled. This level of theory is also the same level of theory used in the SMD solvent model used in this study, and has been shown to give accurate results for the CHELPG method used to calculate electrostatic atomic charges used in determining coulombic intermolecular interactions between the inhibitor and the kinase. It will be shown that level of energy changes involved in molecular distortion and solvation are far higher than any possible gains that might be made using higher level functionals or larger basis sets. Some tests were done to gauge the effect of increased diffuse basis sets (6-31G**) and correlation consistent basis sets (cc-pVDX, and aug-cc-pVDX) on the hydrogen or polar interactions between inhibitor and kinase, with only very small changes being apparent in CHELPG atomic charges. Similarly energy differences between the molecular species using 6-31G* and cc-pVDX basis sets were very similar (see Experimental section).

Hydrogen bonding and polar (electrostatic) interactions $\mathbf{\Delta E_{elect}}$ (or $\mathbf{\Delta E_{hydrophilic}}$ in water) were calculated from the coulombic CHELPG charge interaction between pairs of involved atoms. This allows a direct identification of atoms at the X-ray structure bonding distances which have electrostatic charges which can form ***attractive*** coulombic bonding interactions. The various interactions are shown in Table 1 and 2. The van der Waals interactions in vacuo, $\mathbf{\Delta E_{vdW}}$, are then calculated from equation (2), where "van der Waals interactions" are defined as a global term including all forces that are not hydrogen bonding or other polar interactions. Calculating $\mathbf{\Delta E_{vdW}}$ is difficult, since they include multiple anisotropic forces [26], and solvents have large attenuating effects compared to those in vacuo. All water molecules in the X-ray structures were removed, since it is unknown how many, if any, water molecules dynamically interact with the tyrosine kinase, inhibitor, or tyrosine kinase-inhibitor complex. The errors potentially involved in the treatment of tyrosine kinase, ie treating only those residues immediately surrounding the inhibitor in the complex, removal of all water molecules that were present in the X-ray structures are difficult to evaluate, as they are in all protein-ligand binding studies. The final arbiter of the any computational binding analysis must be evaluation against experimental binding data.

**Table 1. Comparison of binding energies for imatinib and nilotinib with ABL tyrosine kinase with literature data.**

| | **Imatinib Ion –TK** 13 residues | **Imatinib Ion -TK** 11 inner + 9 outer residues | **Imatinib Ion –TK** 11 inner residues | **Nilotinib-TK** 13 residues | **Imatinib-TK (Nilotinib-TK)** Dubey 2011 (Dubey 2010) | **Imatinib-TK** Pricl 2005 | **Imatinib-TK** Lin 2010 | **Imatinib-TK** Alexsandrano v 2010 |
|---|---|---|---|---|---|---|---|---|
| $\Delta E_{int}$ | -332.5 | -397.3 | -214.3 | -170.2 | -90.4 to -101.4 (-74 Nilotinib) | -128.8 | -27.7 | -151.9 |
| $\Delta E_{elect}$ | -39.3 | -50.4 | -50.4 | -44.8 | -18.2 to -26.5 | -57.4 | 0.5 (1.2*) | -126.1 |
| $\Delta E_{vdW}$ | -293.2 | -346.9 | -163.9 | -125.4 | -72.2 to -74.9 | -71.4 | -29.4 | -25.8 |
| $\Delta G_{desolv\text{-}inhib}$ | 87.0 <CDS 2.6> {74.9}# {<CDS 7.7>} | | 87.2 <CDS 2.6> {77.3}# {<CDS 6.2>}# | 18.5 <CDS 10.4> | | | | |
| $\Delta G_{desolv\text{-}tyrkin}$ | 211.5 <CDS 32.6> | | 126.2 <CDS -30.5> | 162.5 <CDS -27.9> | 51.4–61.4 *** <$\Delta G_{NP}$ -8.7 - -8.6> | 101.3*** < $\Delta G_{NP}$ -6.7> | | |
| $\{T\Delta S_{free} - T\Delta S_{bound}\}_{inhib}$ | 9.0 water | | 8.0 water (14.1 gas) | 14.7 water (16.1 gas) | 61.4 - 65.5 | 17.1 | 11.3 | |
| $\{T\Delta S_{free} - T\Delta S_{bound}\}_{tk}$ | 7.8 water | | 2.0 water | 6.7 water | | | 1.4 | |
| $\{\varepsilon_{total\text{-}free} - \varepsilon_{total\text{-}bound}\}_{inhib}$ | 3.0 | | 5.5 (7.9 gas) | 3.2 (5.6 gas) | | | (5.6**) | |
| $\{\varepsilon_{total\text{-}free} - \varepsilon_{total\text{-}bound}\}_{tk}$ | -1.0 water | | 7.3 water | 12.0 water | | | | |
| $\Delta G_{bind}$ | -50.8 | | -10.9 {-20.8} | -10.6 | -10.4 | -10.4 | -9.4 | -15.3 |
| PDB X-ray Structure Ref | 1IEP | 2HYY | 2HYY | 3CS9 | 2HYY | 1IEP | 1IEP | 2HYY |

**Footnotes:**

Crystal structures (2HYY): S.W. Cowan-Jacob, G.F. Floersheimer et al, Acta Crystallogr., 2007, D63, 80. C. Bissantz, B. Kuhn, M. Stahl, J. Med. Chem., 2010, 53, 5061.

Crystal structure (3CS9): E. Weisberg, P.W. Manley et al, Cancer Cell 2005, 7, 129.

Crystal structure (1IEP): B. Nagar, W. Bornmann et al, Cancer Res. 2002, 62**,** 4236.

Protein Data Bank Europe PISA X-ray structure 3CS9, 2HYY, 1IEP.

Hydrophobic pocket: comprised of Ile293, Leu298, Leu354, Val379 surrounding the protonated $NH^+$ of the piperazine moiety of imatinib: T. Asaki, Y. Sugiyama, T. Hamamoto, M. Higashioka,M. Umehara, H. Naito, T. Niwa, Bioorganic & Medicinal Chemistry Letters 2006,**16**,1421.

A. Aleksandrov, T. Simonson, J. Biological Chem. 2010, 285, 13807.

Y. Lin, M. Yilin, J. Wei, B. Roux, www.pnas.org/cgi/doi/10.1073/pnas.1214330110 (repulsive energy*, translational/rotational energy**)

D.K. Dubey, P.R. Ojha, J. Biol. Phys. 2011, 37, 69.

D.K. Dubey, K.A. Chaubey, A. Parveen, P.R. Ojha, J. Biophys. Struct. Biol, 2010, 2, 47.

S. Pricl, M. Fermeglia, M. Ferrone, E. Tamborini, Mol. Cancer Ther., 2005, 4, 1167.

$\mathbf{\Delta G_{bind} = \Delta G_{int} + \Delta G_{desolv\text{-}inhib} + \Delta G_{desolv\text{-}tyrkin} - \{T\Delta S_{free} - T\Delta S_{bound}\}_{inhib}}$ -- $\mathbf{\{T\Delta S_{free} - T\Delta S_{bound}\}_{tk}}$

$\mathbf{\Delta G_{bind}}$ = -170.2 +18.5+162.5–21.4 = -10.6 kcal/mol Nilotinib

$\mathbf{\Delta G_{bind}}$ = -214.3 +87.2+126.2–10.0 = -10.9 kcal/mol Imatinib ion for bound state 2HYY

$\mathbf{\Delta G_{bind}}$ = -214.3 +77.3+126.2–10.0 = -20.8 kcal/mol Imatinib ion for QM optimised state 2HYY

$\mathbf{\Delta G_{bind}}$ = -332.5 +87.0+211.5–16.8 = -50.8 kcal/mol Imatinib ion for bound state 1IEP. The value of -50.8 kcal/mol indicates that the 1IEP structure is so distorted that realistic values of $\mathbf{\Delta G_{bind}}$ cannot be derived.

Experimental binding values for imatinib-tyrosine kinase and nilotinib-tyrosine kinase complexes are -10.4 and -10.6 kcal/mol based on reported $IC_{50}$ values. (Pricl, Corbin [48], Manley 2005)

The terms $\mathbf{\{\varepsilon_{total\text{-}free}\text{-}\varepsilon_{total\text{-}bound}\}_{inhib}}$ and $\mathbf{\{\varepsilon_{total\text{-}free}\text{-}\varepsilon_{total\text{-}bound}\}_{tk}}$ are included in the $\mathbf{\Delta E_{int}}$ term and are shown for reference purposes only.

*** $\mathbf{\Delta G_{solv} = \Delta G_{PB} + \Delta G_{NP}}$ = free energy of complex solvation using MM/PBSA in literature references, Dubey, Pricl

# see text for discussion

**Table 2. Hydrogen bonding, π-π and polar bonding for Imatinib (Gleevec 2HYY), Imatinib (1IEP) and Nilotinib (3CS9) complexes with Tyrosine Kinase**

| Hydrogen or polar bonds | 2HYY Imatinib-TK | | 1IEP Imatinib-TK | | 3CS9 Nilotinib-TK | |
|---|---|---|---|---|---|---|
| | Bond Lengths | Bond Energies | Bond Lengths | Bond Energies | Bond Lengths | Bond Energies |
| Met290**S**-**H**N (CO) H bond | 3.4 (3.4) | 3.5 [3.2] | 3.3 (2.2) | 6.4 | 3.4 (2.8) | 6.9 |
| Glu286C-**O**$^{--}$ - **H**N(CO) H bond | 2.8 (1.9) | 11.5 [11.1] | 3.0 (1.95) | 10.5 | 3.0 (2.2) | 11.6 |
| Ile360C=**O** - **HN**$^{+}$Piperazine H Bond | 3.1 (2.35) | 3.8 [2.0] | 3.1 (2.75) | 0.8 | na | |
| His361C= **O** - H**N**$^{+}$Piperazine H Bond | 2.7 (2.75) | 4.5 [5.9] | 2.7 (2.85) | 3.6 | na | |
| Asp381N**H** - **O**C(NH) H bond | 3.0 (2.3) | 8.5 [8.0] | 2.9 (1.7) | 7.6 | 3.1 (2.6) | 5.9 |
| *Arg362 - HN$^{+}$Piperazine CH-π Hydrophobic* | | | | | na | |
| *Gly321 - H**N**$^{+}$Piperazine CH-π Hydrophobic* | | | | | na | |
| *Val289 - NH$^{+}$Piperazine CH-π Hydrophobic* | | | | | na | |
| *Ile313 - Phenyl CH-π Hydrophobic* | | | | | | |
| *Leu248 - Pyridine CH-π Hydrophobic* | | | | | | |
| Ala269C=**O** - **H**C$_{27}$Ph H Bond | 3.9 | 0.1 [0.1] | | | 3.4 | 0.0 |
| Met318N**H** - **N** Pyridinyl H Bond | 2.95 (2.35) | 6.2 [6.4] | 2.9 (1.75) | 0.8 | 3.0 (2.35) | 5.5 |
| Thr315**O**H - **H**N(Ar)(Pyrim) H bond # | 3.1 (1.7) | 8.0 | 2.9 (1.75) | 8.0 | 2.9 (1.75) | 8.0 |
| Phe317 - **N** Pyridinyl π-π interaction | 3.3 | 0.8 [0.7] | 3.3 | 0.4 | 3.7 | 0.8 |
| Tyr253 - **C** Pyrimidinyl π-π interaction | 3.7 | 1.5 [1.4] | 3.6 | 0.8 | 3.7 | 1.2 |
| Phe382 - **N** Pyrimidinyl π-π interaction | 3.4 | 2.0 [1.8] | 3.5 | 0.4 | 3.4 | 1.4 |
| *Ile293 - NH$^{+}$Piperazine Hydrophobic Pocket** | | | | | na | |
| *Leu298 - NH$^{+}$Piperazine Hydrophobic Pocket** | | | | | na | |
| *Leu354 - NH$^{+}$Piperazine Hydrophobic Pocket** | | | | | na | |
| *Val379 - NH$^{+}$Piperazine Hydrophobic Pocket** | | | | | na | |
| His361N**H** - **F**C(F$_2$) polar bond | na | | na | | 3.4 (2.9) | 1.1 |
| Val(O=)**C**CNH$_2$ - **F**C(F$_2$) polar bond | na | | na | | 4.0 | 2.6 |
| Ile**CH**-**F**C(F$_2$) polar bond | na | | na | | 3.4 (2.9) | 0.2 |
| Asp381**NH**-**F**C(F$_2$) polar bond | na | | na | | 3.5 (3.0) | 1.3 |
| | | 50.4 [53.0] | | 39.3 | | 44.8 |

**Footnotes:**

Crystal structures (2HYY): S.W. Cowan-Jacob, G.F. Floersheimer et al, Acta Crystallogr., 2007, D63, 80. C. Bissantz, B. Kuhn, M. Stahl, J. Med. Chem., 2010, 53, 5061.

Crystal structure (3CS9): E. Weisberg, P.W. Manley et al, Cancer Cell 2005, 7, 129.

Crystal structure (1IEP): B. Nagar, W. Bornmann et al, Cancer Res. 2002, 62, 4236.

Protein Data Bank Europe PISA X-ray structure 3CS9, 2HYY, 1IEP.

Bond lengths Å are the X-ray structure values. Values in (..) are the distances between the hydrogen or polar bonded atoms which are shown in column 1 as underlined atoms.

Bond energies are in kcal/mol calculated from coulombic atomic charges.

Hydrophobic pocket: Asaki 2006 identified a hydrophobic pocket comprised of Ile293, Leu298, Leu354, Val379 surrounding the protonated $NH^+$ of the piperazine moiety of imatinib

Values in [..] calculated from 2HYY complex with 9 outer hydrophobic kinase residues (shown italicized) removed

# Thr315**O**H - **H**N(Ar)(Pyrim) H bond: These bond energies are artificially low, as a result of the bond being "buried" away from the MEP grid used in the CHELPG algorithm: the "corrected" value is ~8.0 kcal/mol (see Experimental)

Table 5 has a side by side comparison of tyrosine kinase residues used in the calculations.

na = not applicable

*3.3. Conformational and configurational effects on binding energies*

An estimate of the conformational energies involved when the small ligand inhibitors nilotinib and imatinib move from the free solvated state to the complexed state in the tyrosine kinase binding pocket can be made by assuming the *optimised* lowest energy state for the free molecule is the preferred conformation in solution. The difference in configurational entropies of the free and complexed (bound) states of the inhibitor is $\{T\Delta S_{free} - T\Delta S_{bound}\}_{inhib}$ which represents the conformationally induced entropy change for the inhibitor as it moves from a free unconstrained solvated state to a complexed state. It is clear that both nilotinib and imatinib have many conformers when free, but the lowest energy state would be the dominant conformer. In this study this state is taken to be the quantum mechanically (QM) optimised inhibitor. Comparisons were made with a basic molecular mechanics (MM) optimisation method, to give some insights into the potential deficiencies of using MM based approaches to calculate protein-inhibitor binding energies.

The optimisation of the free inhibitors would also involve some configurational (bond lengths and angles) as well as pure conformational changes, if the starting structure from the X-ray structure is artificially distorted from the norm, as would appear to be the case for the three structures in this study to varying degrees (see 2(1) X-ray structures above).

It is much more difficult to estimate what conformational changes occur with the tyrosine kinase, since many factors are involved in proteins conformations, both short range and long range. Various molecular mechanics (plus docking, molecular dynamics) approaches have been used, but is well known that such calculations do not accurately calculate energies. In this study, an estimate of changes in kinase binding pocket conformation is made by assuming that only close range effects immediately surrounding the ligand are operating (by removing all distant residues and water other than those bound to the kinase by hydrogen or polar bonds, or in close proximity to the bound inhibitor, see Table 5 for a comparative list of residues), and then MM optimising these residues, and assuming this *"free" state* is the lowest energy conformation for those kinase residues. The *bound kinase state* is that taken from the X-ray structure after removing distant residues without any optimisation. The configurational entropy difference between the two values is then $\{T\Delta S_{free} - T\Delta S_{bound}\}_{tk}$. *Clearly this is a crude approximation, which can only give a rough estimate of the entropy changes involving the kinase receptor during binding*. A QM optimisation of the tyrosine kinase is not computationally practicable and in addition there are many possible configurations for the tyrosine kinase in solution which defy easy identification. The error in calculating $\{T\Delta S_{free} - T\Delta S_{bound}\}_{tk}$ is clearly much larger than in the calculation of $\{T\Delta S_{free} - T\Delta S_{bound}\}_{inhib}$.

It has been established [3,27-31] that conformational changes in ligand and protein before and during ligand-protein binding can involve large changes in conformational energy and configurational entropy, which can cancel much of the energy change that drives binding, and must be included when calculating binding affinities. If the ligand adopts a strained conformation in the bound state, the rise of its internal energy on binding will make a positive contribution to and thus oppose binding. Alternatively, if the ligand accesses a wide range of equally stable conformations in the free state, but only a narrow range of conformations in the bound state, its configurational energy will drop and thus oppose binding. Finally, if the bound conformation of the ligand corresponds to an especially stable conformation of the free ligand, then the ligand is preorganized for binding, and both the strain and energy penalties will be low, and vice versa. The same considerations apply to the protein, which is an

example of conformational selection [28] where the conformational ligand-protein ensemble is stabilized more than the separate ligand or protein binding site. A loss (or gain) in conformational energy cancels most of (or adds to) the binding energy, and the net binding free energy is a small difference between two larger numbers.

Perola has shown that ligands rarely bind in their lowest conformation configuration, and conformational strain energies in 60% of cases examined were less than 5 kcal/mol, and in 10% of cases energies over 9 kcal/mol were found. Chen and Foloppe [29] examined a wider set of data, and found similar results, with the "global conformational energy window" being within 15 kcal/mol. GÅ [31]examined a diverse range of proteins, and large number of ligands (233), and found that the conformational free energy change averaged 5.8 kcal/mol, with a maximum of 33.2 kcal/mol. Entropic and enthalpic contributions were significant. There have been criticisms about the magnitude of ligand conformational energy changes during protein binding [32], based on errors induced by force field calculations being inadequate for calculating energies, and errors in using X-ray data which inaccurately portray conformations, distort ring structures, induce steric clashes etc [16].

A correlation was found between acceptable strain energy and ligand flexibility, while there was no correlation between strain energy and binding affinity, thus indicating that expensive conformational rearrangements can be tolerated in some cases without overly penalizing the tightness of binding. The unfolding of hydrophobic ligands during binding, which exposes hydrophobic surfaces to contact with protein residues, could be one of the factors accounting for high reorganization energies [27,33].

*3.4. Solvent effects on binding energies*

The solvation free energies of nilotinib, imatinib ion, and the tyrosine kinase residues in the gas state and water were calculated using Cramer and Truhlar's SMD solvent model. The model is based on the free energy of solvation $\Delta G_S^o = \Delta G_{ENP} + G_{CDS}$ where ENP is the electronic nuclear polarization: the change in the solute free energy due to electrostatic interactions between the solute and the bulk solvent and distortion of the solute's electronic structure in solution. The solvent is modelled as a dielectric continuum. CDS is the cavitation dispersion structure, involving non-bulk solvent electrostatic contributions to the free energy of hydration. The CDS represents first solvation shell effects. It involves atomic surface tension (geometry dependent proportionality constants). The $G_{CDS}$ term has been parameterized using extensive experimental data sets for optimization, and has the advantage of including a realistic experimentally based hydrogen bonding model. The CDS involves cavitation, dispersion, and collective "solvent structure contribution" estimates for partial hydrogen bonding, repulsion, and deviation of the dielectric constant from its bulk value.

Bulk solvent electrostatic interactions are long-range electrostatic polarization effects.

The CDS covers *shorter-range* polarization effects and shorter-range non-electrostatic effects such as cavitation, dispersion, and solvent structural effects (which includes both hydrogen bonding) and exchange repulsion effects. The hydrogen bonding model uses Abraham's solvent model where α is the hydrogen bond acidity and β is the hydrogen bond basicity. The CDS contribution is a sum of terms (with atomic surface tensions that are proportional to the solvent-accessible surface areas of the individual atoms of the solutes).

Implicit solvation models such as the SMD model have been extensively used as a basis for determining solute – solvent free energies in protein and enzyme folding processes. The solvation free

energy is the energy required to transfer a solute molecule from a "vacuum" (or gas phase) into a solvent. Transfer free energies (the difference between the free energy of a solute in water and another solvent, eg n-octanol) have been used to quantify the hydophobicity of drugs and similar molecules.

The SMD model optimization and parameterization included extensive experimental transfer free energy data, which makes it a good model for examining transfer free energies of statins in various solvents. The model also treats ionized and neutral species well. Using the B3LYP/6-31G* level of theory, the SMD model achieves a mean unsigned error of 0.6-1.0 kcal/mol for neutrals and 4 kcal/mol for ions.

The $\mathbf{\Delta G_{desolv\text{-}inhib} + \Delta G_{desolv\text{-}tyrkin}}$ values are shown in Table 1, and $\mathbf{\Delta G_s}$ values for the inhibitors in various solvents are shown in Table 3. Solvation effects are also discussed in terms of hydrophilic and hydrophobic solvation in section 2.6.

**Table 3.** Solvation free energies and molecular parameters for free* nilotinib and imatinib ion in various solvents

| | $\mathbf{\Delta G_s}$ Nilotinib kcal/mol | CDS Nilotinib kcal/mol | $\mathbf{\Delta G_s}$ Imatinib$^+$ kcal/mol | CDS Imatinib$^+$ kcal/mol | Dipole Nilotinib D | Dipole Imatinib$^+$ D |
|---|---|---|---|---|---|---|
| Gas / Vacuum | 0 | 0 | 0 | 0 | 11.5<br><5.7> | 44.0<br>[156] |
| Benzene | -28.2<br><-27.0> | -14.5<br><-16.2> | -57.8<br>[-58.7] | -13.8<br>[-14.9] | 13.8<br><6.7> | 51.5<br>[158] |
| n-Octane | -24.8<br><-24.2> | -13.6<br><-15.2> | -50.3<br>[-51.3] | -12.9<br>[-13.9] | 13.4<br><6.6> | 50.9<br>[158] |
| n-Octanol | -31.8<br><-25.6> | -1.1<br><-1.7> | -81.6<br>[82.5] | -2.9<br>[-4.1] | 17.5<br><8.1> | 54.5<br>[160.4] |
| Water | -29.6<br>{-27.0}<br><-18.5> | 9.1<br>{9.9}<br><10.4> | -87.2<br>[-87.0]<br>{[-86.8]}<br><[-77.3]> | 2.6<br>[2.6]<br>{[2.7]}<br><[6.2]> | 70.5<br>{19.0}<br><8.7> | 55.4<br>[156]<br>{161}<br><154.1> |

**Footnotes:**

free* = Nilotinib (from 3CS9 X-ray structure) and imatinib ion 1IEP in same configuration and conformation as complexed with ABL tyrosine kinase in the X-ray structures but with tyrosine kinase residues removed (see Table 5 for list of residues)

Values in [..] are for imatinib from the 2HYY X-ray structure

Values in {..} are for molecular mechanics optimised structures, values in <..> are for quantum mechanics optimised configurations

### *3.5. Binding energy results*

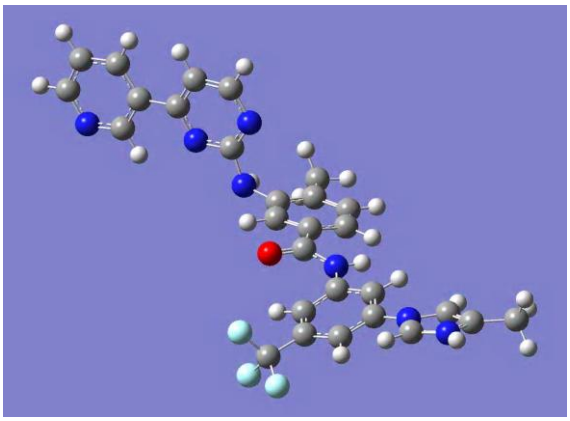
Table 1 gives the binding energy data for nilotinib and imatinib ion with tyrosine kinase using the 3CS9 and 2HYY X-ray structures directly. The structure from 1IEP was also used to gauge what effects the significant structural distortions clearly present would have on binding energy calculations compared to the other structures where the distortions are far less. The table also includes literature data from five recent binding studies of the same complexes using different molecular mechanics – molecular dynamics approaches [11,12,34-36].

The binding energy $\mathbf{\Delta E_{int}}$ for the imatinib ion - tyrosine kinase complex (2HYY) (excluding the 9 outer residues) was -214.3 kcal/mol. The total coulombic bonding energy for 7 hydrogen and 3 π-π bonds is -50.4 kcal/mol, leaving a vdW dispersion energy of -163.9 kcal/mol. Also shown in Table 1 is $\mathbf{\Delta E_{int}}$ -397.3 kcal/mol where all 9 outer kinase residues are included, leaving a total vdW dispersion energy -346.9 kcal/mol.

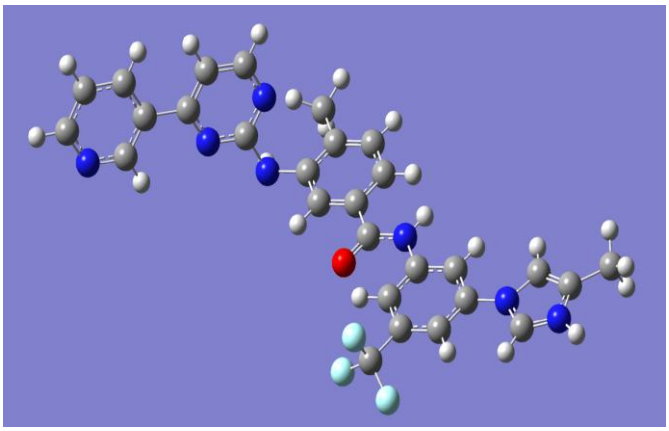
The bare imatinib ion configurational / conformational energy change from the free QM *optimised* to the free unoptimised state (1IEP X-ray structure geometry) is 698.4 kcal/mol (gas), a very substantial destabilisation in molecular energy. The bare imatinib ion from the 1IEP structure is less stable by 506.2 kcal/mol than the same ion from the 3CS9 X-ray structure. Examination of the imatinib molecule in the X-ray structure shows large distortions of bond lengths, angles and even significant distortion of aromatic rings, which appear to be artifacts of the X-ray structural interpretation. This data is consistent with the X-ray structure validation report for 1IEP discussed in 2(1) above. This data clearly demonstrates that distorted X-ray structures give unrealistic results, indicating that care must be exercised when choosing X-ray structures for binding studies. Many literature papers do not clearly define the "pre-treatment" methods used when adopting X-ray structures for binding studies.

The 2HYY X-ray structure indicates that free imatinib ion (extracted from the complexed state) is less stable than the free MM *optimised* imatinib ion by128.8 kcal/mol (gas). This large difference of 128.8 kcal/mol is the mainly the result of a conformational change from the bound configuration which has the pyrimidine-pyridine plane at an $80^o$ angle (about the $N_{31}$-$C_{33}$ bond) to the adjoining phenyl plane, and the piperazine ring is about $60^o$ to the adjoining phenyl ring, whereas the MM optimised structure has the pyrimidine-pyridine plane coplanar to the adjoining phenyl ring with the piperazine ring unchanged at $60^o$ to the adjoining phenyl ring. A full QM energy optimisation of the bound state of imatinib produces an energy stabilisation of 194.2 (gas) or 182.7 (water) kcal/mol and a structure where most of the molecule is almost flat (the pyridine ring is tilted $20^o$ from the adjoining pyrimidine ring and the phenyl group adjoining the piperazine ring is tilted $30^o$ from the -C(O)N(H)- linkage) but the piperazine ring is about $63^o$ to the adjoining phenyl ring. Similarly, MM and full QM energy optimisation of the bound nilotinib in water (from the 3CS9 X-ray structure) produce structures where (a) the MM molecular configuration is similar to the bound configuration, but the pyrimidine – pyridine plane is rotated $150^o$ (about the $N_{31}$-$C_{33}$ bond) compared to $175^o$ to the adjoining phenyl ring in the bound configuration, and (b) the QM optimised structure is almost flat with the pyridine ring tilted at $20^o$ to the adjoining pyrimidine ring. The -$CF_3$-$C_6H_3$-N(H)-C(=O)-plane is tilted $28^o$ to the adjoining toluoyl ring, and the pyrrole ring is tilted $18^o$ from the $CF_3$-C6$H_3$-N(H)-C(=O)-plane. The QM configuration is more stable by -74.9 kcal/mol compared to the MM optimised configuration, which is -56.4 kcal/mol more stable than the bound state. (Figures 2 and 3) These data indicate that some degree of geometrical distortions are present in both the bound forms of nilotinib and more so in imatinib in the X-ray structures, consistent with the X-ray quality reports (see discussion under X-ray structures above). These distortions are clearly far smaller than those observed in the 1IEP structure. *Clearly a major conclusion from these studies is that optimisations of structures where substantial electronic interactions are operating across the molecular structure (as in nilotinib and imatinib) using empirical molecular mechanics force fields may give erroneous results compared to using a full quantum mechanical optimisation method which can account for full configurational and conformational electronic effects. Such errors would be amplified when adding solvent interactions to gas phase calculations.*

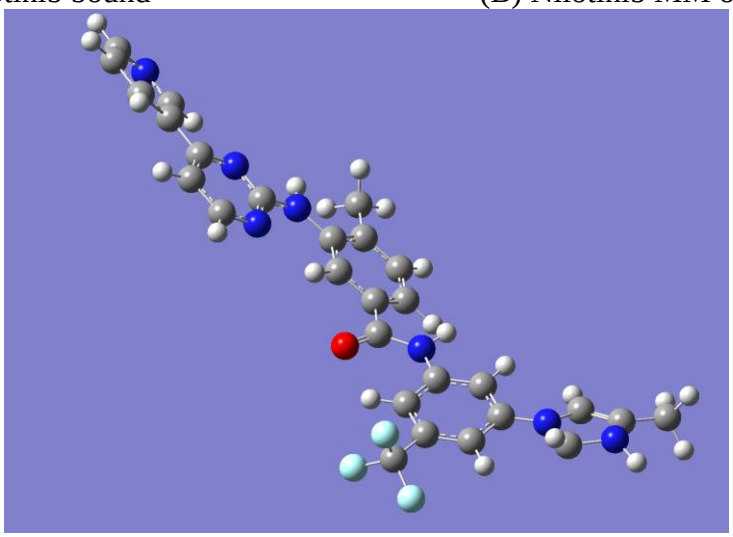
**Figures 2.** Nilotinib in (**A**) bound (complexed with tyrosine kinase), (**B**) optimised by molecular mechanics (MM), (**C**) optimised by quantum mechanics (QM) configurations



(A) Nilotinib bound

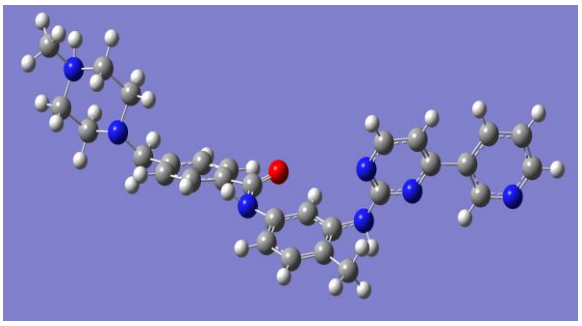


(B) Nilotinib MM optimised

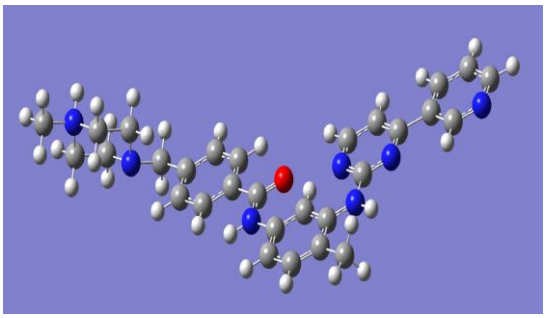


(**C**) Niltonib QM optimised

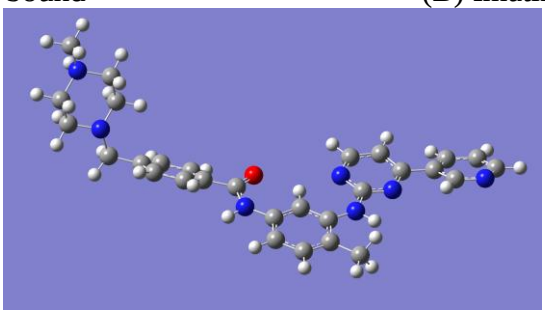
**Figure 3.** Imatinib in (A) bound (complexed with tyrosine kinase), (B) optimised by molecular mechanics (MM), (C) optimised by quantum mechanics (QM) configurations



(A) Imatinib bound



(B) Imatinib MM optimised



(C) Imatinib QM optimised

The calculated gas dipole moments (Table 3) for the free unoptimised imatinib ion (1IEP) is 44.0D compared to the 2HYY structure value of 156.0D, reflecting large charge differences in the molecular structures. The free QM *optimised* imatinib ion (2HYY) has a dipole moment of 151.5D. It is interesting that the dipole moment of the uncharged nilotinib increases from 11.5D to 70.5D from the gas to water phase, indicating that significant polar charge interaction between water and nilotinib in the first CDS solvent shell (which does not occur with larger polar solvents such as n-octanol). This notion is supported by the large CDS water solvation value of 9.1 kcal/mol included in the $\mathbf{\Delta G_s}$ -29.6 kcal/mol (Table 3).

The calculated $\mathbf{\Delta E_{int}}$ for the nilotinib ion - tyrosine kinase complex (3CS9) was -170.2 kcal/mol. The total coulombic bonding energy for the 6 hydrogen bonds, 4 polar bonds, and 3 π-π interactions is -44.8 kcal/mol leaving a vdW dispersion energy of -125.4 kcal/mol.

The nilotinib configurational / conformational energy change from the MM (QM) free optimised state to the bound state in vacuo (as per the 3CS9 X-ray structure geometry) is MM +11.0 (QM -56.4) kcal/mol. The free MM *optimised* nilotinib is *less* stable than the conformation of the complexed nilotinib from the X-ray structure by 11.0 kcal/mol (gas phase), but in water, this difference is MM -15.9 (QM -74.9) kcal/mol. The free *optimised* nilotinib has a gas dipole moment of MM 10.8D (QM 5.7D) compared to the free unoptimised nilotinib value of 11.5D.

These above data for nilotinib and imatinib are indicative of a small configurational distortion in the 3CS9 X-ray structure, compared to the very large distortions found for imatinib in the 1IEP structure, and to a lesser extent in the 2HYY structure. It is also clear that molecular mechanics optimisations cannot fully optimise structures as well as ab initio quantum mechanical methods, which can account for electronic effects which dominate in various conformations. Also large differences in molecular charge distribution (as reflected in dipole moments) and intertwined solvent effect interactions occur when QM optimisation methods are used instead of MM optimisations. It is unclear how such effects

have impacted the literature binding studies shown in Table 1 which all use MM force field based methods.

Table 1 also shows the differences in free energy of water solvation **ΔG** data for the inhibitors and tyrosine kinase in water calculated with the PCM/SMD solvent model. The differences in *configurational entropy* ***TΔS*** for the inhibitors and tyrosine kinase from QM thermochemical calculations at 298K in water are also shown. The differences in total thermal energies $\varepsilon_{total} = \varepsilon_{translational}+\varepsilon_{rotational}+\varepsilon_{vibrational}+\varepsilon_{electronic}$-are also shown for reference only as these terms are included in $\Delta E_{int}$.

The $\Delta G_{desolv\text{-}inhib}$ for the imatinib ion – TK complex (2HYY) in Table 1 show two values, 87.2 and 77.3 kcal/mol, which are the values for the bound state and the QM optimised state of imatinib. It was shown above that the bound configuration of the imatinib ion in water is less stable than the QM optimised state by 182.7 kcal/mol. Using the former solvation value gives a $\Delta G_{bind}$ -10.9 kcal/mol, (experimental value -10.4 kcal/mol) whereas the latter values gives $\Delta G_{bind}$ – 20.8 kcal/mol. This observation indicates that the desolvation process of the imatinib ion just prior to complexing then allows, during the complexation process, a higher energy conformation (that closely resembles the bound conformation) that fits the into kinase receptor pocket. This result contrasts to the situation for nilotinib where the $\Delta G_{desolv\text{-}inhib}$ 18.5 kcal/mol for the QM optimised state gives $\Delta G_{bind}$ -10.6 kcal/mol (compared to the experimental value -10.6 kcal/mol). The bound conformation for nilotinib is less stable than the QM optimised conformation in water by 74.9 kcal/mol. Very similar results were observed for the 1IEP structure in Table 1 (although this result is partly compromised by the structural distortions in the starting X-ray structure). Atwell [37] has found that imatinib can tightly bind in a trans conformation, as well as in a weakly bound cis conformation, which may be related to the $\Delta G_{desolv\text{-}inhib}$ data discussed above.

Comparison of the desolvation data of imatinib ion and the neutral nilotinib indicates that "pre-organisation" of the inhibitor prior to binding with tyrosine kinase is driven by the thermodynamics of desolvation and the kinetics of the available conformational states that can fit into the (possibly dynamic) kinase receptor pocket. Despite the large thermodynamic desolvation penalties for both inhibitors, the imatinib ion would be expected to undergo slower rates of desolvation and accompanying conformational change than that for nilotinib. Partial desolvation and kinetically driven and controlled intramolecular conformational rearrangement would also be likely, particularly around the quaternary $N^+$ piperazinyl moiety as it complexes to the tyrosine kinase within the Ile293, Leu298, Leu354, Val379 hydrophobic pocket. Effectively desolvation would be the slow rate determining kinetic step, followed by rapid intramolecular conformational processes of the inhibitors as binding occurs to the kinase receptor.

As nilotinib is known to be ca. 30 more effective than imatinib in treating CML, and conformational selection has been shown to be a dominant driver of inhibitor effectiveness, these data are consistent with an intertwined desolvation – conformational selectivity process for nilotinib and imatinib in their particular tyrosine kinase binding pockets. In essence a balance of thermodynamic desolvation and intramolecular conformational kinetic control of the inhibitors is occurring such that the less stable conformations of the inhibitors becomes the dominant product in the complexes. If this conclusion is substantially correct, it implies that drug design of similar CML inhibitors must consider conformational flexibility coupled with solvation properties.

The binding energy for the 1IEP structure where the structural distortions are large (and far in excess of those in the 2HYY and 3CS9 structures) shows that these artefacts do have a significant effect on $\mathbf{\Delta E_{int}}$ as expected (difference of -118.2 kcal/mol between the 1IEP and 2HYY structures, although the number of residues is slightly different, 13 versus 11 respectively, see Table 5), and the $\mathbf{\Delta E_{elect}}$ values (difference of -11.1 kcal/mol between 1IEP and 2HYY structures). The calculated binding energies for the (unoptimized, no "pre-treatment") bound complexes are -50.8 and -10.9 kcal/mol for the 1IEP and 2HYY structures respectively (experimental value -10.4 kcal/mol). The main difference in the component energies that contribute to the overall binding energies of the 1IEP versus 3CS9 structures is the $\mathbf{\Delta E_{int}}$ term, and to a lesser extent $\mathbf{\Delta E_{elect}}$, whereas the solvation energies and configurational entropies are fairly similar. *This side by side comparison shows that the large difference in calculated binding energy is due to some steric clashes (unusually large van der Waals repulsive interaction) between the imatinib and the tyrosine kinase residues, which produce an abnormally high* $\mathbf{\Delta E_{int}}$ *which is largely comprised of* $\mathbf{\Delta E_{vdW}}$ *interactions*. These results clearly show that X-ray structures cannot necessarily be trusted to be totally accurate in all details, and that "pre-treatment" to remove structural distortions (bond angles, lengths, steric clashes (unnatural overlap of two non-bonding atoms), addition of missing atoms, etc) prior to calculating binding energies may be important if the X-ray structure has a poor validation report (see section 2(1) above). However, "pre-treatment" may not necessarily remove all distortions, and possibly could induce unwanted structural distortions in the process. It is clear that much care must be used when using X-ray structures for calculational purposes.

Inspection of the results in Table 1 show consistently close calculated binding energies with experimental results, despite quite different calculational approaches being used, and vastly different component energies that contribute to the overall binding energies. The literature results are all molecular mechanics – molecular dynamics studies, and the reasons for the different calculated inputs to the binding energies are unknown but may be due to (a) differing "pre-treatments" of the X-ray structural data may have been used, such as optimisation, addition of missing atoms (eg H atoms), "pre-correction" of anomalous bond lengths, angles etc, particularly where the 1IEP X-ray structure was used; structural anomalies are easy to detect in the inhibitor molecules, but far more difficult to detect in protein chains (b) different residual tyrosine kinase residues surrounding the inhibitors may have been used in the calculations, changing values for $\mathbf{\Delta E_{int}}$ and $\mathbf{\Delta E_{vdW}}$ (c) different calculated inhibitor charges from MK/RESP (d) different treatment of solvation effects, including implicit and explicit effects (e) different parameter optimisation of conformers used in the force field etc. The binding energy results in Table 1 appear to suggest that as long as a consistent methodology is applied to the thermodynamic binding cycle, the overall outcome is close to the experimental values (as long as structural distortions in the X-ray structures are largely removed in the "pre-treatment" process), even though the component contributing energies are very different, and possibly compensating errors in the individual calculational methods are present. Variations in the solvation method used by the individual studies are likely to be significant, as the solvation energies are large, and the calculation of non-polar solvation terms can vary, particularly if different empirical energy factors are used with the solvent accessible surface areas to calculate the implicit non-polar solvation terms, or different empirical solvent interaction potentials are used in explicit solvation energy calculations (see section 2.6 below).

It is clear from Table 1 that van der Waals interactions (and solvent effects) are the dominant forces driving binding energies. With the exception of the study [12], the $\Delta E_{vdW}$ (hydrophobic) term is far larger than the $\Delta E_{elect}$ (hydrophilic) term. The $\Delta E_{vdW}$ values from QM calculations are much larger than those derived from MM methods, which appear anomalously low. This is consistent with most small molecule – protein interactions where the hydrophobic interaction is the dominant factor over the hydrophilic (or hydrogen and polar bonding) interaction [5]. Table 1 shows the $\mathbf{\Delta E_{int}}$ value for the imatinib ion – kinase complex (2HYY) where an additional 9 residues are added, including the 4 residues identified as a pocket surrounding the charged quaternary piperazinyl $N^+$ atom. These 9 residues add an extra 183.0 kcal/mol of van der Waals interaction.

Separate (implicit) solvation energies are reported by Pricl [34] and Dubey [35] for the imatinib complex 51.4 and 108.0 kcal/mol, with $\Delta G_{NP}$ values of -8.7 and -6.7 kcal/mol respectively. These values are calculated from $\Delta G_{solv} = \Delta G_{PB} + \Delta G_{NP}$ using the MM/PBSA method. $\Delta G_{PB}$ is the polar solvation term. $\Delta G_{NP}$ the non-polar solvation term, which includes cavity creation in water, and the vdW interaction between the non-polar parts of the kinase and water. $\Delta G_{NP}$ is analogous to the transfer free energy when moving a "non-polar" molecule from vacuum to water. Lin [13] and Aleksandrov [12] used explicit solvation models in their binding energy calculations and no separate solvation energies are reported. Table 1 shows the separate QM PCM/SMD desolvation energies for the inhibitors and the tyrosine kinase, with the CDS values. The CDS value for imatinib is vastly different from the literature $\Delta G_{NP}$ values since it includes CDS involves cavitation, dispersion, and collective "solvent structure contribution" estimates for partial hydrogen bonding, repulsion, and deviation of the dielectric constant from its bulk value. It is unclear how explicit solvent models incorporate physical solvent interactions, and what accuracy is involved. The PCM/SMD solvent model (which is *parameterised* with extensive experimental data, including hydrogen bonding) has been tested against a large number of experimental data, and the accuracy is known. *It is clear from Table 1 that solvent effects (along with van der Waals interactions) are the dominant factors in determining binding energies*.

It is also possible to get an estimate of how solvent effects interact with the configurational entropies for imatinib ion and nilotinib. The $\mathbf{\{T\Delta S_{free} - T\Delta S_{bound}\}_{inhib}}$ values for nilotinib and imatinib ion in water (and in vacuo) are 14.7 (16.1) and 8.0 (14.1) kcal/mol respectively. The values for the neutral nilotinib are about the same within error, but for the positively charged imatinib the difference is significant, indicating a significant solvation contribution to the configurational entropy.

It is clear from this study that the active species that binds to tyrosine kinase is the species which has a protonated quaternary $N^+$ on the terminal nitrogen of the piperazine ring. Calculations with a neutral imatinib did not give realistic binding energies. This result is consistent with the known pKa of imatinib in blood serum.

*3.6. Hydrophilic and hydrophobic nature of inhibitor interaction with tyrosine kinase*

The binding of a ligand and a protein can be also portrayed in terms of hydrophilic (hydrogen and polar bonding) and hydrophobic interactions, as these descriptions are often used clinically when describing the properties of drugs. Hydrophilicity relates to water (or blood serum) solubility, and hydrophobicity relates to the clinical ability of drugs to pass through cell membranes (such as crossing the blood brain barrier) or lipophilic solubility in fatty tissues. *As lipophilicity increases, there is an increased probability of binding to hydrophobic protein targets other than the desired one, and therefore, there is more potential for toxicity.* Hydrophilicity and hydrophobicity / lipophilicity are important clinical features of any administered drug. The relationship to binding can be expressed by:

$$\mathbf{\Delta E_{int} = \Delta E_{hydrophilic} + \Delta E_{hydrophobic}} \text{ (water)}$$

Much work has been done in attempting to find solvents that can mimic the interior properties of proteins or biological membranes. Octanol has been the most used solvent, with the widespread use of water-n-octanol partition coefficients (log P, log D) being widely used in drug design. However the laboratory determined coefficients are compromised by the known fact that n-octanol contains ~2.3 M of water. Other solvents have been suggested to better represent the interior of proteins, including cyclohexane, 1,9 decadiene, and n-octane [41-45]. Partition co-efficients of a non-polar solute (log S) or an ionized solute (log D) between water and n-octanol are experimentally determined, and the free energy, enthalpy and entropy can be determined by isothermal calorimetry.

Implicit solvation models have been used to calculate the transfer free energy of solvation $\Delta G_s$ (defined as the difference between the energy of the solute in the gas phase and in a particular solvent) of solute solvent interactions. Hence the transfer of a drug from a water environment (such as blood serum) to a binding pocket of a protein or enzyme can be mimicked by determining the transfer free energy of solvation from water to a solvent such as n-octanol, cyclohexane or n-octane which is taken as a proxy for the interior of the protein.

The hydrophobic effect is a major factor in the energetics of the folding of globular proteins with hydrophobic interior cores. The solvent accessible surface area (SASA) of the solute or drug has been widely used as a proxy for the extent of the hydrophobic effect when a drug binds to a protein. The extent of this effect lies between 5 and 45 cal/($Å^2$ mol) [46]. Ladbury et al [47] have examined the SCORPIO data base of published isothermal titration calorimetric results for a range of protein and small ligand interactions with changes in solvation (using the polar and non-polar solvent accessible surface area of the ligand and changes resulting from the protein-ligand complexation). Most interactions were enthalpy driven. The strongest correlation was between non-polar surface area burial upon complexation and the binding free energy, consistent with an entropy driven process (with $T\Delta S^o$ being about double $\Delta H^o$). However the free energy contribution per unit area buried was only about 30-50% of previous similar studies of transfer free energies of small ligands. The transfer of an ion from water to a nonpolar media with dielectric constant of ~3 (lipid bilayer) or 4 to 10 (interior of proteins) costs significant energy [40,41].

Lomize et al [49] have shown that atomic solvation parameters (ASP), which are solvation transfer energies, of a number of small molecules from water to the interior of a large number of proteins that: (a) aliphatic and aromatic groups are closely related to solvent transfer energies from water-

cyclohexane partition coefficients, and (b) sulphur, nitrogen and oxygen groups are intermediate between water-octanol and water-cyclohexane partitioning coefficients.

To assess which solvents might be a proxy for the hydrophilic interior of tyrosine kinase, a number of solvents were examined. The solvation free energies of nilotinib in various solvent are shown in Table 3 calculated using Cramer and Truhlar's SMD solvent model [9]. Table 3 shows the solvation free energies for neutral nilotinib and imatinib ion in water and various solvents, including n-octanol (the basis of log D measures) which has both polar and non-polar characteristics, to n-octane (non-polar with no hydrogen bonding capability) and benzene (non-polar, but has solvent π interacting features). Transfer free energies between water and n-octane (which has no hydrogen bonding capability) were examined to measure hydrophobicity

Literature calculations of solvent effects using the PBSA model $\Delta G_{solv} = \Delta G_{PB} + \Delta G_{NP}$ include the estimation of the $\Delta G_{NP}$ term using an empirical hydrophobicity factor based on the calculated solvent accessible surface area (SASA). For the imatinib – kinase complex, values used vary from 5.4, 7.2 and 50 cal/(mol/Å$^2$) respectively [12,34,35,] to give $\Delta G_{solv}$ values of 51.4-61.4 and 101.3 [34,35].

Using the values in Table 4, which contain the solvent accessible surface areas and buried surface area for the inhibitors, tyrosine kinase and complexes taken from the literature X-ray studies, it is possible to calculate the empirical hydrophobic proxy factors that would be required to give the transfer free energies (the differences between the CDS values for water and n-octane, in the QM optimised states using the PCM/SMD solvent model) for the inhibitors taken from Table 1 (20.1 and 25.6 kcal/mol for the imatinib ion and nilotinib respectively). The reverse calculated hydrophobic proxy factors are 28.0 and 34.6 cal/(mol/Å$^2$) for the imatinib ion and nilotinib respectively. The value for imatinib is very different from the arbitrary PBSA values of 5.4, 7.2 or 50 cal/(mol/Å$^2$) used in the literature [34,35,12]. It is noted that the SMD solvent model uses exposed surfaces of atoms to calculate the non-polar or non-electrostatic component of the free energy of solvation, and these values are *parameterized* with a large number of experimentally determined solvent properties, including macroscopic surface tensions, refractive index, hydrogen bonding acidity and basicity etc. *These data suggest that a major error in inhibitor – kinase binding energy calculations lies in the non-polar solvation terms of the solvation calculations.* Table 4 also contains similar proxy factors for the tyrosine kinase, however these values are only indicative estimates, as the residues used in the model are not a complete representation of the full kinase structure.

**Table 4.** Transfer free energies and solvent accessible surface area hydrophobic proxy factors for inhibitor – tyrosine kinase complexes

| | Imatinib Inhibitor 2HYY | Nilotinib Inhibitor 3CS9 | Tyrosine kinase Interface Chain A 2HYY | Tyrosine kinase Interface Chain A 3CS9 |
|---|---|---|---|---|
| Inhibitor Polar Surface Area* $Å^2$ | 86 | 98 | | |
| Inhibitor Log P | 4.4 | 4.4 | | |
| Inhibitor Major Species at pH 7.4 | +1 | 0 | | |
| Hydrophobic Surface Area as % Total SASA** | | | 56.3 | 54.3 |
| Complex Solvent Accessible Surface Area (SASA) $Å^2$ | 793 | 803 | 576 | 570 |
| Complex Buried Surface Area $Å^2$ | 718 | 742 | | |
| Transfer Free Energy Water-Octane CDS[#] kcal/mol | 20.1 | 25.6 | 48.0 | 44.9 |
| SASA Hydrophobic Proxy Factor[##] cal/(mol/$Å^2$) | 28.0 | 34.5 | 14.8 | 14.5 |

**Footnotes:**

* PubChem: http://pubchem.ncbi.nlm.nih.gov

** POPS calculation of total solvent accessible surface of kinase: http://mathbio.nimr.mrc.ac.uk/wiki/POPS F. Fraternali, L. Cavallo, Nucleic Acids Res.2002, **30,** 2950.

3CS9 and 2HYY refer to data taken from these PDB X-ray structures

[#] Transfer Free Energy Water-Octane CDS values: taken from Table 3 are the differences between the CDS values for water and n-octane, in the QM optimised states (see text)

[##] SASA Hydrophobic Proxy Factor: calculated in cal/(mol/$Å^2$) which would be required to multiply the buried surface / solvent accessible surface areas of the inhibitors or tyrosine kinase to equate to the transfer free energies CDS values in row 8

**Table 5.** Amino acid residues used in binding energy calculations from 2HYY, 1IEP and 3CS9 X-ray structures.

| 2HYY X-ray Structure | | 1IEP X-ray Structure | | 3CS9 X-ray Structure | |
|---|---|---|---|---|---|
| LEU | 248 | | | | |
| TYR | 253* | | | | |
| ALA | 269* | | | | |
| GLU | 286* | | | | |
| VAL | 289 | TYR | 253 | TYR | 253 |
| MET | 290* | GLU | 286 | ALA | 269 |
| ILE | 293 | VAL | 289 | GLU | 286 |
| LEU | 298 | MET | 290 | MET | 290 |
| ILE | 313 | VAL | 299 | ILE | 293 |
| THR | 315* | THR | 315 | THR | 315 |
| PHE | 317* | PHE | 317 | PHE | 317 |
| MET | 318* | MET | 318 | MET | 318 |
| GLY | 321 | ILE | 360 | HIS | 361 |
| LEU | 354 | HIS | 361 | VAL | 379 |
| ILE | 360* | LEU | 370 | ALA | 380 |
| HIS | 361* | ASP | 381 | ASP | 381 |
| ARG | 362 | PHE | 382 | PHE | 382 |
| VAL | 379 | | | | |
| ASP | 381* | | | | |
| PHE | 382* | | | | |

**Footnote:**

2HYY: 11 inner* + 9 outer residues

## 4. Conclusions

The binding energies of imatinib and nilotinib to tyrosine kinase have been determined by quantum mechanical (QM) computations, and compared with literature binding energy studies using molecular mechanics (MM) or molecular mechanics / molecular dynamics methods.

An assessment of potential errors in the computational methods has been made which indicates the following critical factors:

- Much care must be taken when using published X-ray structures as the starting basis for determining binding energies, as it has been shown that two PDB structures 1IEP and 2HYY for the same imatinib-tyrosine kinase complex give vastly different binding energies owing to artificial structural distortions in the 1IEP structure. The structural distortions are steric clashes which give unrealistically high van der Waals energies. The 2HYY structure which has far fewer structural distortions gives a binding energy close to the experimental value.
- Surprisingly, the very different QM and MM/MD computational methods give binding energies close to experimental values, despite very different component contributions to the overall binding energy.

- It has been shown by comparing QM and MM optimised structures for nilotinib and imatinib ion in water for the free and complexed states that the MM optimisation gives a very different configuration to the QM optimisation for nilotinib, whereas the imatinib ion gives similar configurations. This finding reinforces the known weakness of empirical MM methods in determining the lowest energy configurational conformation, implying MM based binding energies may be suspect as a result.
- Solvation energies have been shown to be a major component contributing to the overall binding energy. The established and validated QM based solvent model (PCM/SMD) used in this study gives quite different values from those used in the implicit PBSA solvent models used in the MM based literature studies, which use an arbitrary empirical solvent accessible surface area factor, cal/(mol/Å$^2$), to calculate the non-polar solvent contribution. A major error in inhibitor – kinase binding energy calculations lies in the non-polar solvation terms of the solvation calculations.
- Solvent transfer free energies have been calculated using the PCM/SMD model, and the required empirical solvent accessible surface area factors for nilotinib and imatinib ion to give the transfer free energies have been reverse calculated. These values are very different from those used in the MM PBSA studies, suggesting this area is a source of error in MM PBSA calculations.
- Comparison of the desolvation data of imatinib ion and the neutral nilotinib indicates that "pre-organisation" of the inhibitor prior to binding with tyrosine kinase is driven by the thermodynamics of desolvation and the kinetics of the available conformational states that can fit into the (possibly dynamic) kinase receptor pocket. An intertwined desolvation – conformational selectivity process for nilotinib and imatinib in their particular tyrosine kinase binding pockets is a balance of thermodynamic desolvation and intramolecular conformational kinetic control which results in the less stable conformations of the inhibitors becoming the dominant product in the complexes.
- The configurational entropy (TΔS) in water has been calculated for the imatinib ion from the free and bound configurations, and is similar to the MM literature values for imatinib (with one exception). As the QM or MM optimised imatinib ion gives similar conformations, the agreement with the literature TΔS values is expected. However the same result would not apply to nilotinib as the QM optimised structure is very different from the MM structure. These values fall within the literature range for TΔS for a wide range of protein-drug complexes

**Acknowledgments**

**Author Contributions**

**Conflicts of Interest**

The author declares no conflict of interest.

**References and Notes**

1. American Cancer Society, 2013, "Targeted Therapies for Chronic Myeloid Leukaemia", http://www.cancer.org/cancer/leukemia-chronicmyeloidcml/detailedguide/leukemia-chronic-myeloid-myelogenous-treating-targeted-therapies

2. Okuno, S.H., 2011, "The Use of Tyrosine Kinase Inhibitors for Gastrointestinal Stromal Tumors (GIST)", http://www.onclive.com/publications/contemporary-oncology/2011/spring-2011/the-use-of-tyrosine-kinase-inhibitors-for-gastrointestinal-stomal-tumors-gist#sthash.5cJlC3xm.dpuf http://www.onclive.com/publications/contemporary-oncology/2011/spring-2011/the-use-of-tyrosine-kinase-inhibitors-for-gastrointestinal-stomal-tumors-gist
3. Gilson M.K., Zhou H.X., Calculation of Protein Ligand Binding Affinities, Ann. Rev. Biophys. Biomol. Struct., 2007, 36, 21.
4. Mobley D.A., Dill K.A., Binding of Small-Molecule Ligands to Proteins:''What You See'' Is Not Always ''What You Get'', Structure, 2009, 17, 489.
5. Kuriyan J., Konforti B., Wemmer D., The Molecules of Life, Garland, 2008, Ch 12.
6. Bissantz C., Kuhn B., Stahl M., A Medicinal Chemist's Guide to Molecular Interactions, J. Med. Chem., 2010, 53, 5061.
7. Mackerell A.D., Empirical Force Fields for Biological Macromolecules:Overview and Issues, J. Comput. Chem. 2004, 25, 1584.
8. Carpinelli M., Colonna G., D'Angola A., Molecular partition function: vibrational, rotational and electronic contributions, Ch. 5, Fundamental Aspects of Plasma Chemical Physics, Springer Series on Atomic, Optical, and Plasma Physics
9. Marenich A.V., Cramer C.J., Truhlar D.J., Universal Solvation Model Based on Solute Electron Density and on a Continuum Model of the Solvent Defined by the Bulk Dielectric Constant and Atomic Surface Tensions, J. Phys. Chem B, 2009, 113, 6378.
10. Brameld K.A., Kuhn B., Reuter D.C., Stahl M., Small molecule conformational preferences derived from crystal structure data. A medicinal chemistry focused analysis, J. Chem. Inf. Model. 2008, 48, 1.
11. Seeliger M.A., Nagar B., Filipp F., Xixian C., Henderson M.N., Kuriyan J., c-Src Binds to the Cancer Drug Imatinib with an Inactive Abl/c-Kit Conformation and a Distributed Thermodynamic Penalty Structure, 2007, 15, 299.
12. Aleksandrov A., Simonson T., Conformational selection governs the binding preferences of imatinib for several tyrosine kinases J. Biological Chem. 2010, 285, 13807.
13. Lin Y., Yilin M., Wei J., Roux B., Explaining why Gleevec is a specific and potent inhibitor of Abl kinase, www.pnas.org/cgi/doi/10.1073/pnas.1214330110
14. Sigfridsson E., Ryde U., Comparison of Methods for Deriving Atomic Charges from the Electrostatic Potential and Moments, J. Comput. Chem, 1998, 19, 377.
15. Breneman C.M., Wiberg K., Determining atom-centered monopoles from molecular electrostatic potentials, J. Comp. Chem. 1990, 11, 361.
16. Davis A., Stgallay S., Kleywegt G., Limitations and lessons in the use of X-ray structural information in drug design, Drug Discovery Today, 2008, 13, 831.
17. Thompson H.P.G., Day G.M., Which conformations make stable crystal structures? Mapping crystalline molecular geometries to the conformational energy landscape, Chem. Sci**.**, 2014, **5**, 3173.
18. Baker E., Hubbard R., Hydrogen-Bonding in Globular-Proteins, Prog. Biophys. Mol. Biol., 1984,44, 97.
19. Hubbard R.E., Hydrogen Bonds in Proteins: Role and Strength, Encyclopedia of Life Sciences, 2001, 1, McMillan, www.els.net.

20. Morozov A.V., Kortemme T., Tsemekhman K., Baker D., Close agreement between the orientation dependence of hydrogen bonds observed in protein structures and quantum mechanical calculations, PNAS, 2004, 101, 6946.
21. Weisberg E., Manley P.W. et al, Characterization of AMN107, a selective inhibitor of native and mutant Bcr-Abl, Cancer Cell 2005, 7, 129.
22. Manley P.WCowan-Jacob., S.W., Fendrich G., J. Mestan, 2005, Blood (ASH Annual Meeting Abstracts) 106: Abstract 3365, American Society of Hematology
23. Scott L.R., Stigliano A.F., 2013, TR-2013-10, "A disruptive dipole-dipole alignment promotes a stable molecular association", Department of Computer Science and Mathematics, University of Chicago, http://www.cs.uchicago.edu/research/publications/techreports/TR-2013-10.
24. Cowan-Jacob S.W., Floersheimer G.F., Furet A., Liebetanz P., Rummel J., Rheinberger G., Centeleghe P., Manley P.W., Structural biology contributions to the discovery of drugs to treat chronic myelogenous leukaemia, Acta Crystallogr., 2007, D63, 80.
25. Williams D.H., Stephens E., O'Brien D.P., Zhou M., Understanding Noncovalent Interactions: Ligand Binding Energy and Catalytic Efficiency from Ligand-Induced Reductions in Motion within Receptors and Enzymes, Angew. Chem., Int. Ed. 2004, 43, 6596.
26. Roy D., Marianski M., Maitra N.T., Dannenberg J.J., Comparison of some dispersion-corrected and traditional functionals with CCSD(T) and MP2 ab initio methods: Dispersion, induction, and basis set superposition error, J. Chem. Phys. 2012, 137, 134109.
27. Perola E,. Charifson P.S, Conformational Analysis of Drug-Like Molecules Bound to Proteins: An Extensive Study of Ligand Reorganization upon Binding, *J. Med. Chem.*, 2004, 47, 2499.
28. Csermely P., Palotai R., Nussinov R.. Induced fit, conformational selection and independent dynamic segments: an extended view of binding events, Trends Biochem Sci. 2010, 35, 539.
29. Chen I.J., Foloppe N., Conformational Sampling of Druglike Molecules with MOE and Catalyst: Implications for Pharmacophore Modeling and Virtual Screening, J. Chem. Inf. Model. 2008, 48, 1773.
30. Boström J., Norrby P.O., Liljefors T., Conformational energy penalties of protein-bound ligands, J Comput. Aided Mol. Des., 1998, 12, 383.
31. Gao C., Park M.S., Stern H.A., Accounting for Ligand Conformational Restriction in Calculations of Protein-Ligand Binding Affinities, Biophys. J., 2010, 98, 901.
32. Butler K.T., Javier Luque F., Barril X., Toward accurate relative energy predictions of the bioactive conformation of drugs, J. Comput. Chem., 2009, 30 , 601.
33. Stockwell G.R., Thornton J.M., Conformational diversity of ligands bound to proteins, J. Molec. Biol., 2006, 356, 928.
34. Pricl S., Fermeglia M., Ferrone M., Tamborini E., T315I-mutated Bcr-Abl in chronic myeloid leukaemia and imatinib: insights from a computational study, Mol. Cancer Ther., 2005, 4, 1167.
35. Dubey D.K., Ojha P.R., J. Biol. Phys. Binding free energy calculation with QM/MM hybrid methods for Abl-Kinase inhibitor 2011, 37, 69.
36. Dubey D.K., Chaubey K.A., Parveen A., Ojha P.R., Comparative study of inhibition of drug potencies of c-Abl human kinase inhibitors: A computational and molecular docking study, J. Biophys. Struct. Biol, 2010, 2, 47.
37. Atwell S. et al, Mechanism of signal transduction: Revealed by the Structure of Spleen Tyrosine Kinase: A Novel Mode of Gleevec Binding, J. Biol. Chem., 2004, 279, 55827.

38. Nagar B., Bornmann W. , Pellicena P. , Schindler T., Veach D.R., Miller W.T.,. Clarkson B, Kuriyan J., Crystal Structures of the Kinase Domain of c-Abl in Complex with the Small Molecule Inhibitors PD173955 and Imatinib (STI-571), Cancer Res, 2002,62.4236.
39. Asaki T., Sugiyama Y., Hamamoto T., Higashioka M., Umehara M., Naito H., Niwa T., Design and synthesis of 3-substituted benzamide derivatives as Bcr-Abl kinase inhibitors, Bioorganic & Medicinal Chemistry Letters, 2006,**16**,1421
40. Baldwin R.L., Weak Interactions in Protein Folding: Hydrophobic Free Energy, van der Waals Interactions, Peptide Hydrogen Bonds, and Peptide Solvation, Chapter 6, Protein Folding Handbook, ed. J. Buchner, T.Kiefhaber, published online: 31 Jan 2008,Wiley-VCH Verlag GmbH & Co. Weinheim. 2005, 127-162.
41. Baldwin R.L., Energetics of Protein Folding, J. Mol. Biol. 2007, 371, 283.
42. Cozzini P., Fornabaio M., Marabotti A., Abraham D.J., Kellogg G.E., Mozzarelli A., Free Energy of Ligand Binding to Protein: Evaluation of the Contribution of Water Molecules by Computational Methods, Curr. Med. Chem. 2004, 11, 1345.
43. Gitlin I., Carbeck J.D., Whitesides G.M., Why Are Proteins Charged? Networks of Charge–Charge Interactions in Proteins Measured by Charge Ladders and Capillary Electrophoresis, Angew. Chem. Int. Ed. 2006, 45, 3022.
44. Kukic P., Nielsen J.E., Electrostatics in proteins and protein–ligand complexes, Future Med. Chem. 2010, 2, 647.
45. Fong C.W., Statins in therapy: Understanding their hydrophilicity, lipophilicity, binding to 3-hydroxy-3-methylglutaryl-CoA reductase, ability to cross the blood brain barrier and metabolic stability based on electrostatic molecular orbital studies, Eur. J. Med. Chem. 2014, 85, 661.
46. Sharp K.A., Nicholls A.,. Fine R.F, Honig B., Reconciling the magnitude of the microscopic and macroscopic hydrophobic effects, Science, 1991, 252, 106.
47. Olsson T.S.G., Williams M.A., Pitt W.R., Ladbury J.E., The Thermodynamics of Protein-Ligand Interaction and Solvation: Insights for Ligand Design, J. Mol, Biol., 2008, 384, 1002.
48. Corbin A.S., Buchdunger E., Furet P., Druker B.J., Analysis of the Structural Basis of Specificity of Inhibition of the Abl Kinase by STI571 J Biol. Chem. 2002, 277, 32214.
49. Lomize A.L., Reibarkh M.Y., Pogozheva I.D., Interatomic potentials and solvation parameters from protein engineering data for buried residues, Protein Sci. 2002, 11, 1984
50. Martin F., Zipse H., Charge Distribution in the Water Molecule - A Comparison of Methods, J. Comp. Chem. 2005, 26, 97.
51. Kubelka J., Population Analysis, www.uwyo.edukubelka-chempopulation_analysis.pdf
52. Riley K.E., Holt B. T., Merz K.M., Critical Assessment of the Performance of Density Functional Methods for Several Atomic and Molecular Properties, J. Chem. Theory Comput. 2007, 3, 407.
53. Sapse A.M., Molecular Orbital Calculations for Biological Systems, Applications of DFT to Biological Systems, p. 102, 1998, Oxford UP, New York.
54. Sponer J., Lankas F., Substituent Effects on Hydrogen Bonds in DNA, Ch 18, Computational Studies of RNA and DNA, Springer 2006, Dordrecht, The Netherlands p 482
55. Li R., Hongwei F., Gang X., Xu X., Yijing Y., Performance of Several Density Functional Theory Methods on Describing Hydrogen-Bond Interactions, J. Chem. Theory. Comput. 2009, 5, 86.
56. van der Wijst T.,. Guerra C.F, Swart M., Bickelhaupt F.M., Performance of various density functionals for the hydrogen bonds in DNA base pairs, Chem. Phys. Lett., 2006, 426, 415.

57. Rayne S., Forest K., Accuracy of computational solvation free energies for neutral and ionic compounds: Dependence on level of theory and solvent model, Nature Proceedings, (2010) http://dx.doi.org/10.1038/npre.2010.4864.1
58. Rizzo R.C., Aynechi T., Case D.A., Kuntz I.D., Estimation of absolute free energies of hydration using continuum methods: accuracy of partial charge models and optimization of nonpolar contributions, J. Chem. Theory Comput. 2006, 2, 128.
59. Treitel N., Shenhar R., Aprahamian I., Sheradsky T., Rabinovitz M., Calculations of PAH anions: When are diff use functions necessary? Phys. Chem. Chem. Phys. 2004, 6, 1113.
60. Bauza A., Quiñonero D., Deyà P.M , Frontera A, Is the use of diffuse functions essential for the properly description of noncovalent interactions involving anions? J. Phys. Chem. A, 2013, 117, 2651.
61. Bauza A., Quiñonero D., Deyà P.M., Frontera A., Long-Range Effects in Anion–π Interactions, New Journal of Chemistry 2013, *37* (9), 2636.
62. Matczak P., Łukomska M., Assessment of various density functionals for intermolecular N → Sn interactions, Comput. Theor. Chem., **2014**, 1036, 31.
63. Dougherty, D.A., Cation-**π** Interactions Involving Aromatic Amino Acids, J. Nutrition, 2007, 137. 1504S.
64. Atomistry: http://fluorine.atomistry.com/pdb3cs9.html