Acta Crystallographica Section D
Biological Crystallography
ISSN 1399-0047

Sunitinib: from charge-density studies to interaction with proteins

Maura Malińska,[a]* Katarzyna N. Jarzembska,[a] Anna M. Goral,[a,b] Andrzej Kutner,[c] Krzysztof Woźniak[a] and Paulina M. Dominiak[a]*

[a]Department of Chemistry, University of Warsaw, Pasteura 1, 02-093 Warsaw, Poland, [b]College of Inter-Faculty Individual Studies in Mathematics and Natural Sciences, University of Warsaw, Warszawa, Poland, and [c]Pharmaceutical Research Institute, Rydygiera 8, 01-793 Warszawa, Poland

Correspondence e-mail:
mmalinska@chem.uw.edu.pl,
pdomin@chem.uw.edu.pl

Received 10 October 2013
Accepted 31 January 2014

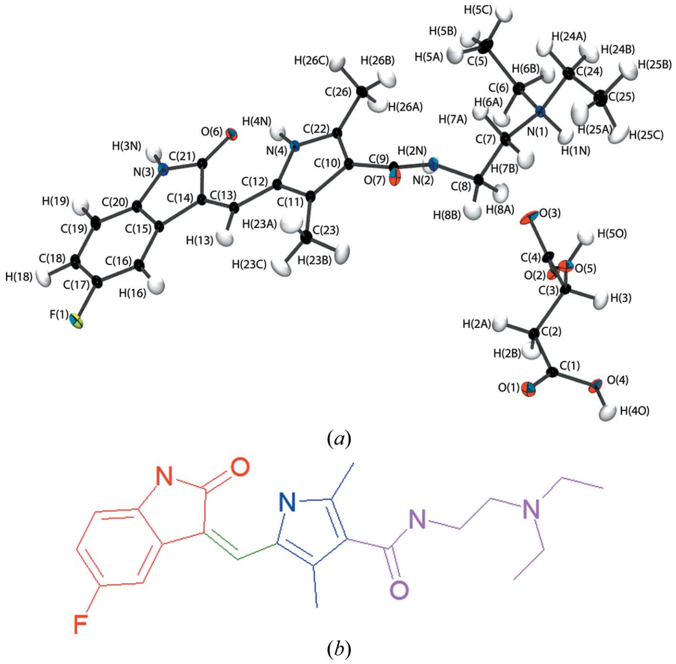
Protein kinases are targets for the treatment of a number of diseases. Sunitinib malate is a type I inhibitor of tyrosine kinases and was approved as a drug in 2006. This contribution constitutes the first comprehensive analysis of the crystal structures of sunitinib malate and of complexes of sunitinib with a series of protein kinases. The high-resolution single-crystal X-ray measurement and aspherical atom databank approach served as a basis for reconstruction of the charge-density distribution of sunitinib and its protein complexes. Hirshfeld surface and topological analyses revealed a similar interaction pattern in the sunitinib malate crystal structure to that in the protein binding pockets. Sunitinib forms nine preserved bond paths corresponding to hydrogen bonds and also to the $C—H\cdots O$ and $C—H\cdots\pi$ contacts common to the VEGRF2, CDK2, G2, KIT and IT kinases. In general, sunitinib interacts with the studied proteins with a similar electrostatic interaction energy and can adjust its conformation to fit the binding pocket in such a way as to enhance the electrostatic interactions, *e.g.* hydrogen bonds in ligand–kinase complexes. Such behaviour may be responsible for the broad spectrum of action of sunitinib as a kinase inhibitor.

1. Introduction

Sunitinib malate {*N*-[2-(diethylamino)ethyl]-5-[(*Z*)-(5-fluoro-2-oxo-1,2-dihydro-3*H*-indol-3-ylidene)methyl]-2,4-dimethyl-1*H*-pyrrole-3-carboxamide L-malate; SUM; Fig. 1} is currently approved by the US Food and Drug Administration as a drug for the treatment of advanced renal cell carcinoma and is the second-line treatment for gastrointestinal stromal tumours (Goodman *et al.*, 2007; Rock *et al.*, 2007; Sun *et al.*, 2003). Sunitinib (SU) competes with ATP for binding within the intracellular domain of receptor tyrosine kinases (RTKs; Goodman *et al.*, 2007; Rock *et al.*, 2007). These include vascular endothelial growth factor receptors (VEGFR types 1 and 2), platelet-derived growth factor receptors (PDGFR-α and PDGFR-β), a stem-cell factor receptor (KIT), FMS-like tyrosine kinase 3 (FLT3), a glial cell-line derived neurotrophic factor receptor (RET) and finally a receptor of macrophage colony-stimulating factor (CSF1R) (Mendel *et al.*, 2003; Demetri *et al.*, 2006; Osusky *et al.*, 2004; Abrams *et al.*, 2003; Murray *et al.*, 2003; O'Farrell *et al.*, 2003; Schueneman *et al.*, 2003). In the Protein Data Bank (PDB; http://www.pdb.org; Berman *et al.*, 2000), as of 14 June 2013, six crystal structures of protein kinase (PK) complexes with SU were found and were thus selected for further investigations. Four of these proteins belong to three different protein tyrosine kinase families, *i.e.* IL-2-inducible T-cell kinase (IT kinase; ITK; PDB entry 3miy; 1.67 Å resolution; Kutach *et al.*, 2010), KIT kinases

doi:10.1107/S1399004714002351

[PDB entries 3g0e (1.60 Å resolution) and 3g0f (2.60 Å resolution), where 3g0f is the D816H mutant of 3g0e (Gajiwala *et al.*, 2009); KIT mutant] and a vascular endothelial growth factor receptor 2 (VEGFR-2; PDB entry 4agd; 2.81 Å resolution; McTigue *et al.*, 2012). The two remaining proteins are serine/threonine-specific protein kinases, one of which is human phosphorylase kinase gamma 2 (G2 kinase; G2K; PDB entry 2y7j; 2.50 Å resolution; Structural Genomics Consortium, unpublished work) and the other is cyclin-dependent kinase 2 (CDK2; PDB entry 3ti1; 1.99 Å resolution; Martin *et al.*, 2012). Additionally, the KIT protein complex with adenosine diphosphate (ADP) (PDB entry 1pkg; 2.90 Å resolution; Mol *et al.*, 2003), where the protein is characterized by the same sequence as in KIT, has been deposited in the PDB and is included in the studies. All selected proteins share the typical architecture of the protein kinase-like superfamily (Hanks & Hunter, 1995; Cowan-Jacob, 2006), *i.e.* Alpha Beta 2-layer Sandwich (CATH code 3.30.200.20; N-lobe) and Mainly Alpha Orthogonal Bundle (CATH code 1.10.510.10; C-lobe). The lobes are connected by a flexible hinge region, which enables their changes in relative orientation (Fig. 2). The ATP and SU molecules bind to the cleft between the N-lobe and the C-lobe, partially surrounded by a Gly-rich loop with a conserved GXGXΦG motif (where G is a glycine residue, X is any residue and Φ is a phenylalanine or tyrosine residue), a hinge region and the activation loop (A-loop).



(*a*)

(*b*)

Figure 1
(*a*) Molecular structure and atomic labelling of SUM. Thermal displacement ellipsoids are drawn at the 50% probability level (Farrugia, 1997). (*b*) Division of the SU molecule into four fragments: the fluorinated oxindole ring (fragment 1, red), the methylene bridge (fragment 2, green), the pyrrole ring with two methyl groups (fragment 3, blue) and the amide group connected to the triethylamino fragment (fragment 4, magenta).

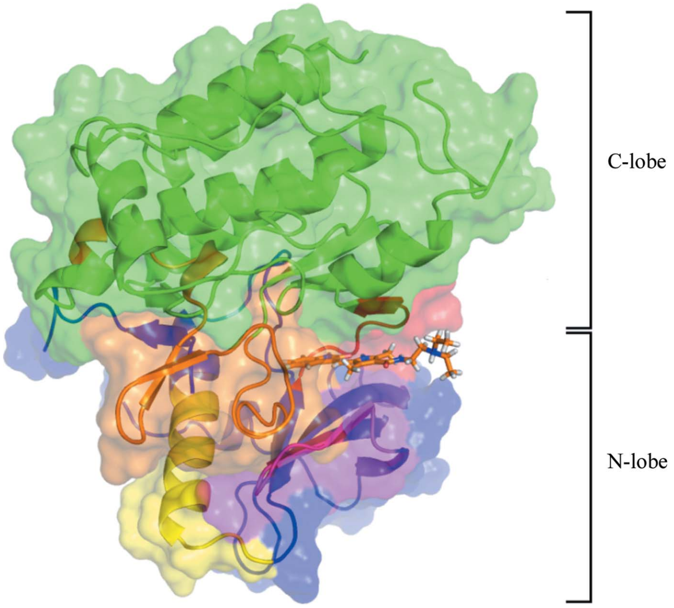


Figure 2
General features of PK architecture showing the C-lobe (green) and the N-lobe (blue). The hinge region is indicated in red, the A-loop in orange, the Gly-rich loop in magenta and the αC helix in yellow.

PKs may adopt different forms, *i.e.* an activated and inactivated form (Huse & Kuriyan, 2002). The general features of the active form are conserved owing to its catalytic function. Among others, the catalytic activity requires proper orientation of the highly conserved DFG motif (D, aspartic acid; F, phenylalanine; G, glycine), in which the side chain of aspartate is situated near the ATP-binding site (DFG_{in} conformation). The DFG triad is localized at the beginning of the A-loop and is frequently called the magnesium-binding site. Additionally, the proper orientation of the lysine side chain (αC helix) responsible for transfer of the γ-phosphorylate group of ATP is governed by the formation of a salt bridge with glutamate (β5 sheet of the N-lobe). Flipping of the DFG motif (DFG_{out} conformation) changes the ATP-binding site and exposes the additional much less conserved hydrophobic pocket (back pocket). Access to the back pocket is controlled by the size of the so-called gatekeeper residue. The competitive inhibitors of ATP bind to the same site, forming at least one hydrogen bond to the residues of the hinge region.

To obtain a deeper understanding of the interactions that are present in molecular complexes, *i.e.* beyond geometrical considerations following standard crystal structure determination, analysis of charge-density distribution is desirable. The charge density can either be computed by employing quantum-mechanical methods or be measured experimentally using high-resolution X-ray diffraction methods. To obtain an experimental charge density in a quantitative way, the multipole model (MM; Supporting Information §S1[1]; Hansen–Coppens formalism) is usually used, instead of the so-called independent atom model (IAM), to fit experimental structure factors.

The IAM is the most frequently used approach in structural refinement of X-ray diffraction data in the field of small-

---

[1] Supporting information has been deposited in the IUCr electronic archive (Reference: DZ5313).

molecule and protein crystallography. The crystal charge density in this model is treated as a sum of atom-centred fragment densities which are assumed to be neutral and spherical. Thus, the IAM does not take into account any charge transfer and deformation of the valance charge density resulting from chemical bonding, the presence of lone electron pairs, or intramolecular and intermolecular interactions. Back in 1978, Hansen and Coppens proposed a more complex, aspherical atom MM, which enables the description of aspherical charge-density features on the basis of theoretical or experimental structure factors (Hansen & Coppens, 1978). The MM provides a more accurate analytical representation of a charge-density distribution. To refine MM parameters effectively, high-resolution X-ray measurement (to at least 0.5 Å resolution) is required, as it ensures better deconvolution of thermal motion and a higher data-to-parameter ratio. This is, however, still difficult to achieve for most protein crystals. Currently, there are 349 protein structures deposited in the PDB which were measured to an X-ray resolution of <1 Å. Among these, the highest resolution data were collected for the small protein crambin (46 residues), *i.e.* to 0.48 Å resolution (Schmidt *et al.*, 2011). The first attempts to refine a 0.66 Å resolution human aldose reductase structure with aspherical structure factors have already been reported by Lecomte *et al.* (2004).

Fortunately, it has been shown that atoms with similar chemical environments usually do not differ much in terms of charge-density description. Therefore, the idea of transferable atomic multipole parameters (pseudo-atom charge-density models) between different molecules has been introduced (Brock *et al.*, 1991). These observations initiated the creation of aspherical atom databases such as ELMAM (Domagała *et al.*, 2012), UBDB (Jarzembska & Dominiak, 2012) and Invariom (Dittrich *et al.*, 2013) (see Bąk *et al.*, 2011 for a comparison of databases). UBDB, which has been used in our analysis, employs the *LSDB* program (Volkov, Li *et al.*, 2004) to assign atom types to a given molecule and transfers the MM parameters, which improves the representation of charge density with respect to the IAM. On the basis of the modelled charge density, it is subsequently possible to obtain the Coulombic interaction energy (electrostatic interaction energy; $E_{\rm es}$) using, for instance, the exact potential and multipole model (EPMM) method (Volkov, Koritsanszky *et al.*, 2004). An alternative method of protein charge-density reconstruction from transferred pseudoatom densities has also been proposed (Mebs *et al.*, 2010). In this approach, a given macromolecular structure is reduced to a substructure around an active site and the desired pseudoatom densities are then computed using the DFT approach.

Despite the importance of SU in antitumour therapy, its crystal structure, as the malate, has only recently been published (Sidoryk *et al.*, 2013). The molecular geometry provides the foundation for further energy analyses. In turn, more advanced crystallographic studies employing high-resolution X-ray measurements led to a charge-density distribution model, which then enables the investigation of some electronic effects and interactions present in crystals. A comprehensive charge-density and energy analysis provides valuable information about molecular binding properties and the nature of interactions. Therefore, in this contribution, we present an experimental charge-density study of SU malate (SUM) supplemented by energy and crystal-packing investigations. Our analysis is additionally combined with exploration of the previously described pharmaceutical aspects of SU (Faivre *et al.*, 2007). The goal was to compare the SU-mediated interactions in the studied SUM crystal and in the available complexes with proteins. The quantitative characterization of interactions in the SUM crystal structure is based on an experimental charge-density distribution model and is additionally supported by intermolecular energy calculations conducted within the *ab initio* approach (Boys & Bernardi, 1970; Simon *et al.*, 1996). However, in the case of the studied protein complexes high-resolution data are not available and *ab initio* methods are not applicable. Thus, here we utilized the aspherical atom databank (Jarzembska & Dominiak, 2012) to reconstruct the desired charge-density distribution, as on this basis it was possible to estimate the electrostatic interaction

Table 1
Experimental details for the SUM data collection.

Crystal data	
Chemical formula	$C_{26}H_{33}F_1N_4O_7$
Molecular mass (Da)	532.56
Crystal system, space group	Monoclinic, $P2_1$
Temperature (K)	90
Unit-cell parameters (Å, °)	a = 7.5531 (2), b = 8.3710 (3), c = 20.2929 (6), α = 90, β = 98.454 (2), γ = 90
V (Å^3)	1269.12 (7)
Z	2
Radiation type	Mo $K\alpha$
μ (mm^{-1})	0.10
Crystal size (mm)	0.35 × 0.17 × 0.05
Data collection	
Diffractometer	Bruker KAPPA APEX II ULTRA
Absorption correction	Multi-scan
Resolution (sinθ/λ) (Å^{-1})	1.14
No. of reflections [$I > 2\sigma(I)$]	
Measured	83597
Unique	29456
Observed	26333
$R_{\rm int}$	0.03
$\langle I/\sigma(I)\rangle$†	27 (5.8)
Multiplicity†	6.5 (6.0)
IAM refinement	
$R[F^2 > 2\sigma(F^2)]$	0.033
$wR(F^2)$	0.079
S	1.00
No. of reflections	29456
No. of parameters	349
No. of restraints	1
Absolute structure	Flack (1983)
Flack parameter	0.0 (1)
Multipole refinement	
No. of reflections [$I > 3\sigma(I)$]/parameters	24460/760
R^1/wR^1 [for $I > 3\sigma(I)$]	0.022/0.029
R^2/wR^2 [for $I > 3\sigma(I)$]	0.023/0.034
GooF [for $I > 3\sigma(I)$]	0.914
Largest residual density peak and hole	0.28, −0.24

† Values in parentheses are for the highest resolution shell (>1 Å).

energy between the chosen molecular fragments in these macromolecular systems.

2. Experimental

2.1. Data collection

Single-crystal high-resolution X-ray diffraction data collection for SUM was performed to about 0.44 Å resolution at 90 K on a Bruker AXS KAPPA APEX II ULTRA diffractometer. Indexing and integration were performed with the original Bruker APEX software (*SAINT* v.7.68A). A multi-scan absorption correction was applied in the scaling procedure using the *SORTAV* program (Blessing, 1987). Conventional spherical atom refinement was carried out with *SHELX*-97 (Sheldrick, 2008), applying the full-matrix least-squares on F^2 method for all data sets. The least-squares multipole refinement for SUM was based on F^2 with resolution limited to 0.44 Å to obtain 100% data completeness. The obtained lattice parameters and the final *R* indices are shown in Table 1. The high-resolution 90 K X-ray single-crystal diffraction experiment and the subsequent charge-density distribution modelling and analysis were carried out so as to obtain quantitative information regarding the nature of the interactions in the SUM crystal lattice. The charge-density distribution was evaluated by means of the Hansen–Coppens formalism (Hansen & Coppens, 1978). The details of the data collection and refinement are available in Table 1 and in the Supporting Information. The final Fourier difference maps are featureless (Supporting Information). The largest residual electron-density peaks, which amount to 0.28 and −0.24 e Å^{-3}, are located in the vicinity of the O(3) atom. Additionally, the thermal ellipsoid of this atom is visibly elongated.

2.2. Multipole refinement

The multipole refinement of SUM based on the Hansen and Coppens formalism (Hansen & Coppens, 1978) was accomplished using the *XD*2006 program suite (Volkov, Macchi *et al.*, 2006). Atomic positions and mean-square displacement parameters of non-H atoms were refined using the high-order X-ray diffraction data ($\sin\theta/\lambda > 0.6$). Atomic coordinates *x*, *y*, *z* and anisotropic displacement parameters (ADPs) were fixed after the high-order refinement. The C—H bond distances were fixed at the averaged distances for such bond types taken from single-crystal neutron diffraction data (Allen & Bruno, 2010). The hydrogen thermal displacement parameters were estimated using the *SHADE*2 server (Madsen, 2006; Munshi *et al.*, 2008) and then fixed. Appropriate symmetry restrictions were applied to the population of multipoles for selected atoms (Supplementary Table S3). For the F and O atoms multipoles up to hexadecapole level were refined during the refinement procedure, and up to the octupole level for the remaining non-H atoms. For the H atoms, only monopoles, bond-directed dipoles and bond-directed quadrupoles were used. Kappa parameters were refined for non-H atoms when all multipoles had already been refined. Some restrictions for the κ and κ' parameters were applied; for further details, see Supplementary Table S3. In the case of H atoms, the κ and κ' parameters were fixed at the recommended values (Volkov *et al.*, 2001). After these steps, the full-matrix refinement of atomic positions, ADPs for the non-H atoms and population of multipoles was conducted. The refinement was performed in a stepwise manner and any step was considered as complete when full convergence was reached, *i.e.* when the maximum shift/s.u. (standard uncertainty) ratio after a given refinement cycle was lower than 10^{-5}. The Hirshfeld rigid-bond test (Hirshfeld, 1976) was applied after each step. The DMSDA (differences of mean-squares displacement amplitudes) values were all lower than 0.0010 Å^2.

2.3. AIM integrated atomic charges

All calculations of the integrated properties were performed using the *TOPXD* program from the *XD*2006 package (Volkov, Macchi *et al.*, 2006). All grid parameters are presented in the Supporting Information (Supplementary Table S9). An error in the integrated atomic Lagrangian of lower than 10^{-3}–10^{-4} indicates acceptable accuracy of the numerical integration. Integration of the electron density over topological atomic basins (Bader, 1994) was performed for all atoms of SUM. The sum of charges integrated over the atomic basins in both molecules is close to zero and amounts to 0.001. The individual atomic contributions are given in Supplementary Table S9. The error in the calculated cell volume is 0.8 Å^3, which amounts to 0.06% of the unit-cell volume. The properties of covalent and interaction bond critical points (BCPs) are reported in Supplementary Tables S7 and S8, respectively.

2.4. Computations

Theoretical analysis of the analysed molecular dimers was carried out in *Gaussian* 09 (Frisch *et al.*, 2009) using DFT within the supermolecular approach. For the DFT calculations, the B97D (Chai & Head-Gordon, 2008), wB97XD (Grimme, 2006) and M06 (Wood *et al.*, 2006) functionals were used. Several basis sets were tested, with the aim of reducing the basis-set superposition error. Finally, the 6-311G** basis set (Frisch *et al.*, 1984; McLean & Chandler, 1980) was applied and the reported interaction energies were calculated by the standard counterpoise method (Boys & Bernardi, 1970; Simon *et al.*, 1996).

Energy scans over the torsion angle C(11)—C(10)—C(9)—O(7) were performed using 18 steps of 10° starting from the final geometry retrieved from the crystal structure at the B97D (Chai & Head-Gordon, 2008)/6-311G** (Frisch *et al.*, 1984; McLean & Chandler, 1980) level of theory. After each change of the torsion angle, constrained optimization of all other geometrical parameters in the isolated SU conformers was performed.

2.5. Electrostatic calculations (dimers)

Two types of calculations were performed to obtain $E_{\rm es}$ for the selected dimers. Firstly, the EPMM method (Volkov, Koritsanszky *et al.*, 2004) was applied, which allows computation of $E_{\rm es}$ between two molecular charge distributions represented within the Hansen–Coppens electron-density formalism (Hansen & Coppens, 1978). It combines a numerical evaluation of the exact Coulomb integral for short-range interatomic interactions (less than 4.5 Å) with a Buckingham-type multipole approximation for the long-range contacts. Secondly, the exact $E_{\rm es}$ was obtained by virtue of the *SPDFG* program (Volkov, King *et al.*, 2006). $E_{\rm es}$ calculations were performed on the basis of wavefunctions obtained with the *Gaussian* 09 program (Frisch *et al.*, 2009) using the B97D (Chai & Head-Gordon, 2008) functional with the 6-311G** (Frisch *et al.*, 1984; McLean & Chandler, 1980) basis set.

2.6. Preparation of protein structures

For all of the analysed PDB structures, H atoms were added using the *Reduce* program (Word *et al.*, 1999) and water molecules were then removed. The orientations of OH, SH, NH_3^+, Met CH_3 and the side chains of Asn, Gln and His were optimized. Arg, Lys, Asp and Glu residues were treated as ionized. X—H bond lengths were extended to the standard neutron diffraction values (Allen & Bruno, 2010) and fixed in the case of all performed calculations.

2.7. Hirshfeld surface analysis

For the purpose of our investigations, the amino acids closest to the SU molecule (6 Å) were separated from the protein structures. The percentage contributions to the Hirshfeld surface from all types of interactions for the SUM crystal and the VEGRF2, CDK2, ITK (chain *A*), KIT and G2K (chain *A*) complexes are 99, 36, 57, 38, 63 and 43%, respectively. For comparative purposes, all percentage contributions to the Hirshfeld surface were scaled to 100%.

2.8. Electrostatic calculations (ligand–protein complexes)

Pseudoatom databanks allow reconstruction of the electron density of macromolecular systems for which experimentally derived geometries are available. In this study, we used the UBDB (Jarzembska & Dominiak, 2012) together with the *LSDB* program to transfer the multipole parameters of the atom types stored in the UBDB to the studied protein–SU complexes and the protein–ADP complex. The two missing atom types from the SU moiety, *i.e.* the quaternary amine-group N and H atoms, were modelled and added to the databank to enable the analysis. All amino-acid residues and SU molecules were scaled independently to their formal charges after the databank transfer. In the case of the analysed KIT kinases, the triethylamino group is highly disordered and thus is not localized in the structures. The scaled multipole parameters for the SU atoms found in KIT and KIT mutant structures were transferred from the VEGFR2 complex. None of the 24 selected residues discussed in the presented paper is disordered.

The EPMM method was used to compute electrostatic interactions. The Buckingham approximation was employed

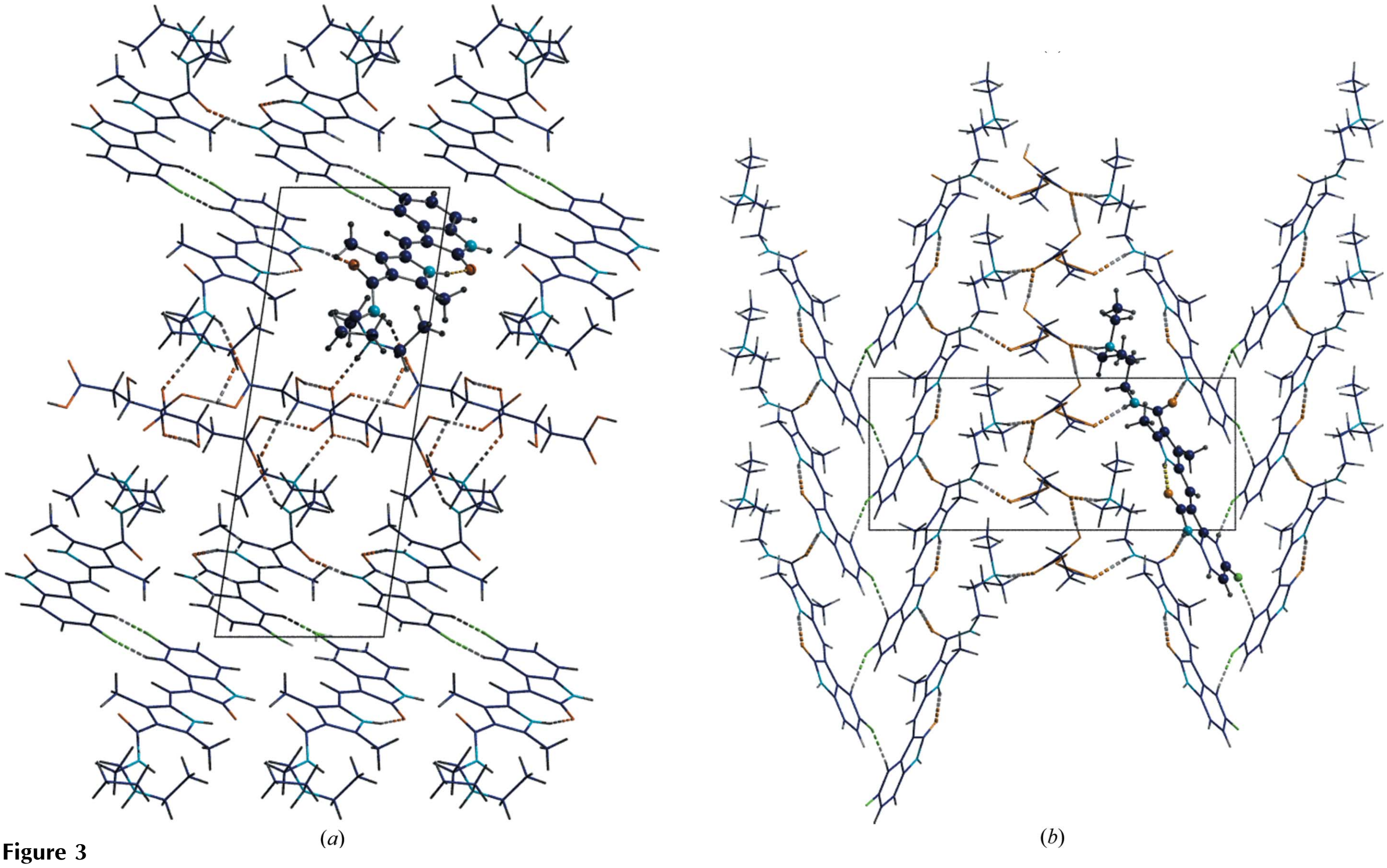
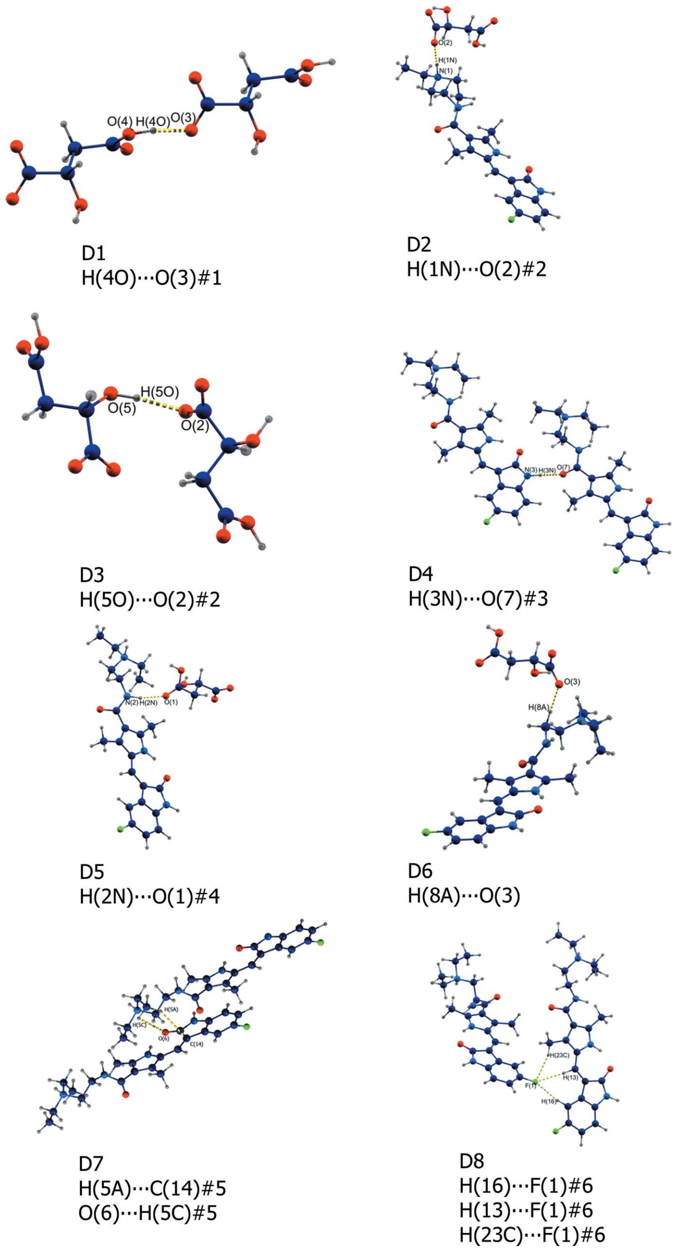


Figure 3
Crystal packing of SUM molecules: (*a*) view along the *y* axis, (*b*) view along the *x* axis. Hydrogen bonds and C—H···F contacts are shown as dotted lines.

Table 2
Selected HB interactions from the SUM crystal structure together with the donor–acceptor distance $R(\mathrm{D}\cdots\mathrm{A})$ and the corresponding values of charge density $\rho(\mathbf{r})$ and its Laplacian $\nabla\rho(\mathbf{r})$ at BCPs for hydrogen bonds found between an H atom and an acceptor atom.

E stands for the interaction energy calculated using the empirical formula proposed by Espinosa *et al.* (1998).

		$R(\mathrm{D}\cdots\mathrm{A})$ (Å)	$\rho(\mathbf{r})$ (e Å^{-3})	$\nabla^2\rho(\mathbf{r})$ (e Å^{-5})	E (kcal mol^{-1})
D1	O(4)—H(4O)···O(3)#1	2.4930 (9)	0.44 (8)	4.2 (3)	−24
D2	N(1)—H(1N)···O(2)#2	2.7577 (5)	0.28 (5)	2.2 (1)	−12
D3	O(5)—H(5O)···O(2)#2	2.763 (3)	0.23 (6)	0.2 (1)	−7
D4	N(3)—H(3N)···O(7)#3	2.808 (2)	0.21 (5)	1.5 (1)	−7
D5	N(2)—H(2N)···O(1)#4	2.945 (2)	0.18 (1)	1.0 (1)	−5
D6	C(8)—H(8A)···O(3)	3.319 (1)	0.10 (3)	1.2 (1)	−3
Intra	N(4)—H(4N)···O(6)	2.682 (3)	0.25 (4)	3.29 (7)	−11



D1
H(4O)···O(3)#1

D2
H(1N)···O(2)#2

D3
H(5O)···O(2)#2

D4
H(3N)···O(7)#3

D5
H(2N)···O(1)#4

D6
H(8A)···O(3)

D7
H(5A)···C(14)#5
O(6)···H(5C)#5

D8
H(16)···F(1)#6
H(13)···F(1)#6
H(23C)···F(1)#6

Figure 4
Selected dimer motifs extracted from the SUM crystal structure.

in order to estimate the electrostatic potential at distances greater than 4.5 Å. As VEGRF and ITK consist of four and two independent protein chains in the asymmetric unit, respectively, the chains were analysed separately owing to their different SU arrangements. In these cases, the EPMM calculation results obtained for different chains were finally averaged.

2.9. Electrostatic potential analysis

All electostatic potentials (ESPs) were calculated using the *XDPROP* (Volkov, Macchi *et al.*, 2006) program and visualized in *PyMOL* (v.1.3r1; Schrödinger). The charge-density distribution for all studied kinases was reconstructed with the aid of the UBDB, as was also performed for the electrostatic energy calculations (Jarzembska & Dominiak, 2012). For the purpose of comparison, the missing residues were deleted to obtain identical sequences. The terminal residues were completed by H atoms or methyl groups to achieve chemically sensible groups and a neutral formal charge of residues.

3. Results and discussion

3.1. Quantification of interactions present in the SUM crystal

SUM crystallizes in the monoclinic space group $P2_1$ with two moieties in the asymmetric part of the unit cell, *i.e.* the SU cation containing a protonated quaternary amine group and the malic acid anion (Fig. 1). The crystal molecular packing seems to be governed by electrostatic interactions among molecular ions together with an extensive net of hydrogen bonds (HBs; Supplementary Tables S1 and S2; Sidoryk *et al.*, 2013). Some weaker intermolecular interactions are present, mainly between the SU molecules (Fig. 3). Nevertheless, pure geometrical information is often not sufficient to draw meaningful conclusions on the relative importance of the mentioned interactions.

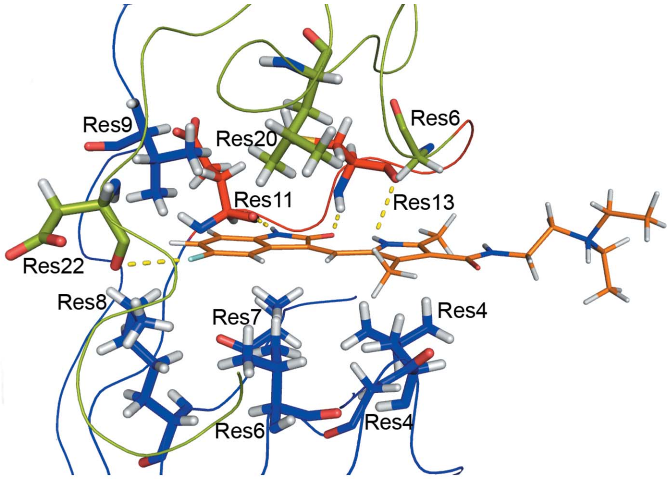
Fortunately, it was possible to collect SUM single-crystal X-ray diffraction data to 0.44 Å resolution and, on this basis, to derive a physically reasonable model of the SUM electron-density distribution (for details, see Supporting Information). In order to better quantify hydrogen bonds and other weak interactions, we analysed the obtained charge density of SUM in terms of the quantum theory of atoms in molecules (QTAIM; Bader, 1994). We used the values of the electron density, $\rho(r)$, and its Laplacian, $\nabla^2\rho(r)$, at bond critical points (BCPs) found between atoms forming hydrogen bonds (Espinosa *et al.*, 1998; Abramov, 1997; see Supporting Information). According to the results (Table 2), the strongest hydrogen bond occurs between the H(4O) and O(3)#1 atoms[2] (Fig. 4, dimer D1), both belonging to malate anions, with an interaction energy equal to −24 kcal mol^{-1}. The origin of its strength can be explained by the phenomena discussed in the literature on the topic of strong hydrogen bonds, *i.e.* charge-assisted hydrogen bonds (CAHBs; Gilli *et al.*, 2004; Góra *et al.*,

---

[2] Symmetry operators: #1, $x+1, y, z$; #2, $-x+1, y+1/2, -z+1$; #3, $x-1, y-1, z$; #4, $x-1, y, z$; #5, $x, y+1, z$; #6, $-x+1, y-1/2, -z+2$; #7, $-x+1, y+1/2, -z+2$.

2013) and resonance-assisted hydrogen bonds (RAHBs; Gilli *et al.*, 1989; Bertolasi *et al.*, 1991). In the case of D1, the negative charge on the proton-accepting carboxylate group increases the strength of this (−)CAHB remarkably. The second strongest intermolecular hydrogen bond, H(1N)$\cdots$O(2)#2, occurs between SU and malate molecules. It was found to be similar in strength to the intramolecular H(4N)$\cdots$O(6) bond present in SU, −12 and −11 kcal mol^{-1} for intermolecular and intramolecular HBs, respectively. However, these interactions differ in nature: the former connects the charged donor and acceptor groups [(±)CAHB], whereas the latter hydrogen bond involves the conjugated double-bond system (RAHB; Gilli *et al.*, 1989; Bertolasi *et al.*, 1991). The remaining hydrogen bonds are noticeably weaker, although still significant, with an average interaction energy equal to −6 kcal mol^{-1}.

In a more global approach to intermolecular interactions, the interaction energy is computed on the basis of the charge-density distribution of the whole molecule, *i.e.* not restricted solely to local descriptors such as properties of BCPs. Such an approach considers not only contributions from hydrogen bonding but also from long-range electrostatic interactions present between any molecular charge-density fragments. Having the experimental charge-density distribution of a crystal, we are able to compute the electrostatic part of the total interaction energy for any molecular complex present in the crystal lattice (Fig. 4). The attractive electrostatic interactions are naturally found in the heterodimers of SU and malate ions, and are especially strong in these linked together *via* well oriented and relatively strong hydrogen bonds, such as D2 (Table 3). All homodimers are characterized by repulsive electrostatic interactions; however, for the dimer with the strongest hydrogen bonding, D1, the electrostatic repulsion is least pronounced. Generally, attractive and repulsive electrostatic energy components tend to cancel out. Nevertheless, the resulting electrostatic lattice energy for the SUM crystal is stabilizing and amounts to −235 kcal mol^{-1}.



(*a*)

```
Res               1   5   6    7   8    9   10      17   19   20  21  24
VEGFR2-4AGD:(837)KPLGR(5)V(17)A(1)K(30)V(16)VEFCKFGN(1)ST(108)L(9)CDFG(1049)
CDK2  -3TI1:(  7)EKIGE(5)V(12)A(1)K(30)V(15)FEFLHQ-D---LK( 55)L(9)ADFG(148)
ITK   -3MIY:(366)QEIGS(5)V(11)A(1)K(27)V(15)FEFMEHGC(1)SD( 43)L(9)SDFG(504)
G2K   -2Y7J:( 27)DVIGR(6)V(12)A(1)K(37)I(16)FDLMRKGE(1)FD( 42)L(9)SDFG(174)
KIT   -3G0E:(592)KTLGA(5)V(17)A(1)K(30)V(15)TEYCCYGD(1)LN(118)L(9)CDFG(813)
MutKIT-3G0F:(592)KTLGA(5)V(17)A(1)K(30)V(15)TEYCCYGD(1)LN(118)L(9)CDFG(813)
                    :*    *    *   *    :     :    :           *    ***
```

(*b*)

Figure 5
(*a*) SU molecule in the binding pocket of VEGFR2 surrounded by conserved amino acids. Res23 and Res24 are omitted for clarity. Short contacts are marked as yellow dotted lines. The H$\cdots$H interactions are omitted for clarity. The protein chain is shown as a ribbon diagram. Green, red and blue colours indicate the C-lobe, N-lobe and hinge regions according to Fig. 2. (*b*) Structure-based sequence alignment of the binding pocket. Asterisks indicated identical amino acids and colons indicate conserved amino acids. The first line represents the numbering scheme of residues within the binding pocket (Res1–Res24)

Table 3
Total interaction energy (E) and electrostatic interaction energy (E_{es}) obtained for selected dimers extracted from the SUM crystal structure.

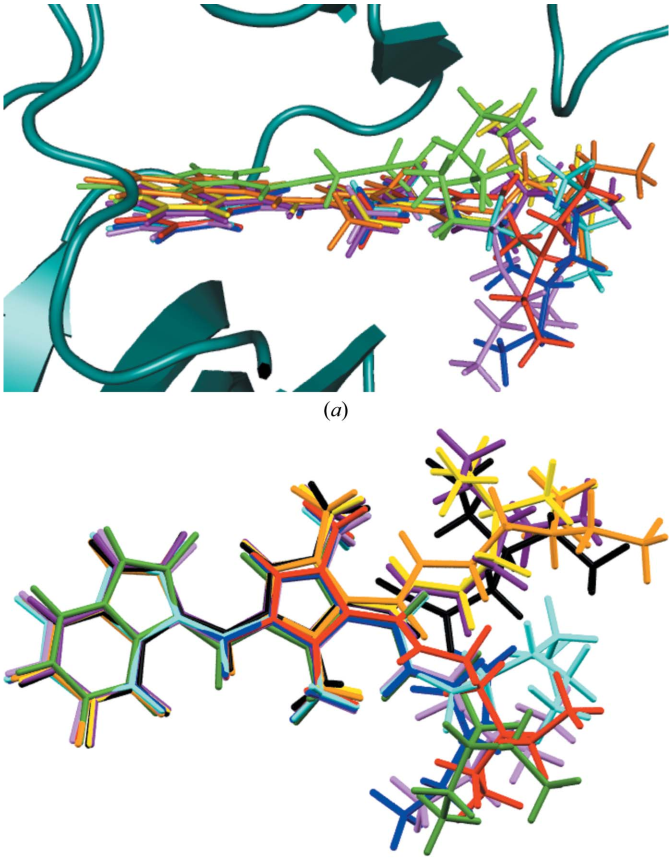
		E_{es}† (kcal mol^{-1})	E_{es}‡ (kcal mol^{-1})	E_{es}§ (kcal mol^{-1})	E¶ (kcal mol^{-1})
D1	H(4O)$\cdots$O(3)#1	18	10	15	29
D2	H(1N)$\cdots$O(2)#2	−89	−83	−87	−96
D3	H(5O)$\cdots$O(2)#2	50	42	49	47
D4	H(3N)$\cdots$O(7)#3	17	13	17	15
D5	H(2N)$\cdots$O(1)#4	−58	−72	−63	−64
D6	H(8A)$\cdots$O(3)	−65	−55	−64	−75
D7	O(6)$\cdots$H(5C)#5	34	35	30	17
D8	H(16)$\cdots$F(1)#6	19	18	17	16

† E_{es} obtained with the exact potential (Volkov, King *et al.*, 2006) from monomer charge distributions expressed in terms of Gaussian-type basis functions (B97D/6-311G**). ‡ E_{es} obtained with the EPMM method (Volkov, Koritsanszky *et al.*, 2004) between two molecular charge distributions within the Hansen–Coppens model (Hansen & Coppens, 1978) refined against experimental data. § E_{es} obtained with the EPMM method between two molecular charge distributions reconstructed with the aid of the UBDB. ¶ Interaction energy obtained by the supermolecular method with basis-set superposition error (BSSE) correction (Boys & Bernardi, 1970; Simon *et al.*, 1996) for gas-phase dimers [B97D (Chai & Head-Gordon, 2008)/6-311G** (Frisch *et al.*, 1984; McLean & Chandler, 1980].

Currently, it is impossible to reliably estimate total interaction energy on the basis of the experimental charge-density distribution. Therefore, to check how the electrostatic energy relates to the total interaction energy for the analysed dimers, we carried out quantum-mechanical calculations for molecular dimers in the gas phase (Fig. 4). A linear correlation is observed between the total energy and the electrostatic contribution. For the purpose of the present study, two other E_{es} calculations were additionally performed: the first using EPMM applied to charge densities of SU and malate molecules reconstructed on the basis of the UBDB, and the second using the quantum-mechanical supermolecular approach (Table 3). All three methods gave similar E_{es} values. All of the above results indicate that the pseudoatom databank approach, followed by topological analysis and EPMM calculations, may provide very useful information about the interactions present in protein complexes of SU.

3.2. Geometrical aspects of the interaction of SU with protein kinases

SU belongs to the type I inhibitors (Zuccotto *et al.*, 2010; Liu & Gray, 2006), which are characterized by a broad spectrum of action, *i.e.* they can inhibit PKs in the active and the inactive form. Currently, more than 70% of mammalian kinase structures stored in the PDB are in the DFG$_{in}$ conformation and 22% constitute intermediate structures (Kufareva &


(*a*)

(*b*)

Figure 6
(*a*) The result of superposition of SU molecules (drawn as lines) from all protein complexes studied and one selected protein (VEGFR2, drawn as a cartoon representation). (*b*) Overlay of SU molecules retrieved from the SUM crystal structure (black) and protein–ligand complexes of ITK (yellow and purple), VEGFR2 (orange), G2K (violet, blue, cyan and red) and CDK (green) proteins.

Abagyan, 2008). The SU–kinase complexes analysed in this study have various orientations of the DFG triad and represent both forms. VEGFR2, KIT and KIT mutant adopt the DFG$_{\rm out}$ conformation, whereas the CDK2, G2K and ITK kinases have the DFG$_{\rm in}$ conformation similar to the active state of kinases.

Structural alignment of all kinases discloses 24 residues that form the SU binding pocket (Fig. 5*a*), which contains eight identical and four conserved amino acids (Fig. 5*b*, Supplementary Table S20). To ease further analysis, these 24 residues have been numbered Res1–Res24, and this numbering scheme is used during discussion of the results. The backbone of Res13 (cysteine/methionine/lysine) forms two hydrogen bonds to the SU molecule in all studied protein complexes. Therefore, Res13, despite its variability, was treated as a conserved residue. Res22, Res23 and Res24 form the highly conserved DFG motif.

The fluorinated oxindole ring, methylene bridge and pyrrole ring of SU (Fig. 1*b*, fragments 1, 2 and 3) are enclosed in the protein binding pocket, while the aliphatic diethylaminoethyl tail lies on the protein surface and its spatial position differs from one protein to another (Fig. 6*a*). Two main conformers with different C(11)—C(10)—C(9)—O(7) torsion angles can be distinguished (Fig. 6*b*). The first is common to the SUM crystal and the ITK and VEGFR protein–SU complexes (with a torsion angle from -42° to 0°) and the second to the CDK2 and G2K complexes (-160° to -140°). These two conformers correspond well to two minima found after the energy scan over the torsion angle at -166° and -12°, for which the energy difference amounts to 1.4 kcal mol^{-1} in favour of the former. The energy difference is small, but the gas-phase rotation barrier amounts to 20 kcal mol^{-1}. Nevertheless, such a barrier might be overcome by the formation of favourable interactions.

The interactions of SU with the protein surroundings in complexes of SU with kinases were quantified *via* Hirshfeld surface analysis (Hirshfeld, 1976; Spackman & Jayatilaka, 2009) and compared with the corresponding results obtained for SU molecules retrieved from the SUM crystal lattice (Fig. 7 and Supplementary Fig. S3). Additionally, one of the natural products of protein kinases, ADP, as well as the counterion of SU in the SUM crystal structure, malate, were analysed.

It appears that H···H interactions are the dominating interactions, which is very common for crystals of organic molecules (Fig. 7). SU is surrounded by aliphatic residues forming van der Waals (vdW) contacts in the protein binding pocket. These are the residues located above and below the fluorinated oxindole ring, *i.e.* Res3, Res4, Res6, Res7, Res16 and Res20. They interact

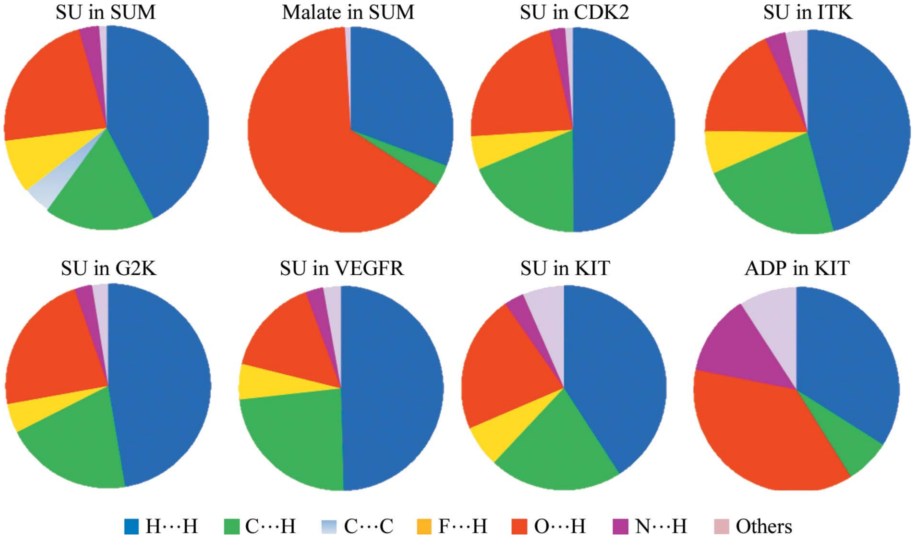


Figure 7
Proportional pie diagram with percentage contributions of H···H (blue), C···H (green), C···C (light blue), F···H (yellow), O···H (red) and N···H (purple) contacts and other (pink) interactions to the Hirshfeld surface found between SU, malate and ADP in SUM, CDK2, ITK (chain *A*), G2K (chain *A*), VEGFR2 and KIT.

via H$\cdots$H and C—H$\cdots$O contacts that correspond to the D7 dimer in the SUM crystal lattice. The next most abundant contacts result from hydrogen bonding (35% of the area of the SU Hirshfeld surface). SU forms the shortest hydrogen bonds to Res11 and Res13. The former corresponds to the H(3N)$\cdots$O(7)#3 interaction in the case of the SUM crystal structure, whereas the latter corresponds to the O(6)$\cdots$H(7B)—C(7) interaction, both of which are present in D4 (Fig. 8).

The relative contribution of particular contacts for ADP bound to the active form of KIT (Fig. 7) clearly differs from the corresponding result for SU. Owing to the different conformations of the KIT molecule when ADP is bound, and also to the different character of the ADP moiety itself with respect to SU (and the different molecular mutual orientations in the active site), the surroundings of ADP molecules are distinct from those of SU. Generally, as the ADP molecule is larger than SU, it is also more tightly packed in the protein binding pocket. Naturally, the most efficient packing is present in the SUM crystal structure.

3.3. SU$\cdots$PK interaction from the charge-density perspective: QTAIM

Reconstruction of SU–protein charge densities with the aid of the aspherical atom databank enables QTAIM analysis similar to that performed for the SUM crystal. In all of the studied protein complexes, intermolecular bond paths were found between SU and eight amino acids (Res3, Res10, Res11, Res12, Res13, Res14, Res16 and Res20) located at the beginning of the Gly-rich loop and at the hinge region with the gatekeeper residue (Res10; see Supporting Information). Taking a closer look at the mentioned interactions, it appears that bond paths exist between the same atoms for all but the Res10 and Res12 residues mentioned above (Fig. 8, Table 4). There are two C—H$\cdots\pi$-type and one O$\cdots$H—C-type conserved interactions with Res3. Next, there are hydrogen-bond contacts between SU and the residues located at the hinge region. The backbone atoms of Res11 and Res13 act as acceptors and donors for hydrogen bonds. Furthermore, Res14 is an acceptor of a weak hydrogen bond of the O$\cdots$H—C type, while Res16 and Res20 form C—H$\cdots\pi$ interactions with the SU molecule. Beside the above conserved bond paths, paths characteristic for particular SU–protein complexes were found depending on the surrounding amino acids and the conformation of the SU molecule.

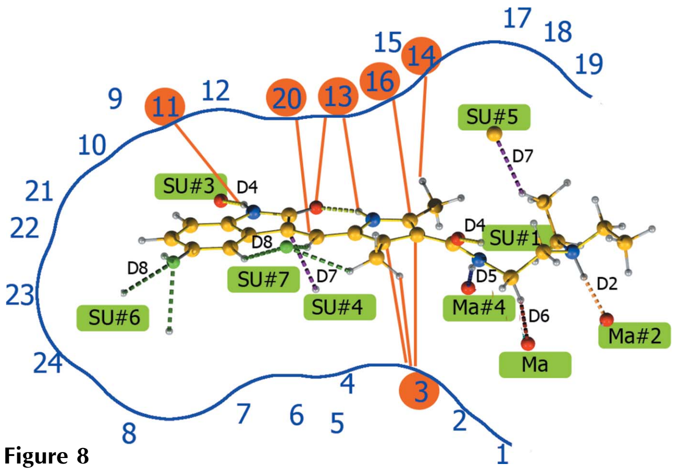


Figure 8
Schematic picture of SU contacts from the SUM crystal structure (dotted lines with dimer number D1, D2, D3, *etc.*) and the 24 selected corresponding amino acids (Res1–Res24; Res description omitted for clarity) from the protein binding pocket. Common bond paths for all the studied complexes are pinpointed as solid orange lines. Abbreviations used: SU, sunitinib; Ma, malate anion.

The aforementioned hydrogen bonds to Res11 and Res13 are the canonical contacts described in the literature as being the most important for ATP/ADP or inhibitor binding. According to the values of interaction energies obtained using the approach of Espinosa *et al.* (1998), these three hydrogen bonds are the strongest. The interaction energy for the hydrogen bond to Res11 amounts to ~−6 kcal mol^{-1} (with the exception of G2K), which corresponds well to the value obtained for the same type of interaction in D4 in the SUM crystal lattice. The interaction energy with Res13 varies slightly among the protein complexes, and is replaced by a weaker O$\cdots$H—C contact (D4) in the SUM crystal. After summation of the interaction energies of fragments 1–3 of SU, the interaction energy is similar for all of complexes studied (−35 kcal mol^{-1} on average). Interaction strengths with the remaining fragment 4 of SU also seem to be similar, with the exception of the CDK2 complexes. This is mainly owing to stronger hydrogen bonds to Res3 and an additional very strong hydrogen bond to Res17 characterized by an interaction energy of −22 kcal mol^{-1}. The greater number of bond paths in the CDK2–SU complex corresponds well to the highest percentage of the Hirshfeld surface covered by all types of interactions (57%). This value is two times greater than in the case of, for example, VEGFR2. Adoption of the less favourable conformation of the C(11)—C(10)—C(9)—C(7) torsion angle is compensated by the formation of a stronger hydrogen bond with Res3 in the G2K and CDK2 complexes and with Res17 in CDK2. The above analysis shows that major contacts can easily be detected using the QTAIM (Bader, 1994) approach for ligand–protein complexes and are supported by the first rough estimation of interaction energy.

3.4. SU$\cdots$PK interaction from the charge-density perspective: EPMM

The comparative part of electrostatic energy analysis is focused mainly on the fluorinated oxindole ring, methylene bridge and pyrrole ring of SU (fragments 1–3; Table 5), owing to the lack of coordinates for the aliphatic diethylaminoethyl tail (fragment 4; Fig. 1*b*) for the KIT and KIT mutant crystals. However, the most interesting issues concerning the interactions with the tail fragment are additionally discussed.

Hydrogen bonds formed by the sunitinib molecule with the backbone of Res11 and Res13 are, as already mentioned, a characteristic feature of competitive inhibitors of ATP (Toledo *et al.*, 1999; Noble *et al.*, 2004). Interestingly, the electrostatic energy of Res11 and Res13 with fragments 1–3 of

Table 4
Interaction energy calculated using the empirical formula proposed by Espinosa *et al.* (1998) for bond paths common to all SU–kinase complexes.

For proteins with more than one molecule in the asymmetric unit, the energy results were averaged.

Residue	Bond path PK$\cdots$SU	CDK2 (3ti1)	ITK (3miy)	G2K (2y7j)	VEGFR2 (4agd)	KIT (3g0e)	KIT mutant (3g0f)
3	$C-H\cdots\pi$	−1	−2	−3	−2	−1	−1
3	$C-H\cdots\pi$	−1	0	−1	−2	−1	−1
3	$O\cdots H-C$	−7	−1	−4	0	−2	−1
11	$O\cdots H-N$	−5	−4	−13	−6	−5	−5
13	$N-H\cdots O$	−7	−14	−9	−6	−12	−13
13	$O\cdots H-N$	−4	−1	−4	−2	−3	−1
14	$O\cdots H-C$	−3	−2	−1	−1	−1	0
16	$C-H\cdots\pi$	—	−2	−1	−1	−1	−1
20	$C-H\cdots\pi$	−2	−1	−1	−1	−1	−1

Table 5
Selected electrostatic interaction energies for the SU fragments 1–3 (according to Fig. 1*b*) with the closest amino acids (Res1–Res24).

The structure-based sequence alignment is as listed in Fig. 5(*b*). For further details, see Supplementary Tables S21–S35. For proteins with more than one molecule in the asymmetric unit the energy results were averaged. Phe (Res23) from the DFG motif is differently located in space depending on the protein conformation.

Residue	Conserved residues	CDK2 (3ti1)	ITK (3miy)	G2K (2y7j)	VEGFR2 (4agd)	KIT (3g0e)	KIT mutant (3g0f)
1		−10	−1	−10	9	11	11
2		8	−8	−1	−1	−1	−1
3	Ile/Leu	−3	−4	−6	−6	−6	−5
4	Gly	1	1	1	1	1	1
5		−8	0	8	7	0	−1
6	Val	0	0	−1	−1	−3	−3
7	Ala	0	−1	0	0	1	1
8	Lys	7	4	6	5	10	7
9	Val/Ile	0	1	1	1	1	1
10		−4	−2	−2	0	−2	−1
11	Glu/Asp	−23	−21	−30	−24	−22	−23
12		0	0	0	0	−1	−1
13		−11	−16	−16	−12	−16	−16
14		−3	−11	10	8	0	0
15		0	−1	7	−1	−1	−1
16	Gly	—	0	0	1	0	0
17		−13	−1	−13	2	−11	−11
18		0	0	0	0	−2	−3
19		7	−11	−10	1	1	1
20	Leu	−2	−1	−1	−1	0	0
21		0	0	0	0	−2	−3
22	Asp	−9	−5	−6	−5	−6	−6
23	Phe	0	0	0	−2	0	0
24	Gly	0	0	0	0	0	0
$\sum$(1–24)	Fragment 1–3	−61	−75	−60	−16	−48	−51
$\sum$(1–24)	Fragment 1–4	−156	−162	−117	49	—	—

SU is very similar for almost all of the studied kinases and amounts to ∼−23 and ∼−15 kcal mol^{-1} for Res11 and Res13, respectively. G2K is the only protein which forms a somewhat stronger interaction with Res11 (−30 kcal mol^{-1}). This is mainly owing to the significantly more advantageous geometry of the glutamate (Res11)–SU contact in G2K in comparison to the remaining complexes.

In general, the electrostatic energy of interaction between SU and an amino-acid residue is greatest when the residue is negatively charged or when it forms HBs, which are surely mostly electrostatic in nature. Despite the abovementioned cases of Res11 and Res13, such interactions are observed for the conserved aspartate (Res22, −6 kcal mol^{-1}) or the unconserved aspartate or glutamate at positions Res1, Res2, Res5, Res14, Res17 or Res19 (∼−11 kcal mol^{-1}). Obviously, the opposite situation is found for positively charged residues: the conserved lysine at Res8, (∼6 kcal mol^{-1}) and the unconserved lysine and arginine at Res1, Res2, Res5, Res14, Res15 and Res19 (∼9 kcal mol^{-1} per residue).

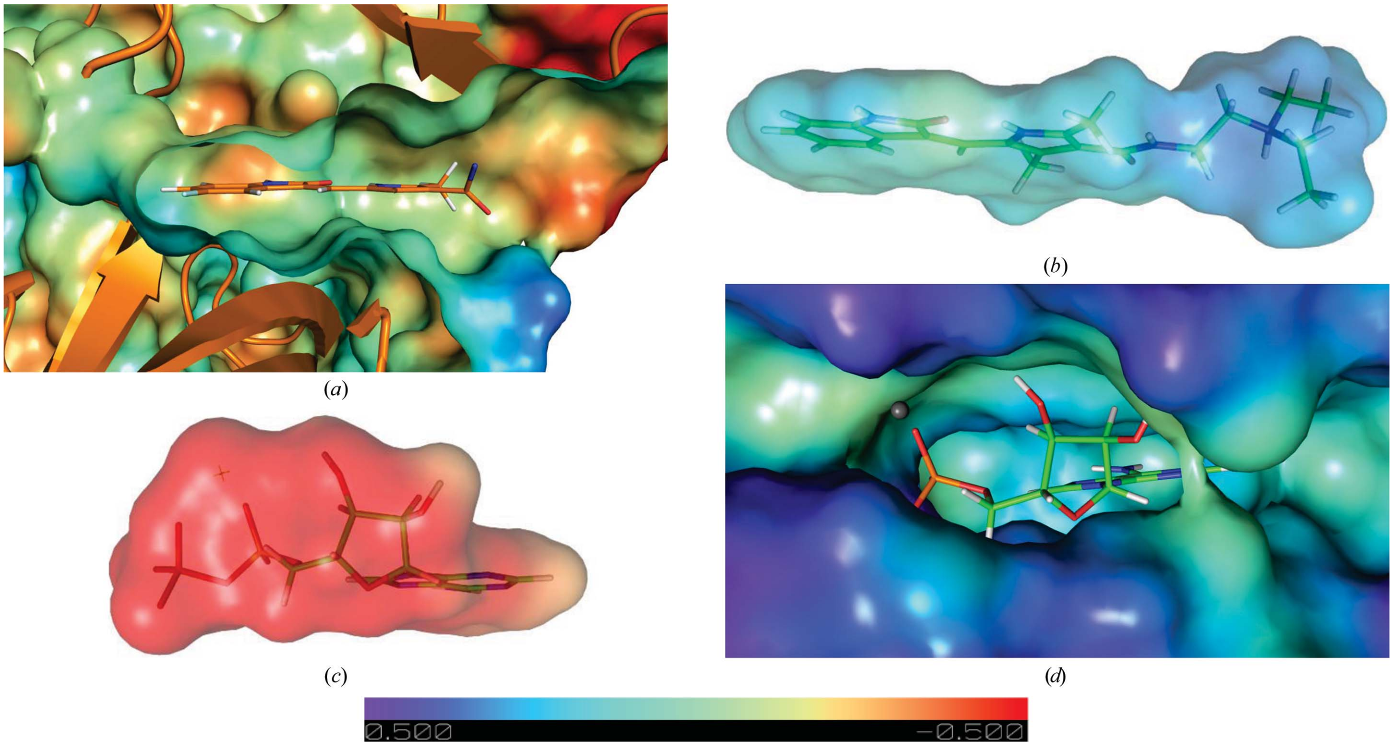
The protein–ligand complexes are further stabilized mostly through van der Waals interactions with the conserved valine (Res6), alanine (Res7), glycine (Res4 and Res16) and leucine/isoleucine (Res3, Res20) residues and the unconserved Res12, Res18 and Res21. Electrostatic contributions to these interactions are negligible, except for Res3 and Res10, and does not exceed ±1 kcal mol^{-1}. The electrostatic contribution to the interaction between SU and Res3 is the most advantageous among the vdW contacts (*i.e.* ∼−5 kcal mol^{-1}) owing to the $C-H\cdots O$ interaction. Electrostatic contributions to Res10$\cdots$SU interaction range from 0 to −4 kcal mol^{-1} and are of stabilizing character.

It is found that the leucine/isoleucine (Res3)$\cdots$SU interactions can be even more significant when fragment 4 of SU is considered (see Supporting Information). This is the case for the CDK2 and G2K crystal structures, in which the strength of this particular interaction is related to the conformation of the SU amide-bond region, namely to the previously analysed torsion angle $C(11)-C(10)-C(9)-O(7)$. The less advantageous SU conformation present in these two protein–ligand complexes, with the $N(2)-H(2N)$ bond pointing towards the leucine/isoleucine carbonyl O atom, enables the formation of a relatively strong hydrogen bond to the protein. However, it has previously been shown in the topological analysis, in which the bond path to Res3 is present for all the studied complexes, that $E_{\rm HB}$ is characterized by the greatest stabilizing character for these two complexes (∼−10 kcal mol^{-1} compared with ∼−3 kcal mol^{-1} for the rest of the complexes). The opposite orientation of the amide group leads to slightly positive electrostatic energy values obtained by the EPMM method. Additionally, the negatively charged amino acids stabilize the SU complexes. These are the aspartate/glutamine residues (Res17 and Res19). Among such types of contact, the aspartate (CDK2)–fragment 4 interaction is the strongest (−79 kcal mol^{-1}). Likewise, the Espinosa approach indicated this interaction as the strongest (−22 kcal mol^{-1}). The flexible fragment 4 adjusts its conformation to fit the pocket and to enhance the interaction energy (CDK2 and G2K).

The SU (fragment 1–3) interaction energy with the 12 conserved residues (including identical residues) is very similar for all of the studied complexes (−42 kcal mol^{-1} on average). Interestingly, the values are not far from these obtained from the approach of Espinosa and coworkers (Supplementary Table S18). Taking into account all 24 selected residues located at the binding pocket, the energy of interaction with fragment 1–3 of SU is close to −59 kcal mol^{-1} for CDK2, ITK, G2K, KIT and KIT mutant. The VEGFR2–

SU complex stands out with a less negative $E_{\rm es}$. In this case, none of the negatively charged residues are present among the SU nearest neighbours, as is reflected by the $E_{\rm es}$ value, indicating weaker electrostatic interactions (-16 kcal mol^{-1}). Considering all four fragments, the dissimilarity is even more pronounced (49 kcal mol^{-1}). This result disagrees with the inhibitor-constant values ($K_{\rm i}$) found in the literature. The VEGFR2 complex is characterized by the smallest $K_{\rm i}$ among the studied proteins. However, the significant dynamic behaviour of the studied kinases in solution and the different techniques for obtaining the $K_{\rm i}$ are the major disadvantage of this kind of comparison. Moreover, the $K_{\rm i}$ values reported in the literature for the studied complexes are within a small range, *i.e.* from 3.9 n*M* for VEGFR (Martin *et al.*, 2012) to 20 n*M* for ITK (Kutach *et al.*, 2010) and 22 n*M* for KIT (Gajiwala *et al.*, 2009). Regarding the electrostatics, the SU cation should not bind to the catalytic subunit of VEGFR2. A repeated calculation with an unprotonated SU molecule (SU charge equal to zero) suggests that this form is preferred. The most pronounced difference in $E_{\rm es}$ concerns Res1 (Lys838) located in the close proximity of fragment 4. However, a hydrogen bond is not present between these two moieties. The $E_{\rm es}$ with 24 selected residues for the unprotonated SU (fragments 1–3) amounts to -48 kcal mol^{-1} and this value is similar to the results obtained for the rest of the analyses of PKs (Supplementary Table S35). At this point, it should be noted that our calculations do not account for counterions and water molecules.

As described above, SU competes with ATP for binding, and therefore SU binds in and around the region occupied by the adenine ring of ATP (known as the adenine region). However, in contrast to ATP, SU does not require the DFG motif located in the activation loop to adopt the DFG$_{\rm in}$ conformation for binding (type I inhibitor). The studied complexes represent a wide range of possible PK conformations. However, the $E_{\rm es}$ of interaction for fragments 1–3 of the SU molecule is similar regardless of the kinase conformation. This is because the aspartate (Res22) always forms an O$\cdots$F contact (-6 kcal mol^{-1} on average), regardless of its conformation, and phenylalanine (Res23) interacts *via* H$\cdots$H contacts, in which electrostatics is not the main component. The $E_{\rm es}$ for Res22 in the ADP–KIT complex is 55 kcal mol^{-1}, which is compensated by interaction with Res10 (-159 kcal mol^{-1}). In the ADP–KIT complex, $E_{\rm es}$ for the phenylalanine (Res23) is also close to zero. Undoubtedly, ADP only binds to protein with the DFG$_{\rm in}$ conformation, in which Res23 occupies the back pocket and is further away from the binding site. The adenine region is responsible for connecting the ligand to the hinge region by three hydrogen bonds with Res11 and Res13, and the C—H$\cdots$O weak hydrogen bond with Res14 revealed by the topological analysis. For the purpose of comparison only common residues



(*a*) (*b*) (*c*) (*d*) (*e*)

Figure 9
Illustration of the complementarity of the electrostatic potential (e Å^{-1}) mapped onto the van der Waals surface in the binding site of KIT obtained from UBDB for the binding pocket: (*a*) ESP of the inactive conformation of the KIT protein from the SU–KIT complex (PDB entry 3g0e), (*b*) ESP of the SU molecule, (*c*) ESP of the ADP molecule with Mg^{2+} cation, (*d*) ESP of the active conformation of the KIT protein from the ADP–KIT complex (PDB entry 1pkg), (*e*) ESP scale for all figures. Ligands are shown in stick representation and the protein backbone is shown as a cartoon. The maximum and minimum values of the ESP on the vdW surface are -0.17 and -1.05 e Å^{-1}, respectively, for ADP with Mg^{2+}, and 0.00 and 0.44 e Å^{-1}, respectively, for SU.

(Res3–Res13, Res16, Res17 and Res20–Res22) were taken into account owing to the different orientations of ligands and conformations of the KIT proteins (the compared proteins have the same sequence; PDB entries 3g0e and 1pkg). Considering only the overlapping part (the oxindole ring in SU and adenine ring of ADP), the results are comparable. The interaction energy between fragment 1 and KIT amounts to $-26\ \text{kcal mol}^{-1}$, whereas for the adenine ring it is $-22\ \text{kcal mol}^{-1}$ (Supplementary Table S36). Taking into account the residues forming the ADP-binding pocket, the interaction energy amounts to $-202\ \text{kcal mol}^{-1}$, which is of the same magnitude as the values obtained for PK–SU complexes.

3.5. Electrostatic potential analysis

Electrostatic potential (ESP) analysis allows us to understand the nature of noncovalent interactions, which may significantly contribute to crystal formation and ligand binding. Therefore, obtaining a reliable ESP is crucial and can be useful in the drug-design process. However, accurate theoretical calculations are still inaccessible for systems as large as proteins, and thus the pseudoatom databank approach can be very attractive, as it enables a reliable reconstruction of charge-density distribution and ESP for proteins and protein–ligand complexes in a reasonable time.

Firstly, the ESP was obtained for an isolated molecule extracted from the crystal lattice. The ESPs derived for SU in the SUM crystal, SU obtained from the refined multipole model and SU reconstructed with the aid of the UBDB (Jarzembska & Dominiak, 2012) are very similar. The ESP values mapped onto the vdW surface change from positive values around the quaternary N atom to slightly negative values at the F atom. The significant difference between the ESP of the SU molecules is the noticeable polarization of the H atoms for the SUM crystal owing to stronger interactions, shorter interatomic contacts *etc.* (Suplementary Fig. S5).

The opposite charge of SU and ATP molecules dramatically differentiates their ESPs, although both molecules bind to the same binding pocket. To explore this phenomenon, an analysis of the ESP was carried out for the VEGFR2 (Supplementary Fig. S6) and KIT proteins with SU and ADP as ligands (Fig. 9).

It appears that the inactive state of VEGFR2 is characterized by a positive ESP (Supplementary Fig. S6) of the binding pocket and this corresponds well with the positive $E_{\rm es}$ for interaction between SU and this protein ($+49\ \text{kcal mol}^{-1}$). The ESP mapped onto the vdW surface obtained for an unprotonated SU molecule fits better to the positive binding pocket of VEGFR2 (Supplementary Fig. S6*c*).

SU molecules prefer the inactivated state of the KIT kinase (DiNitto *et al.*, 2010) with the DFG_{out} conformation and the activation loop closing the pocket, which prevents ATP binding. The SU molecule, with a significant positive ESP, fits well to the smaller pocket, especially to the negative area in the hinge region, and complements it (Figs. 9*a* and 9*b*). The opposite situation is observed in the KIT–ADP complex (with kinase in the active form), where the negative electrostatic potential of the ATP molecule conforms to the positive ESP of the binding pocket of the active KIT kinase form. The conformational change between the active and inactive forms of KIT kinases leads to the outstanding change in the ESP of the two forms.

4. Conclusions

Electrostatic forces are one of the most important factors that contribute to the formation of crystal structures and protein–ligand complexes. In this contribution, for the first time the aspherical atom databank approach, Hirshfeld surface analysis and quantum theory of atoms in molecules were used as a foundation for the analysis of interactions between an inhibitor and a protein. Comparative studies regarding the electrostatic interaction energy calculation and electrostatic potential analysis can explain the binding preferences.

Here, we have presented a wide-context study of the protein kinase inhibitor sunitinib. High-resolution X-ray data provided insight into the electronic nature of the sunitinib malate crystal structure and thus enabled quantitative analysis of the intermolecular interaction strength. The experimental charge distribution of sunitinib malate reported here led to energy results that were consistent with theoretical calculations. The strongest hydrogen bonds in the crystal lattice can be explained by the charge-assisted hydrogen bond and the resonance-assisted hydrogen bond. Finally, the experimental charge density of sunitinib was compared with the analogous representation derived on the basis of the aspherical atom databank, which once again supported its applicability to the reconstruction of charge distributions. A set of selected kinase–protein complexes were explored using the UBDB-derived charge density, as this allows analysis of the electrostatic properties of VEGRF2, CDK2, G2K, KIT and IT kinase complexes. The Hirshfeld surface analysis revealed similar patterns for interactions occurring between sunitinib molecules in the sunitinib malate crystal and in the protein binding pockets. However, as expected, the packing of sunitinib in the sunitinib malate crystal is more condensed. The analysis of interactions by means of interatomic distances and van der Waals radii is insufficient. QTAIM analysis of the charge density is able to cast more light on interactions in ligand–protein complexes. This approach was helpful in detecting the most conserved interactions between the ligand and the residues forming the binding pocket and revealed a characteristic pattern of interactions that includes more interactions than are pinpointed in the literature: three hydrogen bonds, two weak $\text{C—H}\cdots\text{O}$ interactions and four $\text{C—H}\cdots\pi$ contacts with the corresponding residues. These residues are the gatekeeper residues and the other four residues of the hinge region, and also either leucine or isoleucine from the Gly-rich loop. As indicated by the simple empirical formula proposed by Espinosa *et al.* (1998), the head of the sunitinib molecule interacts with similar strength among the studied structures, but in the case of the CDK2 complex this interaction is significantly stronger.

A more general approach which also takes into account long-range interactions is the combination of the aspherical atom databank concept and the exact potential multipole method, which enables the calculation of pure electrostatic interaction energies. Considering all 24 selected residues forming the binding pocket, the energetic results fall into a small range of interaction energies for the sunitinib head, where only the VEGFR2 kinase is characterized by a significantly weaker interaction. The results of the additional exact potential multipole method calculation suggest that sunitinib interacts with VEGFR2 in its unprotonated form. Also taking into account the tail of sunitinib, the energetic results are more diverse; however, they are still comparable. This can be explained by the conformational flexibility of sunitinib, especially in the side-chain region. A sunitinib molecule can adjust its conformation to enhance its electrostatic interaction with kinases. For instance, in the CDK2 structure it adopts a less favourable conformation so as to saturate the hydrogen-bond donors and acceptors of Asp86 and Leu134.

In contrast to the ATP molecule, sunitinib is small and fits into the binding pocket irrespective of the activation state of the protein kinase (a small pocket for the inactive kinases VEGRF2 and KIT and a larger pocket for the active kinases CDK2, G2K and IT). The interaction energy with the highly conserved DFG triad responsible for closing the binding pocket is similar for all complexes studied. The KIT kinase protein structure in the inactive form (with sunitinib) was compared with its active form hosting an ADP molecule. Both ligands interact similarly with the residues of the hinge region, forming hydrogen bonds of comparable interaction energy. Moreover, their entire electrostatic interaction energies with the corresponding KIT binding pockets are in the same range. In the KIT kinases, the reversible conformational change $\mathrm{DFG_{out}}$ to $\mathrm{DFG_{in}}$ involves modification of the electrostatic potential. More negative and more positive values of the electrostatic potential mapped onto van der Waals surfaces are exhibited by the inactive and active forms, respectively. These features are in agreement with the electrostatic potentials of sunitinib and the ATP molecule. These results explain the preference for the inactive form of KIT over the active form by the sunitinib molecule.

5. Related literature

The following references are cited in the Supporting Information for this article: Abramov (1997), Boys & Bernardi (1970), Chai & Head-Gordon (2008), Espinosa *et al.* (1998), Frisch *et al.* (1984, 2009), Hansen & Coppens (1978), McLean & Chandler (1980), O'Hagan & Rzepa (1997), Simon *et al.* (1996), Spackman & Jayatilaka (2009) and Volkov, King *et al.* (2006).

MM thanks the National Science Centre for financial support within grants UMO-2011/01/N/ST4/03616 and UMO-2013/08/T/ST4/00494. KNJ would like to thank the Foundation for Polish Science for financial support within the START program. PMD acknowledges the Polish Ministry of Science and Higher Education for financial support through Grant Iuventus Plus No. IP2010 0076/70. AMG was supported by the EU through the European Social Fund, contract No. UDA-POKL.04.01.01-00-072/09-00. The authors gratefully acknowledge the Interdisciplinary Centre for Mathematical and Computational Modelling in Warsaw (grant No. G33-14) for providing the computer facilities where most of the *Gaussian* 09 calculations were conducted.

References

Abramov, Yu. A. (1997). *Acta Cryst.* A**53**, 264–272.

Abrams, T. J., Lee, L. B., Murray, L. J., Pryer, N. K. & Cherrington, J. M. (2003). *Mol. Cancer Ther.* **2**, 471–478.

Allen, F. H. & Bruno, I. J. (2010). *Acta Cryst.* B**66**, 380–386.

Bader, R. F. (1994). *Atoms in Molecules: A Quantum Theory.* Oxford University Press.

Bąk, J. M., Domagała, S., Hübschle, C., Jelsch, C., Dittrich, B. & Dominiak, P. M. (2011). *Acta Cryst.* A**67**, 141–153.

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. (2000). *Nucleic Acids Res.* **28**, 235–242.

Bertolasi, V., Gilli, P., Ferretti, V. & Gilli, G. (1991). *J. Am. Chem. Soc.* **113**, 4917–4925.

Blessing, R. H. (1987). *Crystallogr. Rev.* **1**, 3–58.

Boys, S. & Bernardi, F. (1970). *Mol. Phys.* **19**, 553–566.

Brock, C. P., Dunitz, J. D. & Hirshfeld, F. L. (1991). *Acta Cryst.* B**47**, 789–797.

Chai, J. D. & Head-Gordon, M. (2008). *Phys. Chem. Chem. Phys.* **10**, 6615–6620.

Cowan-Jacob, S. W. (2006). *Cell. Mol. Life Sci.* **63**, 2608–2625.

Demetri, G. D. *et al.* (2006). *Lancet*, **368**, 1329–1338.

DiNitto, J. P., Deshmukh, G. D., Zhang, Y., Jacques, S. L., Coli, R., Worrall, J. W., Diehl, W., English, J. M. & Wu, J. C. (2010). *J. Biochem.* **147**, 601–609.

Dittrich, B., Hübschle, C. B., Pröpper, K., Dietrich, F., Stolper, T. & Holstein, J. J. (2013). *Acta Cryst.* B**69**, 91–104.

Domagała, S., Fournier, B., Liebschner, D., Guillot, B. & Jelsch, C. (2012). *Acta Cryst.* A**68**, 337–351.

Espinosa, E., Molins, E. & Lecomte, C. (1998). *Chem. Phys. Lett.* **285**, 170–173.

Faivre, S., Demetri, G., Sargent, W. & Raymond, E. (2007). *Nature Rev. Drug Discov.* **6**, 734–745.

Farrugia, L. J. (1997). *J. Appl. Cryst.* **30**, 565.

Flack, H. D. (1983). *Acta Cryst.* A**39**, 876–881.

Frisch, M. J., Pople, J. A. & Binkley, J. S. (1984). *J. Chem. Phys.* **80**, 3265.

Frisch, M. J. *et al.* (2009). *Gaussian* 09, Revision A.1. Gaussian Inc., Wallingford, Connecticut, USA.

Gajiwala, K. S. *et al.* (2009). *Proc. Natl Acad. Sci. USA*, **106**, 1542–1547.

Gilli, G., Bellucci, F., Ferretti, V. & Bertolasi, V. (1989). *J. Am. Chem. Soc.* **111**, 1023–1028.

Gilli, P., Bertolasi, V., Pretto, L., Ferretti, V. & Gilli, G. (2004). *J. Am. Chem. Soc.* **126**, 3845–3855.

Goodman, V. L. *et al.* (2007). *Clin. Cancer Res.* **13**, 1367–1373.

Góra, R. W., Maj, M. & Grabowski, S. J. (2013). *Phys. Chem. Chem. Phys.* **15**, 2514–2522.

Grimme, S. (2006). *J. Comput. Chem.* **27**, 1787–1799.

Hanks, S. K. & Hunter, T. (1995). *FASEB J.* **9**, 576–596.

Hansen, N. K. & Coppens, P. (1978). *Acta Cryst.* A**34**, 909–921.

Hirshfeld, F. L. (1976). *Acta Cryst.* A**32**, 239–244.

Huse, M. & Kuriyan, J. (2002). *Cell*, **109**, 275–282.

Jarzembska, K. N. & Dominiak, P. M. (2012). *Acta Cryst.* A**68**, 139–147.

Kufareva, I. & Abagyan, R. (2008). *J. Med. Chem.* **51**, 7921–7932.

Kutach, A. K., Villaseñor, A. G., Lam, D., Belunis, C., Janson, C., Lok, S., Hong, L.-N., Liu, C.-M., Deval, J., Novak, T. J., Barnett, J. W., Chu, W., Shaw, D. & Kuglstatter, A. (2010). *Chem. Biol. Drug Des.* **76**, 154–163.

Lecomte, C., Guillot, B., Muzet, N., Pichon-Pesme, V. & Jelsch, C. (2004). *Cell. Mol. Life Sci.* **61**, 774–782.

Liu, Y. & Gray, N. S. (2006). *Nature Chem. Biol.* **2**, 358–364.

Madsen, A. Ø. (2006). *J. Appl. Cryst.* **39**, 757–758.

Martin, M. P., Alam, R., Betzi, S., Ingles, D. J., Zhu, J. Y. & Schönbrunn, E. (2012). *Chembiochem*, **13**, 2128–2136.

McLean, A. D. & Chandler, G. S. (1980). *J. Chem. Phys.* **72**, 5639.

McTigue, M., Murray, B. W., Chen, J. H., Deng, Y.-L., Solowiej, J. & Kania, R. S. (2012). *Proc. Natl Acad. Sci. USA*, **109**, 18281–18289.

Mebs, S., Lüth, A. & Luger, P. (2010). *Bioorg. Med. Chem.* **18**, 5965–5974.

Mendel, D. B. *et al.* (2003). *Clin. Cancer Res.* **9**, 327–337.

Mol, C. D., Lim, K. B., Sridhar, V., Zou, H., Chien, E. Y. T., Sang, B.-C., Nowakowski, J., Kassel, D. B., Cronin, C. N. & McRee, D. E. (2003). *J. Biol. Chem.* **278**, 31461–31464.

Munshi, P., Madsen, A. Ø., Spackman, M. A., Larsen, S. & Destro, R. (2008). *Acta Cryst.* A**64**, 465–475.

Murray, L. J., Abrams, T. J., Long, K. R., Ngai, T. J., Olson, L. M., Hong, W., Keast, P. K., Brassard, J. A., O'Farrell, A. M., Cherrington, J. M. & Pryer, N. K. (2003). *Clin. Exp. Metastasis*, **20**, 757–766.

Noble, M. E. M., Endicott, J. A. & Johnson, L. N. (2004). *Science*, **303**, 1800–1805.

O'Farrell, A.-M., Abrams, T. J., Yuen, H. A., Ngai, T. J., Louie, S. G., Yee, K. W. H., Wong, L. M., Hong, W., Lee, L. B., Town, A., Smolich, B. D., Manning, W. C., Murray, L. J., Heinrich, M. C. & Cherrington, J. M. (2003). *Blood*, **101**, 3597–3605.

O'Hagan, D. & Rzepa, H. (1997). *Chem. Commun.*, pp. 645–652.

Osusky, K. L., Hallahan, D. E., Fu, A., Ye, F., Shyr, Y. & Geng, L. (2004). *Angiogenesis*, **7**, 225–233.

Rock, E. P., Goodman, V., Jiang, J. X., Mahjoob, K., Verbois, S. L., Morse, D., Dagher, R., Justice, R. & Pazdur, R. (2007). *Oncologist*, **12**, 107–113.

Schmidt, A., Teeter, M., Weckert, E. & Lamzin, V. S. (2011). *Acta Cryst.* F**67**, 424–428.

Schueneman, A. J., Himmelfarb, E., Geng, L., Tan, J., Donnelly, E., Mendel, D., McMahon, G. & Hallahan, D. E. (2003). *Cancer Res.* **63**, 4009–4016.

Sheldrick, G. M. (2008). *Acta Cryst.* A**64**, 112–122.

Sidoryk, K., Malińska, M., Bańkowski, K., Kubiszewski, M., Łaszcz, M., Bodziachowska-Panfil, M., Kossykowska, M., Giller, T., Kutner, A. & Woźniak, K. (2013). *J. Pharm. Sci.* **102**, 706–716.

Simon, S., Duran, M. & Dannenberg, J. J. (1996). *J. Chem. Phys.* **105**, 11024.

Spackman, M. A. & Jayatilaka, D. (2009). *CrystEngComm*, **11**, 19.

Sun, L. *et al.* (2003). *J. Med. Chem.* **46**, 1116–1119.

Toledo, L. M., Lydon, N. B. & Elbaum, D. (1999). *Curr. Med. Chem.* **6**, 775–805.

Volkov, A., Abramov, Y. A. & Coppens, P. (2001). *Acta Cryst.* A**57**, 272–282.

Volkov, A., King, H. F. & Coppens, P. (2006). *J. Chem. Theory Comput.* **2**, 81–89.

Volkov, A., Koritsanszky, T. & Coppens, P. (2004). *Chem. Phys. Lett.* **391**, 170–175.

Volkov, A., Li, X., Koritsanszky, T. & Coppens, P. (2004). *J. Chem. Phys. A*, **108**, 4283–4300.

Volkov, A., Macchi, P., Farrugia, L. J., Gatti, C., Mallinson, P., Richter, T. & Koritsanszky, T. (2006). *XD2006 – A Computer Program Package for Multipole Refinement, Topological Analysis of Charge Densities and Evaluation of Intermolecular Energies from Experimental and Theoretical Structure Factors.* http://xd.chem.buffalo.edu/.

Wood, G. P. F., Radom, L., Petersson, G. A., Barnes, E. C., Frisch, M. J. & Montgomery, J. A. (2006). *J. Chem. Phys.* **125**, 094106.

Word, J. M., Lovell, S. C., Richardson, J. S. & Richardson, D. C. (1999). *J. Mol. Biol.* **285**, 1735–1747.

Zuccotto, F., Ardini, E., Casale, E. & Angiolini, M. (2010). *J. Med. Chem.* **53**, 2681–2694.