



A rush to explore protein–ligand electrostatic interaction energy with *Charger*

Vedran Vuković, Theo Leduc, Zoe Jelić-Matošević, Claude Didierjean, Frédérique Favier, Benoît Guillot and Christian Jelsch

Acta Cryst. (2021). D77, 1292–1304



IUCr Journals

CRYSTALLOGRAPHY JOURNALS ONLINE

Author(s) of this article may load this reprint on their own web site or institutional repository provided that this cover page is retained. Republication of this article or its storage in electronic databases other than as specified above is not permitted without prior permission in writing from the IUCr.

For further information see <https://journals.iucr.org/services/authorrights.html>

A rush to explore protein–ligand electrostatic interaction energy with *Charger*

Vedran Vuković,^a Theo Leduc,^a Zoe Jelić-Matošević,^b Claude Didierjean,^a Frédérique Favier,^a Benoît Guillot^a and Christian Jelsch^{a*}

^aUniversité de Lorraine, CNRS, CRM2, F-54000 Nancy, France, and ^bDepartment of Chemistry, Faculty of Science, University of Zagreb, 10000 Zagreb, Croatia. *Correspondence e-mail: christian.jelsch@univ-lorraine.fr

Received 17 June 2021

Accepted 13 August 2021

Edited by M. Rudolph, F. Hoffmann-La Roche Ltd, Switzerland

Keywords: electrostatics; polarization; interaction energy; Hansen–Coppens model; protein–ligand interactions; *Charger*.

Supporting information: this article has supporting information at journals.iucr.org/d

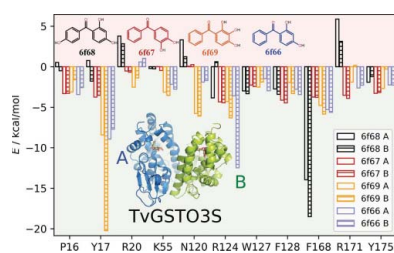
The mutual penetration of electron densities between two interacting molecules complicates the computation of an accurate electrostatic interaction energy based on a pseudo-atom representation of electron densities. The numerical exact potential and multipole moment (nEP/MM) method is time-consuming since it performs a 3D integration to obtain the electrostatic energy at short interaction distances. Nguyen *et al.* [(2018), *Acta Cryst. A* **74**, 524–536] recently reported a fully analytical computation of the electrostatic interaction energy (aEP/MM). This method performs much faster than nEP/MM (up to two orders of magnitude) and remains highly accurate. A new program library, *Charger*, contains an implementation of the aEP/MM method. *Charger* has been incorporated into the *MoProViewer* software. Benchmark tests on a series of small molecules containing only C, H, N and O atoms show the efficiency of *Charger* in terms of execution time and accuracy. *Charger* is also powerful in a study of electrostatic symbiosis between a protein and a ligand. It determines reliable protein–ligand interaction energies even when both contain S atoms. It easily estimates the individual contribution of every residue to the total protein–ligand electrostatic binding energy. Glutathione transferase (GST) in complex with a benzophenone ligand was studied due to the availability of both structural and thermodynamic data. The resulting analysis highlights not only the residues that stabilize the ligand but also those that hinder ligand binding from an electrostatic point of view. This offers new perspectives in the search for mutations to improve the interaction between the two partners. A proposed mutation would improve ligand binding to GST by removing an electrostatic obstacle, rather than by the traditional increase in the number of favourable contacts.

1. Introduction

An investigation of intermolecular interactions involves an intelligent examination of electrostatic interaction energy. One major obstacle in such energy calculations is how to obtain an accurate representation of the total charge distributions of interacting molecules. They are normally associated with theoretical calculation methods such as density-functional theory (DFT).

Full DFT calculations on large systems take a prohibitive amount of time. Approximations such as the kernel energy method give comparable results much faster, as they scale better with the number of atoms concerned (Huang *et al.*, 2005; Mandal *et al.*, 2017). This method has successfully been applied to many different biological systems (see Massa *et al.*, 2019, and references therein).

Charge-density determinations from ultra-high resolution X-ray crystallography provide an experimental counterpart to theoretical studies. The Hansen and Coppens multipolar pseudo-atom model properly represents an electron-density



distribution based on an X-ray diffraction experiment (Hansen & Coppens, 1978). A good model necessitates high-quality crystals, which are nearly impossible to produce in some fields. Biomacromolecular crystallography in particular has this problem. A compromise solution comes from the fact that the multipolar parameters of chemically similar atoms are similar. One can reconstruct protein electron density with ease from building blocks (*i.e.* transferable parameters) from small molecules.

A parameter transfer from a database to a protein structure takes less than a minute on a modern computer. Transferable multipolar parameter databases draw from experimental (for example ELMAM2; Domagała *et al.*, 2012) electron densities, augmented with polarizabilities (Leduc, 2019; Leduc *et al.*, 2019), or from theoretical electron densities [for example INVARIOM (Dittrich *et al.*, 2004) and UBDB (Dominiak *et al.*, 2007)].

Missing atom types pose an obstacle to database transfers. Similar atom types can replace them, or (particularly for theoretical databanks) one can obtain the charge density by an additional DFT calculation. Kumar *et al.* (2019) addressed this problem to some extent by extending UBDB with atom types commonly found in drug molecules.

Calculating the electrostatic energy (E_{elec}) remains a challenge even with a fitting electron-density model to hand, largely because of the overlap between electron clouds (Volkov *et al.*, 2004). The challenge of computation coincides with the great interest of chemists, since atoms with overlapping electron densities are close enough to form chemically meaningful strong contacts.

One should first divert one's attention to a simpler case: atoms that are further away and have a negligible density overlap. They can nevertheless contribute significantly to interaction energies, and more so when they carry considerable charge. A point-charge approximation can describe this kind of interaction well, but the Buckingham approximation (Buckingham, 1959) performs better as it includes multipole moments (MM) along with point charges. These approximations also alleviate the time-cost concerns associated with large biomolecular systems.

Going back to the more relevant case of atoms that are close together, Volkov *et al.* (2004) developed the exact potential (EP) method for the multipolar electron-density model. EP involves a complete numerical 3D integration of the product between the density of one atom and the potential of another atom, and thus handles overlap density directly, albeit slowly. A union of the slow EP method at short distance and the fast MM approximation at long distance gives rise to the precise and reasonably fast EP/MM method (Volkov *et al.*, 2004). Fournier *et al.* (2009) implemented this method in the *VMoPro* program in *MoProSuite*.

To improve the calculation speed further, Nguyen *et al.* (2018) proposed an analytical method to evaluate electrostatic interaction energies between charge densities for the Hansen–Coppens model. This method is largely based on previous mathematical developments by Jones and Weatherford (see Jones, 1993 and references therein) and Löwdin (1956), and

substitutes the numerical EP integrals with analytical ones. We follow their lead in renaming the original method using numerical integration nEP/MM and calling the newer method using analytical integration aEP/MM.

Nguyen *et al.* (2018) promise high accuracy, with a calculation speed improved by two orders of magnitude. Nguyen and Volkov further analysed the precision of aEP/MM and proposed a more precise Fourier-based method in a newer paper (Nguyen & Volkov, 2019), albeit at a certain time cost. They have also recently proposed, in collaboration with P. Macchi, a molecular multipolar moments method that is easily applicable to crystals, as well as corrections for errors in Ewald summations due to density overlap (Nguyen *et al.*, 2020), following in the footsteps of Bojarowski *et al.* (2016). They tested all of the methods on small molecules and oligopeptide benchmark systems.

We have adapted the aEP/MM method to create a new code library called *Charger*. It is an independent implementation of the same method with several enhancements geared towards calculation speed. Our molecular-visualization and electron-density exploration software *MoProViewer* (Guillot *et al.*, 2014) accommodates a simple user interface to *Charger*. It includes a special module for biomacromolecular structure exploration.

Various benchmark systems containing homodimers and heterodimers derived from crystal structures served as the first tests for *Charger*. The molecules contained within these systems ranged from water and methanol to Leu-enkephalin and a dodecapeptide, all containing only C, H, N and O atoms.

Dominiak *et al.* (2009) have already extensively demonstrated that a model from transferred parameters is indeed useful when investigating protein–ligand interactions. Specifically, they used the nEP/MM method to calculate electrostatic interaction energies between a neuraminidase protein and a wide range of its inhibitors.

With this in mind, we wanted to give *Charger* a greater challenge: glutathione transferase (GST), a 2×240 -residue dimeric protein. Schwartz *et al.* (2018) determined four structures of this enzyme, each containing a different benzophenone ligand in the active site. They also measured thermal shifts in the presence of these same ligands, which one can use to determine protein–ligand binding constants and therefore relate to interaction energies (Cimpmperman *et al.*, 2008). These ligands provide a textbook example to gauge our method: chemically similar, they form hydrogen bonds and similar electrostatic interactions with the host protein. The protein–ligand complexes exhibit distinct thermal shifts for each ligand.

Trying to improve protein–ligand binding is another direction in which the quality of *Charger* calculations can be determined. A typical approach seeks to increase pertinent contacts between the protein and ligand (Thapa & Raghavachari, 2019). The biomacromolecular module connected to *Charger* can point to residues that hinder ligand binding by poisoning the electrostatic environment of the active site. We pursued an *in silico* inquiry into this topic using the GST–benzophenone system.

2. Methods

2.1. Charger

Calculations based on quantum chemistry demand the use of computational algorithms. Quantum chemistry often models the world as electron functions that make up orbitals that compose atoms (along with nuclei) that constitute molecules. Computer programs obediently undertake the inhuman drudge of going over the same calculation repeatedly. Our invention, *Charger*, partakes in this tradition.

Charger is a code library written in C that computes the electrostatic interaction energy between molecules. *Charger* works with the Hansen & Coppens (1978) charge-density model,

$$\rho_{\text{atom}}(r) = Z\delta(r) + N_{\text{core}}\rho_{\text{core}}(r) + P_{\text{val}}\kappa^3\rho_{\text{valence}}(\kappa, r) + \sum_{l=0}^{l_{\text{max}}} R_l(\kappa', r)\kappa'^3 \sum_{m=-l}^l P_{lm}d_{lm}(\theta, \varphi). \quad (1)$$

This model contains a point-charge nucleus (the first term), two spherical terms and a multipolar term for the electron density of each atom (defined in more detail in the supporting information). *Charger* has means of calculating (i) the electrostatic energy ($E = q_A q_B / r$) of a pair of point charges, (ii) the exact electrostatic potential (EP, Φ) from a spherical or multipolar density according to Jones (1993), (iii) the energy of a point charge in a potential field [$E = q_A \Phi_B(r_A)$] and (iv) the energy of two interacting electron-density (ρ) distributions,

$$E = \int \int \frac{\rho_A(\mathbf{r}_A)\rho_B(\mathbf{r}_B)}{|\mathbf{r}_A - \mathbf{r}_B|} d\mathbf{r}_A d\mathbf{r}_B. \quad (2)$$

(All equations use atomic units.) *Charger* employs Löwdin's α -function (Löwdin, 1956) to compute this last term when the electron density is expressed as a sum of Slater functions, according to the method proposed by Nguyen *et al.* (2018). They dubbed their method 'analytical exact potential' (aEP), as opposed to the 'numerical exact potential' (nEP) method proposed by Volkov, Koritsanszky and Coppens in 2004 (Volkov *et al.*, 2004).

At short distances, one can calculate the energy using *Charger* as the sum of the following terms:

$$E_{\text{Total}} = E_{\text{nn}} + E_{\text{ns}} + E_{\text{nm}} + E_{\text{ss}} + E_{\text{sm}} + E_{\text{mm}}. \quad (3)$$

The first term represents the nucleus–nucleus energy E_{nn} (for point charges Z of two nuclei, A and B, separated by a distance d),

$$E_{\text{nn}} = Z_A Z_B / d. \quad (4)$$

Two nucleus–electron energy terms E_{ns} and E_{nm} follow [for a point charge Z_A in a spherical $\Phi_B(d)$ or multipolar electric potential $\Phi_B(d)d_{lm}(\theta_A, \varphi_A)$],

$$E_{\text{ns}} = Z_A \Phi_B(d), \quad (5)$$

$$E_{\text{nm}} = Z_A \Phi_B(d)d_{lm}(\theta_A, \varphi_A). \quad (6)$$

The remaining three electron–electron energy terms are E_{ss} , E_{sm} and E_{mm} (two interacting electron-density distributions,

either spherical or multipolar, as defined in the supporting information).

A faster calculation at long distances comes from the substitution of the nuclear and spherical electron terms in the Hansen–Coppens model with a point charge, in the spirit of the Buckingham approximation (Buckingham, 1959). For a pair of interacting atoms at a long distance, we calculate the following sum,

$$E_{\text{Pseudo-Buckingham}} = E_{\text{cc}} + E_{\text{cm}} + E_{\text{mm}}. \quad (7)$$

The first term is a simple point-charge energy similar to E_{nn} ,

$$E_{\text{cc}} = q_A q_B / d, \quad (8)$$

with the charges q of the respective atoms; these charges are defined in the supporting information. It substitutes the $E_{\text{nn}} + E_{\text{ns}} + E_{\text{ss}}$ terms in (3). The point-multipolar term E_{cm} replaces the $E_{\text{nm}} + E_{\text{sm}}$ terms. It is calculated similarly to the E_{nm} term, after substituting z with q ,

$$E_{\text{cm}} = q_A \Phi_B(d)d_{lm}(\theta_A, \varphi_A). \quad (9)$$

The last term is the same multipolar–multipolar term as in (3) (E_{mm}).

The terms E_{cc} and E_{cm} are similar in nature to the terms E_{nn} and E_{nm} . This allows the repurposing of the existing code for a new goal. We call this the pseudo-Buckingham approximation, since it does not involve an evaluation of tensor products, as Buckingham proposed. Nevertheless, the E_{mm} term involves a limited number of computations because the spherical harmonic functions underlying multipoles are orthonormal.

It is common to refer to the Buckingham approximation as multipolar moments (MM) in the literature. We therefore name our combined method analytical exact potential/pseudo-multipolar moments (aEP/pMM). The program library *Charger* provides the functions to calculate energy using both aEP and pMM: library users have the freedom to choose the threshold distances to use the former or the latter.

One important goal of the *Charger* library is calculation speed. A sizeable calculation speedup emerges by excluding very long distance multipolar terms. *Charger* can use a user-adjustable cutoff scheme based on distances and multipole levels (l).

We established two different cutoff schemes (for 'precise' and 'rapid' calculations) based on test calculations. These test calculations revolved around typically high multipolar populations: $P_{lm} = 0.3$ e. The chosen cutoff distances rest on individual multipole–multipole interaction energies for all possible pairs of large multipoles from C and H atoms.

We chose the 'precise' cutoffs as those (rounded-up) distances for which the calculated interaction energy is inferior to 5 cal mol^{-1} (Table 1). This means that in an extreme case of very high multipolar populations, the error due to each excluded multipolar term should be lower than 5 cal mol^{-1} . The error is generally much lower than this in practice.

The 'rapid' scheme cuts deeper: no multipole–multipole interactions are calculated beyond 15 \AA . The two cutoff

Table 1

Cutoffs used in *Charger*.

Charger skips over certain pairs of multipoles at long distance, making the calculation more robust and faster. We propose two cutoff schemes: one that gives precise results (*MoProViewer* default) and another that gives rapid results. Some criteria are common to both cutoff schemes. The precise cutoffs represent (rounded-up) distances for which the calculated interaction energy between a pair of multipoles (stated in the left column, both populations $P_{lm} = 0.3$) is inferior to 5 cal mol⁻¹. The rapid scheme cuts all multipole–multipole interactions calculated beyond 15 Å, a distance chosen as reasonable and proven to be so in subsequent calculations.

Multipoles	Cutoff for precise calculations (Å)	Cutoff for rapid calculations (Å)
Dipole × dipole	40	
Dipole × quadrupole	20	15
Dipole × octupole, quadrupole × quadrupole	18	
Dipole × hexadecapole, quadrupole × octupole		14
Quadrupole × hexadecapole, octupole × octupole, octupole × hexadecapole		12
Hexadecapole × hexadecapole		10

schemes share some cutoff criteria concerning high-level multipoles (Table 1).

We further develop the theoretical background behind these calculations in the supporting information. We have made the aEP/pMM method available in the form of code through the *Charger* library and as a ready-made tool to explore electrostatic interaction energy within *MoProViewer*, our flagship molecular-visualization program (Guillot *et al.*, 2014).

2.2. Applying *Charger* to peptides and proteins

The ultimate proof of a method is in how broad its applications are. One could study many chemical systems with *Charger*, but our interest lies primarily in the biochemical and biostructural domain. We started with several small benchmark systems, and proceeded to an exploration of interactions between a protein and several ligands forming complexes of known structure.

We first benchmarked *Charger* on dimeric systems derived from crystal structures following the lead of Nguyen *et al.* (2018). These systems contained water, glycine, 2-butanone, *N,N*-dimethylacetamide, Leu-enkephalin, nonapeptide, decapeptide and dodecapeptide molecules. These benchmarks give some points of comparison between *Charger* and the code integrated in the *XDPROP* program. Leu-enkephalin is an opioid neurotransmitter pentapeptide. The nona-, deca- and dodecapeptides consist of leucine and 2-methylalanine residues capped with *t*-butoxycarbonyl (N-cap) and methyl (O-cap) groups. Leu-enkephalin and these peptides contain the C, H, O and N atom types often found in biomacromolecules.

The ELMAM2 database does not contain four atom types for 6–8 atoms in the nona-, deca- and dodecapeptides, depending on the peptide composition; their electron density was approximated as spherical (and neutral before application of the electroneutrality constraint). In addition, for comparison purposes, we used the AMBER point-charge model on several benchmark systems for which database transfer was possible. *Charger* calculated electrostatic interaction energies based on these AMBER models as well.

We tested *Charger* further on a larger system: a protein–ligand complex with published experimental structural and thermodynamic data. The glutathione transferase Omega 3S from *Trametes versicolor* (*TvGSTO3S*) is an intriguing model because crystal structures of this 240-residue enzyme have recently been obtained in complex with various hydroxybenzophenone inhibitors (HBPs). Schwartz *et al.* (2018) used thermal shift assays to evaluate the thermal stabilization of *TvGSTO3S* upon binding HBPs.

The studied hydroxybenzophenones have various (in number and position) hydroxyl groups on both phenyl rings (Table 2). Of the eight HBPs assessed, Schwartz and coworkers obtained crystal structures of four *TvGSTO3S*–HBP complexes, which we analysed with *Charger* (PDB entries 6f66, 6f67, 6f68 and 6f69).

These protein structures originate from the Protein Data Bank. The PDB files required preparation: erasure of the C-terminal His tag (which is not present in the native form of the protein) and the addition of three N-terminal missing residues (methionine–serine–serine) using the *Builder* tool in *PyMOL* (Schrödinger). We used the *WHATIF* web server (Hoofst *et al.*, 1996) to ascertain the H-atom positions that form the most optimal hydrogen-bonding network in the protein structure. This procedure is agnostic to any ligands, so we used *MoPro* to add missing H atoms in ligand molecules (Jelsch *et al.*, 2005).

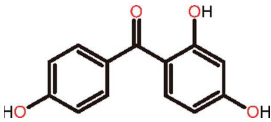
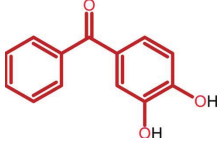
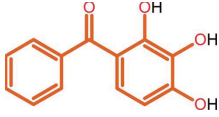
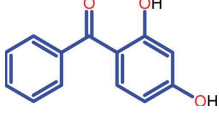
A manual search for prospective protein–ligand hydrogen bonds followed, involving a visual inspection of protein residues with at least one atom within 5 Å of any atom of the ligand. The H-atom positions of these residues and ligands were adjusted manually where deemed necessary in order to maximize the number and the strength of hydrogen bonds. We converted the structure into a *MoPro*-compatible format using *Import2MoPro* (Jelsch *et al.*, 2005).

We developed a multipolar database-transfer tool inside *MoProViewer* (Guillot *et al.*, 2014). This tool has an extension for automatic parameter transfer onto protein structures with appropriate charge assignment (+1 for arginine and lysine, –1 for aspartate and glutamate, 0 for others; there were no doubly protonated histidine residues; all atoms of a charged residue

Table 2

Thermodynamic results obtained for TvGSTO3S with hydroxybenzophenone inhibitors by thermal shift assays (Schwartz *et al.*, 2018).

The PDB codes for the crystal structures of the TvGSTO3S–HBP complexes are indicated along with the experimental thermal shifts (ΔT_d). The colouring of the carbon skeletons of the ligands reflects the colouring in the graphs and structure models throughout this paper.

Compound	Structure	ΔT_d (K)	PDB code
2,4,4'-Trihydroxybenzophenone (2,4,4'-HBP)		2.9	6f68
3,4-Dihydroxybenzophenone (3,4-HBP)		4.4	6f67
2,3,4-Trihydroxybenzophenone (2,3,4-HBP)		5.0	6f69
2,4-Dihydroxybenzophenone (2,4-HBP)		5.7	6f66

share the added charge equally). This new tool transferred experimental multipolar electron-density parameters from the ELMAM2 database (Domagała *et al.*, 2012) to the treated protein structures.

Not so long ago, one could not go beyond the described database transfer. In a recent article, we demonstrated that the transferred electron density complemented with theoretical polarizabilities gives access to polarization energies (Leduc *et al.*, 2019). We built a custom database of theoretical average atomic polarizabilities for ELMAM atoms found in HBPs and proteins. Our computation of the electrostatic interaction and polarization energies of the TvGSTO3S–HBP interactions followed the method described by T. Leduc in his PhD thesis (Leduc, 2019).

After the transfer of atom polarizabilities from this database, the *MoProViewer* polarization module served to compute dipoles induced on the ligand by residues near to the ligand and dipoles induced by the ligand on the same residues (until the self-consistency criterion was satisfied). The residues selected for this calculation are at most 3.5 Å away from the ligand (distance between centres of geometry, value selected by visual inspection). They include most of the active site, but not small, negligible contributions of residues that are further away that would slow down the polarization calculation.

With TvGSTO3S–benzophenone charge densities in hand, we could proceed to energy calculations. TvGSTO3S is a homodimeric enzyme, the crystal structure of which contains two independent monomers, each of which binds a benzophenone and a glutathione ligand. *Charger* calculated total electrostatic interaction and polarization energies between the two subunits and any available ligands. The calculations cover

all possible pairs (first monomer–first ligand, first monomer–second ligand *etc.*). We repeated these calculations with the *VMoPro* program from *MoProSuite* to compare the speed and accuracy of our new algorithm based on analytical EP/MM with the established numerical EP/MM method (Jelsch *et al.*, 2005).

Another reason for interest in the TvGSTO3S protein systems are calculations with S atoms. The benchmark systems from Nguyen *et al.* (2018) contain only CHNO atoms. Since their systems do not contain S atoms, the results do not fully confirm the applicability of analytical EP/MM to protein systems. TvGSTO3S contains several cysteine and methionine residues, and the glutathione ligand also contains an S atom. This allows us to draw more compelling conclusions on the applicability of the method.

We furthermore investigated the degree to which individual amino-acid residues contribute to the electrostatic interaction energy of overall protein–ligand binding. Trying to explain these results prompted us to form presumptive hypotheses. We constructed an *in silico* mutant as a first step towards the experimental verification of one such hypothesis: Arg171 had an unfavourable effect with some ligands.

We chose to mutate Arg171 to glutamine in the structure with PDB code 6f68 (Table 2). *PyMOL* served to generate the Arg171Gln mutation. The glutamine rotamer was chosen so that its side chain occupies a spatial position similar to that of arginine. We transferred the electron-density parameters to the resulting mutant structure and performed the same calculations as described above for the wild-type structures.

In all of these calculations with *Charger*, we applied analytical EP for all atom pairs that are closer than 5 Å apart

and pseudo-Buckingham multipolar moments calculation for all remaining atom pairs. We used both cutoff schemes from Table 1 to determine their quality empirically.

3. Results and discussion

3.1. *Charger* user interface

We integrated the *Charger* library into *MoProViewer* and created a user-friendly interface for it. *MoProViewer* facilitates molecule visualization and atom selection, while the interface handles parameter reorganization to a form understandable by *Charger*.

The options in the *Charger* interface, such as the selection of atomic constituents (for example nucleus, core, valence, multipolar ...), mirror those already available within the *VMoPro* program of *MoProSuite*. The user can select any set of atomic constituents from the Hansen–Coppens multipolar atom formalism (*i.e.* individual terms from equation 1), which then comprise the density used in energy calculations. As an example, this enables an inquiry into the interaction of the total protein density with ligand dipoles.

Furthermore, the *Charger* interface in *MoProViewer* allows control over the Buckingham (MM) domain threshold as a fixed distance or a threshold that differs from the sum of van der Waals radii of current atom pairs by a certain distance. The interface also handles the choice of multipolar cutoffs at very long distances, with the ‘precise’ cutoff criteria (Table 1) set as a default for interaction energy calculations.

We further designed a *Charger*-based calculation module in *MoProViewer* to provide more insight into protein–ligand interactions. It estimates the electrostatic interaction energy

contributions of individual residues to the protein–ligand interaction energy. It automatically selects residues one at a time and calls *Charger* to calculate residue–ligand energies. It is conceivable to use this tool to estimate pertinent amino acids at the interface between two proteins. All of the parameters that apply to a normal *Charger* calculation (Buckingham threshold, multipole–multipole cutoffs, atom composition) apply here as well.

Behind the scenes, the program library *Charger* takes as input molecular-structure and electron-density information through *MoProViewer* or some other interface. The electron density comes mostly from atomic tables describing electrically neutral and spherical atoms, along with some multipolar functions derived empirically to describe nonspherical density.

Once the molecular structure and its electron-density model are in place, *Charger* calculates the electrostatic interaction energy between any two groups of atoms, be it a protein and a ligand, one amino-acid residue and another, two discrete metal–ligand complexes, or even two ligands attached to the same metal centre. It skips unnecessary calculations: those that result in zero because of a fundamental theoretical reason such as symmetry or orthogonality of multipoles.

3.2. *Charger* performance

The first major point to discuss is the place of *Charger* in the universe of electrostatic interaction energy calculations. Its method approaches the core problem from a different perspective (analytical rather than numerical integration). It should therefore give similar results as other (numerical) methods that try to compute the same quantities. The challenges posed are many, from small molecules to large protein–ligand complexes. *Charger* promises faster calculation in all of these cases: a claim that we scrutinize after dealing with calculation accuracy.

We tested the analytical exact potential/pseudo-multipolar moments (aEP/pMM) method accuracy of *Charger* against the established numerical nEP/MM method available in *VMoPro*. In the case of benchmark molecules, we also compared it with available literature data (Nguyen *et al.*, 2018).

The energy results for benchmark calculations are very similar for the standard nEP/MM and our aEP/pMM method (Fig. 1, Table 3 and Supplementary Table S1). The differences are within 0.2 kcal mol^{−1} for the benchmark dimers, and are presumably due to numerical integration errors in the nEP/MM method. The energy values are also largely comparable to those available from Nguyen, Kisiel and Volkov (the aEP/MM method), despite the fact that our calculations relied on electron-density models with a different set of multipolar parameters. The largest relative differences come from dimers with an interaction energy value of ±10 kcal mol^{−1}, which one would expect from significantly different underlying models. Furthermore, *Charger* calculations using AMBER03 charges (Duan *et al.*, 2003) as spherical valence charges in the context of a Hansen–Coppens model agree well for almost all tested dimers (Table 3 and Supplementary Table S2). These results

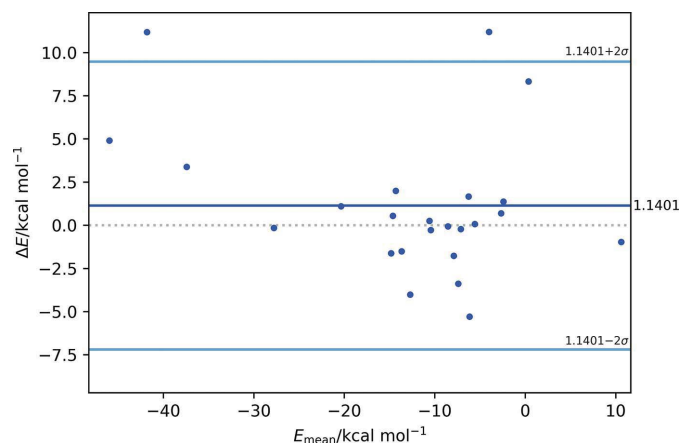


Figure 1

Bland–Altman plot to assess the difference between *Charger* electrostatic interaction energy calculations and those available from Nguyen *et al.* (2018) on benchmark systems. It depicts the difference in results between *Charger*–aEP/pMM (precise cutoff reported in Table 1) and the literature reference (*XDPROP*–aEP/MM, used as a reference: $\Delta E = 0$, dotted grey line) as a function of the magnitude of results. E_{mean} refers to the average value between the *XDPROP* and *Charger* results. Sigmas refer to standard deviations of ΔE values. The values from two Leu-enkephalin and two nonapeptide dimers were omitted for the reasons explained in the fourth and fifth footnotes to Supplementary Table S1.

Table 3

Electrostatic interaction energies of glycine...glycine dimers (in kcal mol⁻¹) for interactions in the benchmark glycine system.

δ refers to relative discrepancy, with the *Charger* precise values as the reference.

Program	Calculation	Gly1	Gly2	Gly3	Gly4	Gly5	Gly6
<i>Charger</i>	Precise	-27.9	-7.2	-19.8	-35.7	10.1	-5.5
<i>Charger</i>	Rapid	-27.9	-7.2	-19.8	-35.7	10.1	-5.5
<i>VMoPro</i>	nEP/MM	-27.8	-7.3	-19.8	-35.9	10.1	-5.6
<i>XDPROP</i>	Nguyen, Kisiel & Volkov paper (NKV)	-27.7	-7.0	-20.9	-39.1	11.1	-5.6
<i>Charger</i>	AMBER spherical charges ($P_{\text{val}} = Q - N_{\text{val}}$)	-23.2	-8.9	-16.7	-30.7	3.8	-4.5
<i>Charger</i>	AMBER point charges	-12.0	-0.4	-11.7	-18.5	6.7	-2.3
	δ (NKV) (%)	0.7	2.9	-5.3	-8.7	-9.0	-1.8
	δ (AMBER spherical charges) (%)	20	-19	19	16	166	22
	δ (AMBER point charges) (%)	133	1700	69	93	51	139

are comparable to those of Kumar *et al.* (2014), who performed an in-depth comparison between AMBER, Hansen–Coppens and several other models on the S66 and JSCH-2005 data sets. However, the agreement is not so good when using AMBER charges as point charges, which yield much lower energy values due to the missing penetration energy.

Calculation times for the aEP/MM and aEP/pMM methods are also of the same order of magnitude (Supplementary Table S3). *Charger* appears faster at first glance, particularly for small systems with many close atoms. A test under exactly the same conditions (same processor and molecular model) is needed before we can draw a more convincing conclusion. A discussion on calculation speed comparison between aEP/pMM and nEP/MM is given below.

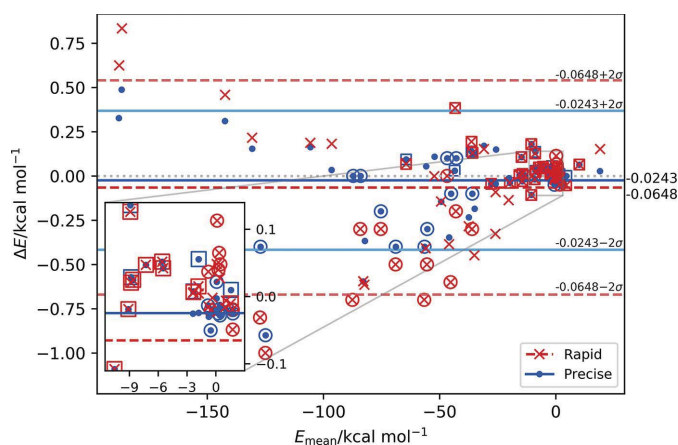


Figure 2

Bland–Altman plot to assess the accuracy of *Charger* electrostatic interaction energy calculations. It depicts the difference in results between the new methods (*Charger*–aEP/pMM, rapid/precise) and an old method (*VMoPro*–nEP/MM, used as a reference: $\Delta E = 0$, dotted grey line) as a function of the magnitude of the results. E_{mean} refers to the average value between the *VMoPro* and *Charger* results. Sigmas refer to standard deviations of ΔE values for the respective calculations. The graph contains energies from benchmark dimers (in squares) and GST–HBP interaction calculations (in circles, polarized density; others, nonpolarized density). The latter interaction energies include electrostatic and electrostatic + polarization energies. The inset magnifies the area around the origin for better assessment of the points clustered in this region. ‘Precise’ and ‘rapid’ refer to cutoff criteria as reported in Table 1.

We wanted to establish whether the results of the aEP/pMM method in *Charger* and the nEP/MM method from *VMoPro* are comparable. We used the same benchmark dimer energies (Supplementary Table S1) along with those between glutathione transferase (GST) and its glutathione and hydroxybenzophenone ligands (electrostatic and electrostatic + polarization energies are given in Supplementary Tables S4 and S5, respectively).

Fig. 2 presents a Bland–Altman plot of all of these energies, using nEP/MM energies from *VMoPro* as a reference. The y axis presents the difference between aEP/pMM and nEP/MM results, while the x axis contains the average of the two. The graph thus presents the difference in results between the two methods as a function of the magnitude of the energy results. It gives a feeling for how different the two methods are for large or small interaction energies (the results of both methods).

Results are on average below the reference *VMoPro*–nEP/MM method (by 0.02 kcal mol⁻¹ for ‘precise’ and 0.06 kcal mol⁻¹ for ‘rapid’ calculations) and most fall within two standard deviations (± 0.39 kcal mol⁻¹ for ‘precise’ and ± 0.60 kcal mol⁻¹ for ‘rapid’ calculations). Rapid calculations give values that are slightly more dissimilar, but are sometimes much closer to the mark.

Concerning the accuracy of the values, there is no direct comparison between aEP/pMM and the external methods here. Leduc *et al.* (2019) compared the nEP/MM method with symmetry-adapted perturbation theory (SAPT) energies obtained from high-level theoretical calculations on the S66x8 set (containing 66 small-molecule dimers at eight different distances). The correlation was strong ($R^2 = 97.0\%$, slope 1.05; intercept set to zero) without adding further empirical parameters. The r.m.s.d. for the whole data set was estimated as 1.4 kcal mol⁻¹. Taken together with the comparison between aEP/pMM and nEP/MM, one should report the energy values in kcal mol⁻¹ up to the first decimal point, while taking care not to overinterpret small differences in estimated energies.

On the other hand, the availability of the two independent protein monomers provides some insights into the variability of the computed energy due to the 3D structure itself. The relative difference $\Sigma|\Delta E|/\Sigma|E|$ of the ligand/glutathione energy values in Supplementary Table S5 is typically 20%.

The variability of electrostatic properties obtained by the use of different pseudoatom multipolar databases was verified by Bąk *et al.* (2011). The electrostatic energies E_{elec} of crystallographic dimers for a series of amino acids and dipeptides were computed using charge densities transferred from the ELMAM2, INVARIOM and UBDB electron-density databases and were complemented by first-principles calculations (MP2 or DZP). The correlation coefficient of the E_{elec} values from the different calculations was in the range 0.79–0.99 and the observed relative difference $\Sigma|\Delta E|/\Sigma|E|$ was about 24% for energy values issued from the three databases.

The GST structure contains one cysteine and several methionine residues, and the glutathione ligand is a tripeptide (L-glutamyl-L-cysteinylglycine). All of these moieties contain S atoms. The aEP/MM method has not been tested on S atoms to the best of our knowledge (Nguyen *et al.*, 2018; Nguyen & Volkov, 2019). Fig. 2 shows the absolute discrepancies between the *VMoPro*-nEP/MM and *Charger*-aEP/pMM results, all of which are lower than 1 kcal mol⁻¹. The largest differences come from calculations with the largest results (over 150 kcal mol⁻¹), corresponding to relative differences of about 1%. These observed differences are attributable to the differences in the methods used to compute these energies. In addition, these results show that *Charger* (and aEP in general) appears to be suitable for calculations with the CHNOS atoms found in protein structures.

Supplementary Tables S3 and S6 contain the wall-clock times required to calculate all of the interaction energies described thus far on an Intel Core i7-8700 (3.20 GHz) processor. We estimate a heavy computation load from the amount of close atom pairs in the calculation, since atoms that are close together elicit the more time-consuming aEP or nEP calculation. Calculation speedup in *Charger* portrayed as a function of close atom pairs (less than 5 Å apart) is shown in Fig. 3 for GST–ligand dimers and in Supplementary Fig. S1 for all examined dimers.

We observe a significant (30–80-fold on average) increase in the speed of *Charger* analytical calculations compared with *VMoPro* numerical integration for GST–ligand dimers. In the case of benchmark systems, the improvement is even larger (up to 150–200-fold). This improvement increases with the number of close atom pairs in the calculation. Closer atoms require more dedication to calculate all interaction integrals, and this afflicts the *VMoPro*-nEP/MM computational time much more severely.

This does not hold true when there are no close atom pairs. Calculating the interaction energy between a GST monomer and the hydroxybenzophenone ligand located in the binding site of the other monomer takes longer for *Charger* (Supplementary Table S6). In this case, *VMoPro* is faster (up to ten times, or 1–3 s in real terms), probably because it includes a proper implementation of the Buckingham potential and *Charger* includes a pseudo-Buckingham approach.

One could therefore imagine that *VMoPro* would be faster even when the ligand has few close interactions (*i.e.* few close pairs of atoms) and many distant interactions. The GST monomer and the glutathione ligand from the binding site of

the other monomer provide such a case: there are as few as six close atom pairs per 10 000. Here, *Charger* is already 4.5 to 15 times faster (precise/rapid calculation) than *VMoPro*-nEP/MM because the analytical calculation for close atom pairs is preponderant and is much faster than numerical calculation.

When comparing cutoffs, ‘rapid’ calculations with a more severe threshold (15 Å for low-order multipoles) are about 20% faster for small systems with many atoms in close contact, such as Leu-enkephalin dimers (77 × 77 atoms; Supplementary Fig. S2). Larger systems with few or zero atoms in close contact run 1.5–3 times faster (GST monomer with benzophenones, 3706 × 26 or 27 atoms, or with glutathione, 3706 × 36 atoms; or a dimer with twice as many atoms).

‘Rapid’ calculation does not suffer from a large loss of accuracy (Fig. 2). Using our processor, the saved time in real terms amounts to a tenth of a second for small systems and between 2 and 11 s for larger systems. The benefits of using a 15 Å threshold for single-point calculations on small to mid-size proteins are modest, but a large-scale calculation (larger protein, database search) would profit significantly.

MoProViewer allows users to select two groups of interacting atoms for energy computation. To explore the energy contributions of individual amino-acid residues to the overall protein monomer–ligand interaction energy, *MoProViewer* silently selects each residue and then uses *Charger* to compute its interaction energy with the chosen ligand. This feature has a very small overhead of 1% in computation time compared with calculating the total energy between the ligand and a whole monomer.

It is possible to verify these individual residue results with the *VMoPro*-nEP/MM method. It already has the capacity to calculate several interaction energies arising from multiple atom selections within one structure. Accessing this capability

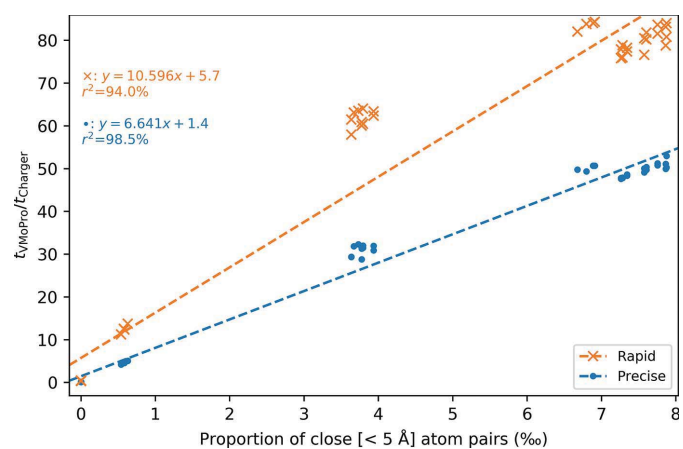


Figure 3 Computation speed enhancement. Improvement in *Charger* calculation speed compared with *VMoPro*-EP/MM. The graph focuses only on data from protein–ligand systems, which have few close atom pairs. The x axis represents the proportion of close atom pairs (in promille, ‰), *i.e.* the number of atoms closer than 5 Å apart divided by the number of all atom pairs. Precise and rapid cutoff criteria are noted in Table 1. Supplementary Fig. S1 shows the full graph with protein–ligand and benchmark calculations.

in *MoProViewer* requires implementing a function to generate the appropriate *VMoPro* scripts from *MoProViewer* selections, which was not undertaken. We expect such a calculation to be nearly as fast as a calculation between the whole protein and the ligand using *VMoPro*.

3.3. Interaction analysis with *Charger*

Having established *Charger* as a reliable tool to calculate electrostatic interaction energy quickly even for biomacromolecules, we turned our attention to investigating these same molecules. The possibility of correlating protein–ligand energies with experimental results (thermal shift assays) piqued our interest.

In the study by Schwartz and coworkers, the crystallographic results seem to have a good agreement with the thermal shift assay results in solution (*i.e.* observed differences in denaturation temperatures; ΔT_d). It was not possible to obtain the crystallographic structures of complexes with the hydroxybenzophenones that had the lowest ΔT_d values. On the contrary, the HBPs with the highest ΔT_d values seemed to be well bound in the active site. Investigating the individual

residue–ligand concept became so appealing that we created a special module in *MoProViewer* dedicated to it.

Firstly, a short introduction to the protein–ligand complex is in order. *Charger* determined the electrostatic energy of the interaction between the glutathione transferase Omega 3S from *Trametes versicolor* (*TvGSTO3S*) and four hydroxybenzophenone inhibitors (HBPs; Table 2). *TvGSTO3S* is a dimeric enzyme; each monomer contains an active site with a glutathione-binding site (G-site) and a hydrophobic pocket for the electrophilic substrate (H-site) (Schwartz *et al.*, 2018).

Hydroxybenzophenones sit in the H-site, where 2,4,4'-HBP has a distinct conformation from those of 2,4-, 2,3,4- and 3,4-HBP (Fig. 4 and Supplementary Fig. S3). Thermal shift assays show that the latter three had a greater thermal stability compared with the 2,4,4'-HBP complex. The crystal structure of this complex (PDB entry 6f68) shows higher thermal motion and the disordered character of this ligand (a main conformation was modelled with an occupancy factor of 0.79) is consistent with its low thermal shift of only $\Delta T_d = 2.9$ K.

Schwartz *et al.* (2018) did not establish a quantitative correlation between the thermal shifts and the ligand-binding strengths. Interestingly, these thermal shifts and *Charger*

energies for monomers and their own ligands are well correlated (correlation coefficient values generally above 85%; see Fig. 5 and Supplementary Fig. S4 and footnotes to Supplementary Table S7). The correlations of thermal shifts with the interaction energies of one monomer and the ligand bound in the second monomer are much poorer (correlation coefficient values generally below 40%). Stronger correlations are expected in the first case, since the ligand within the binding site normally has a stronger influence on the overall protein–ligand complex stability.

The order of thermal shift values measured for the considered complexes with benzophenones nearly perfectly matches that of the electrostatic ligand-binding energies computed with *Charger*. Trend-line slopes (around -6.5 kcal mol⁻¹ K⁻¹ for monomers and -12 kcal mol⁻¹ K⁻¹ for the dimer) indicate that we should detect a different thermal shift by at least 1 K if we calculate a difference of around -6 kcal mol⁻¹ per monomer. As expected, these numbers are higher when considering polarization (around -9 kcal mol⁻¹ K⁻¹ for monomers and -17 kcal mol⁻¹ K⁻¹ for the dimer). An experimentally observable difference in the thermal shift of 1 K should arise for a small structure modification (such as a slight chemical modification of a ligand

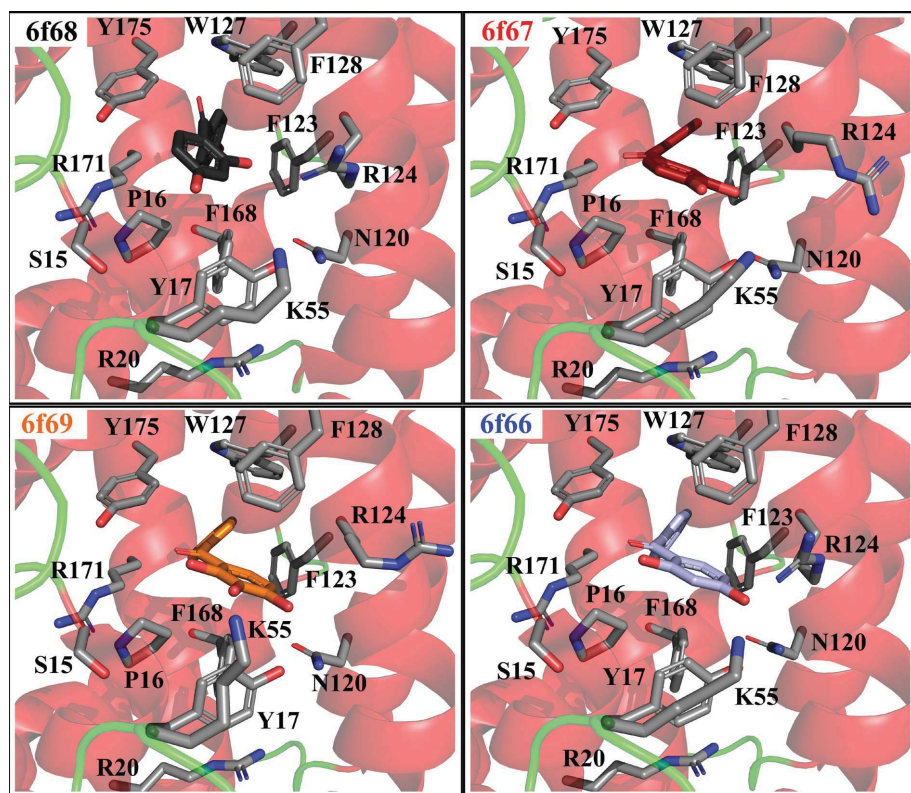


Figure 4
Binding of benzophenones (2,4-, 3,4-, 2,4,4'- and 2,3,4-HBP) in the *TvGSTO3S* H-site (Schwartz *et al.*, 2018; PDB entries 6f66, 6f67, 6f68 and 6f69, respectively). The *TvGSTO3S* H-site is a well delineated deep cavity. Polar residues (Tyr17 and Arg124) at the cavity entrance close to the glutathione binding site and the catalytic residue Ser15 complete the H-site. Side chains and benzophenone ligands are represented as sticks. The conformational difference between the ligand in the PDB entry 6f68 structure and the other ligands resides in the dihedral angles between the carbonyl plane and the aromatic rings (Schwartz *et al.*, 2018). The ligand colouring follows the convention in Table 2.

or a residue mutation) that changes the interaction energy per monomer by roughly -6 or -9 kcal mol $^{-1}$ (without or with polarization). This could serve as a rough guide for ligand design (for example with docking) or a protein mutation study. Thermal shift analysis is a commonly used tool in lead compound identification (screening) during drug discovery and to a lesser extent in lead optimization (Pantoliano *et al.*, 2001; Xing *et al.*, 2019).

A typical analysis of the three-dimensional structure of a macromolecular complex highlights assumed interactions that favour macromolecule–small molecule adhesion. This analysis involves visual inspection and distance measurements. Schwartz *et al.* (2018) performed this ‘manual analysis’ in the complexes of TvGSTO3S with benzophenone ligands. They assumed that the H-site of TvGSTO3S is suited to accommodate HBPs due to the presence of polar residues (Tyr17 and Arg124), aromatic side chains (Phe123, Trp127, Phe128, Phe168 and Tyr175) and the aliphatic part of Arg171 (Fig. 4).

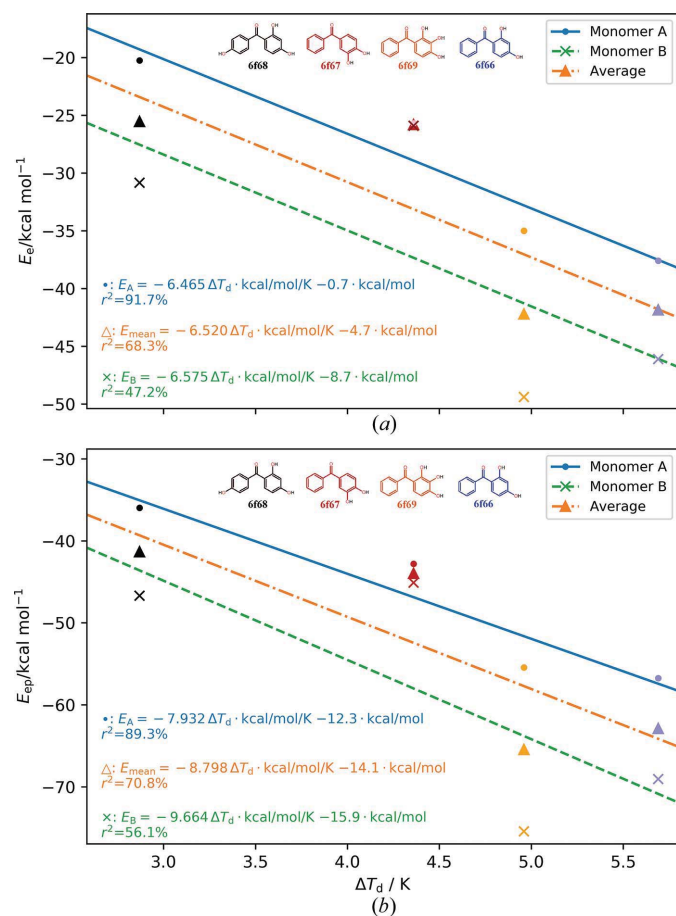


Figure 5

Relationship between estimated experimental thermal shifts and computed energies in GST monomers. Relationship between computed electrostatic interaction energies (a) E_e without polarization and (b) E_{ep} with polarization from *Charger* calculations and estimated experimental thermal shifts for glutathione transferase and four ligands. The quantities E_e and E_{ep} represent interaction energies between each monomer and the ligand bound in its active site, along with their mean values. Thermal shift values are given in Supplementary Table S7.

Charger calculates the contribution (favourable or unfavourable) of each individual residue to the total interaction energy between a ligand and TvGSTO3S. Observing the interaction energy contributions of all 240 individual residues reveals that the two monomers are roughly similar, with a remarkably high degree of intramonomer correlation (Supplementary Table S8). Analysis of the structures indeed revealed no difference between the subunit main chains (mean r.m.s.d. of 0.12 Å). There is even a pronounced correlation between the per-residue interaction energies for monomers hosting different ligands, except for the ligand with a 4'-hydroxyl group (PDB entry 6f68, 2,4,4'-HBP; Supplementary Table S8). The uncommon position of the 4'-hydroxyl ligands within the binding site can explain these weaker correlations.

We went through the individual binding contributions to find the most interesting residues: those with the most pronounced electrostatic interaction energy. We realized that many residues make a minor contribution (Supplementary Fig. S5). Taken at the protein scale, these minor favourable and unfavourable contributions roughly balance themselves out.

We then used a simple filtering criterion: at least one ligand–residue electrostatic interaction energy surpasses ± 3 kcal mol $^{-1}$, corresponding roughly to a 3σ standard deviation (Fig. 6). As expected, the retained residues included most of the residues found by visual inspection (Tyr17, Arg124, Trp127, Phe128, Phe168, Arg171 and Tyr175), with the notable exception of Phe123. Interestingly, *Charger* also identified residues not suspected from a visual search: Pro16, Arg20, Lys55 and Asn120. We inspected these residues closely, and detail here the interactions of Pro16, Tyr17, Phe168 and

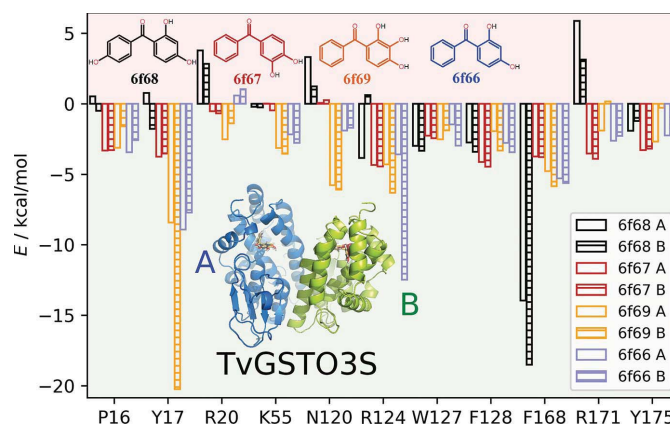


Figure 6

Energy contribution of protein residues to ligand binding. Negative electrostatic (without polarization) interaction energy contributions of individual residues to GST–ligand binding for the two GST monomers (A and B). Residues with no energy contribution larger than ± 3 kcal mol $^{-1}$ are omitted for clarity. Negative numbers on the y axis indicate a favourable binding contribution (green background) and positive numbers indicate unfavourable binding (red background). The legend refers to the PDB codes given in Table 2. The ligands follow the ascending order of their thermal shifts.

Arg171. Other residues have similar energy profiles (Fig. 6) and explanations on the atomic level.

Tyr17 contributes favourably to the binding of 2,4-, 3,4- and 2,3,4-HBP, but not significantly to that of 2,4,4'-HBP. In the first three cases the Tyr17 side chain binds the benzophenone 4-hydroxyl group *via* a hydrogen bond. The particular 2,4,4'-HBP conformation disallows the formation of an equivalent hydrogen bond (Schwartz *et al.*, 2018), so the contribution of this residue becomes negligible.

The contribution of Phe168 is favourable in all cases due to C—H... π interactions. Contrary to the previous case, it contributes much more to 2,4,4'-HBP binding due to an additional hydrogen bond between the Phe168 carbonyl group and the 4'-hydroxyl group (see Supplementary Fig. S6 and Section S2). The 2,4,4'-HBP ligand has a particular binding mode that differs from the other three benzophenones that do not have a 4'-hydroxyl group. The binding of this ligand is not favoured by several nearby residues (Arg20, Asn120 and Arg171), as their interaction energies are unfavourable (the sum of electrostatic interaction energies for the cited residues is +10.1 kcal mol⁻¹, averaged between two monomers; Fig. 6).

However, 2,4,4'-HBP forms a 4'-O—H...O=C hydrogen bond with the Phe168 carbonyl group. The O...O distances of 2.75 and 2.66 Å (monomers A and B) are shorter than 3 Å, which enables them to be classified as strong hydrogen bonds (MacLeod & Rosei, 2014). The balance between these unfavourable contributions and this strong hydrogen bond specific to the 4'-OH substitution might induce this ligand to adopt the specific binding mode observed in the crystal structure.

The calculated electrostatic interaction energy reflects this trend (−16.2 kcal mol⁻¹ for 2,4,4'-HBP, averaged over two binding sites, compared with -4.8 ± 0.9 kcal mol⁻¹ for the 4'-hydrogen ligands, average and standard deviation for three ligands in two binding sites). The value of -4.8 ± 0.9 kcal mol⁻¹ roughly corresponds to the interaction energy between a ligand aromatic ring and the Phe168 phenyl ring, an interaction that is present for all four ligands. The described strong hydrogen bond in the case of 2,4,4'-HBP could be responsible for the additional energy of roughly −11 kcal mol⁻¹. One should keep in mind the desolvation penalty: the effect of hydrogen bonds on ligand-binding strength is often strongly attenuated due to almost isoenergetic hydrogen bonds with the bulk water solvent.

Charger pointed towards an amino acid that was not deemed to be significant by visual inspection in the study by Schwartz and coworkers: Pro16. This residue interacts favourably with 2,4-, 3,4- and 2,3,4-HBP (-5.3 ± 0.9 kcal mol⁻¹, mean \pm standard deviation) and interacts more weakly with 2,4,4'-HBP (−0.8 kcal mol⁻¹ in one binding site and −2.4 kcal mol⁻¹ in the other). A closer contact between Pro16 and 2,4-, 3,4- and 2,3,4-HBP might in part explain these differences. The shortest interatomic distance (from proline C ^{δ} to a ligand O atom) is 4.5 Å for 2,4,4'-HBP and 3.7 Å for the three other benzophenones.

Visual inspection of the *Tv*GSTO3S–HBP crystal structures predicts weak but favourable C—H... π and van der Waals interactions between the HBP ligands and the aliphatic part of the Arg171 side chain (Schwartz *et al.*, 2018). A different interpretation appears under *Charger's* microscope. When considering only permanent multipoles, it indeed predicts a strongly favourable interaction energy between Arg171 and 2,4-, 3,4- and 2,3,4-HBP (-2.6 ± 1.1 kcal mol⁻¹, mean \pm standard deviation). This interaction appears similar to that with Pro16, but is slightly stronger. Surprisingly, and unlike in the

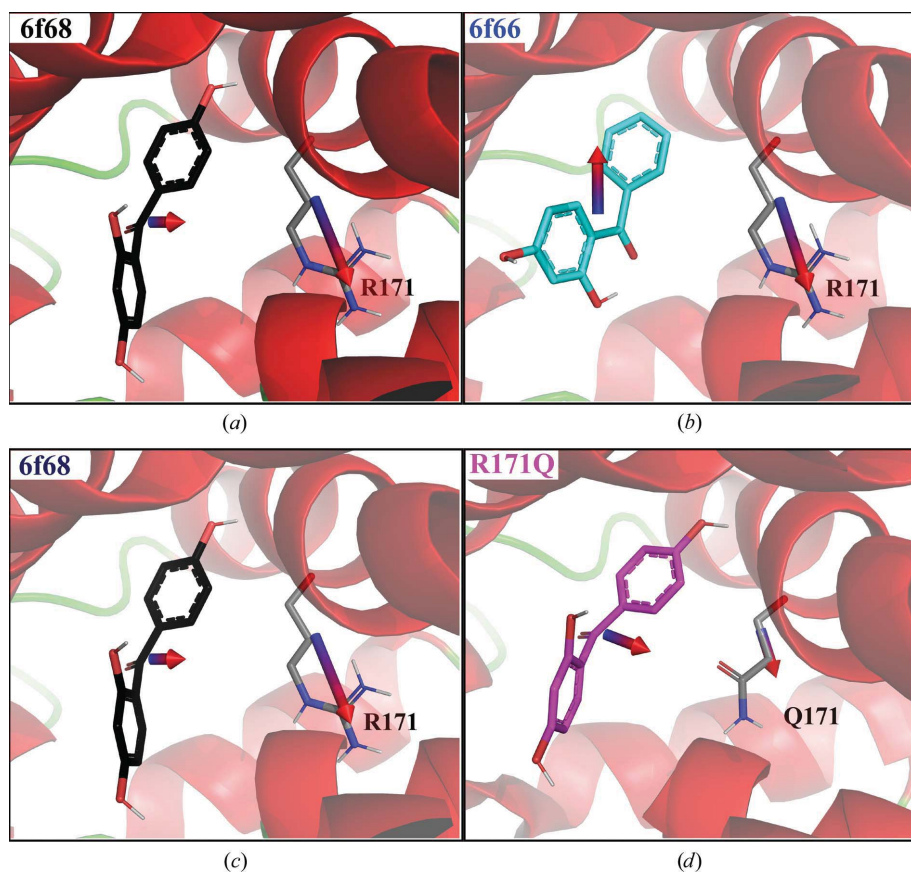


Figure 7

Ligand and residue 171 dipole moments. Strength (arrow length, scaled to 0.21 Å/D) and relative orientation of dipole moments for residue 171 and benzophenone ligands. First row: Arg171 (on the right of both figures) with (a) a 4'-hydroxy (black) and (b) a 4'-hydrogen (orange) ligand. Second row: the 4'-hydroxy ligand (on the left of both figures) with wild-type Arg171 in green, as in (a), and (d) mutated Glu171 in light blue (d). We chose glutamine as a mutation candidate for Arg171 that could test the veracity of our computational method. All figures depict binding site A.

Pro16 case, it detects an unfavourable interaction energy between Arg171 and 2,4,4'-HBP of about $+4.5 \text{ kcal mol}^{-1}$ (Fig. 6).

Calculation and analysis of molecular dipole moments using *VMoPro* and *MoProViewer* provide some explanations of how Arg171 affects ligand binding. The dipole moments of 2,4,4'-HBP and Arg171 are at angles of 63° and 87° in the two monomers (Supplementary Table S9, Fig. 7a). The positive charge distributions of Arg171 and 2,4,4'-HBP seem to interfere. This renders the contact between the ligand and the amino-acid side chain unfavourable. On the other hand, the antiparallelism between the dipole moment of 2,4-, 3,4- and 2,3,4-HBP and that of Arg171 (angles of around $150\text{--}170^\circ$; Supplementary Table S9, Fig. 7b) could signal a favourable Keesom interaction (the interaction between permanent dipoles; Keesom, 1921).

Furthermore, there is a narrow solvent-accessible cavity near the surface of the Arg171 side-chain extremity (Supplementary Fig. S7). This cavity is likely to contain a counterion that would alleviate the influence of Arg171. The counterion would be located further away from the ligand than Arg171, so its influence should be smaller than that of the Arg171 charge.

The addition of polarization density supports the general conclusion. It largely compensates for the negative effect of Arg171 on 2,4,4'-HBP binding (electrostatic interaction energy of about $+0.8 \text{ kcal mol}^{-1}$) and improves the standing of the other ligands (interaction energy of $-7.9 \pm 1.6 \text{ kcal mol}^{-1}$). Although Arg171 does not hinder 2,4,4'-HBP binding severely, as one would conclude without polarization density, a mutation of this residue could still improve ligand binding (see Section 3.4).

3.4. Interaction prediction with *Charger*

The results of the detailed investigation of GST–benzophenone binding prompted us to consider research perspectives. One promising offshoot is a mutation study on GST.

We have tried to improve the binding of 2,4,4'-HBP by an *in silico* mutation. Arg171 intrigued us through its seemingly unfavourable dipole interaction with this ligand. An attenuation of the arginine charge and polar contributions could establish whether this contribution is responsible for the weakly unfavourable binding energy.

We first considered alanine, but dismissed it since replacing the large arginine residue by a small side chain would leave a large void in the active site. We looked for a suitable neutral amino acid of approaching size. We supposed that mutating Arg171 to a glutamine should have an effect on dipole moment due to the introduction of a relatively electronegative O atom and to the loss of the formal charge of arginine.

The dipole moment of an *in silico* mutated Gln171 is much smaller than that of Arg171. The dipole orientation relative to 2,4,4'-HBP remains similar to that found in the wild-type Arg171 (Figs. 7c and 7d, Supplementary Table S9). Furthermore, the addition of polarization density shows how the mutation affects the entire binding pocket. The negative effect of the wild-type Arg171 on 2,4,4'-HBP binding diminishes

when taking polarization into account (average interaction energy of about $+0.8 \text{ kcal mol}^{-1}$ instead of $+4.5 \text{ kcal mol}^{-1}$ without polarization). The Arg171Gln mutant remains favourable: the Gln171–2,4,4'-HBP interaction energy (with polarization) improves to around $-2.1 \text{ kcal mol}^{-1}$ (Supplementary Table S9). This improvement is likely to come from neutralizing the dipole effect, largely due to the lack of charge on glutamine. An analysis of electric field lines underscores this interpretation (Supplementary Fig. S6 and Section S2).

The polarization pronouncedly affects the transferred electron density in the entire binding site. The electrostatic interaction energies (including polarization) with the 2,4,4'-HBP ligand are -41.4 and $-84.0 \text{ kcal mol}^{-1}$ for the wild-type Arg171 monomer average and dimer, compared with a more favourable -50.3 and $-101.6 \text{ kcal mol}^{-1}$ for the mutant Gln171 monomer average and dimer. The projected energy difference should correspond to about a 1 K difference in thermal shift, making Arg171Gln an attractive mutation candidate.

4. Conclusions

The methodology presented here and in Nguyen *et al.* (2018) gives results that are almost identical to the established nEP/MM numerical integration, but in a fraction of the time. *Charger* determines electrostatic interaction energies with a precision within bounds of $\pm 1 \text{ kcal mol}^{-1}$ for a given charge-density model. This remains true even though several of the newly studied systems contain previously uninvestigated sulfur atoms. It has to be recalled, however, that using different charge-density models, such as ELMAM2, UBDB and INVARIOM database transfer, results in much larger variations (Bač *et al.*, 2011) and the accuracy is estimated about at one quarter in relative value. The purpose of *Charger* is to calculate electrostatic energies, so it does not currently have any features to calculate other contributions to the total interaction energy. van der Waals energy-term computation is possible with the *VMoPro* module, using the Lennard–Jones parametrizations proposed in the literature.

The *raison d'être* of *Charger* is speed. Its aEP/pMM calculations are 30–80 times faster than *VMoPro*–nEP/MM. They take several seconds for the largest examined systems: protein–ligand complexes. This makes *Charger* an excellent tool to explore crystal or protein electrostatic interaction energies in real time. One can demonstrably improve the speed of *Charger* easily by using appropriate cutoffs, while guarding great accuracy: for example, in a systematic database search where speed becomes a concern.

We have studied the glutathione–benzophenone interaction in detail using *Charger*. The order of computed *Charger* protein–ligand energies and experimental thermal shifts is identical in nearly all cases, with high correlation coefficients. The slope of the trend line helps to decide which *in silico* investigations to hand over to the laboratory.

We have used *Charger* calculations further to determine the contributions of each residue to benzophenone ligand fixation. *Charger* determines important residues from the electrostatic

point of view: Schwartz *et al.* (2018) already suspected most of them from a visual analysis.

Charger also uncovered residues that are adverse to benzophenone fixation. Particularly interesting was the case of Arg171, which interacts favourably with three benzophenone ligands, but not with 2,4,4'-HBP. We suspected that mutating Arg171 to a noncharged residue such as a glutamine could remove this electrostatic obstacle to binding. This adds an alternative to strategies that aim to improve or increase the number of existing favourable contacts (Thapa & Raghavachari, 2019).

An *in silico* Arg171Gln mutant meets our expectations. The polarization correction of the transferred charge density (Leduc, 2019; Leduc *et al.*, 2019) helped to determine the importance of this mutation. If the production and purification of the Arg171Gln mutant turns out to be feasible, thermal shift assays could assess this prediction in the future. Further electrostatic energy studies on numerous protein–ligand systems with more thermodynamic data such as binding constants and enthalpy of binding are needed to assess the importance of polarizing charge densities based on transferred multipolar parameters.

One can easily integrate *Charger* into different programs. It supports Slater-type basis functions in general, and Hansen–Coppens charge-density models in particular. We have released the library code under the permissive Apache 2.0 licence to encourage integration into other projects. Our flagship program *MoProViewer* now comes with a simple and intuitive *Charger* interface. Both the *Charger* and *MoProViewer* codes can be found at <https://crm2.univ-lorraine.fr/lab/fr/software/mopro/>.

5. Related literature

The following references are cited in the supporting information for this article: Clementi & Roetti (1974), Koga *et al.* (1993, 2000) and Michael & Volkov (2015).

Funding information

VV is grateful for a PhD fellowship from MENSRT.

References

Bąk, J. M., Domagała, S., Hübschle, C., Jelsch, C., Dittrich, B. & Dominiak, P. M. (2011). *Acta Cryst.* **A67**, 141–153.
 Bojarowski, S. A., Kumar, P. & Dominiak, P. M. (2016). *ChemPhysChem*, **17**, 2455–2460.
 Buckingham, A. D. (1959). *Q. Rev. Chem. Soc.* **13**, 183–214.
 Cimperman, P., Baranauskienė, L., Jachimovičiūtė, S., Jachno, J., Torresan, J., Michailovienė, V., Matulienė, J., Sereikaitė, J., Bumelis, V. & Matulis, D. (2008). *Biophys. J.* **95**, 3222–3231.
 Clementi, E. & Roetti, C. (1974). *At. Data Nucl. Data Tables*, **14**, 177–478.

Dittrich, B., Koritsánszky, T. & Luger, P. (2004). *Angew. Chem. Int. Ed.* **43**, 2718–2721.
 Domagała, S., Fournier, B., Liebschner, D., Guillot, B. & Jelsch, C. (2012). *Acta Cryst.* **A68**, 337–351.
 Dominiak, P. M., Volkov, A., Dominiak, A. P., Jarzemska, K. N. & Coppens, P. (2009). *Acta Cryst.* **D65**, 485–499.
 Dominiak, P. M., Volkov, A., Li, X., Messerschmidt, M. & Coppens, P. (2007). *J. Chem. Theory Comput.* **3**, 232–247.
 Duan, Y., Wu, C., Chowdhury, S., Lee, M. C., Xiong, G., Zhang, W., Yang, R., Cieplak, P., Luo, R., Lee, T., Caldwell, J., Wang, J. & Kollman, P. (2003). *J. Comput. Chem.* **24**, 1999–2012.
 Fournier, B., Bendeif, E. E., Guillot, B., Podjarny, A., Lecomte, C. & Jelsch, C. (2009). *J. Am. Chem. Soc.* **131**, 10929–10941.
 Guillot, B., Enrique, E., Huder, L. & Jelsch, C. (2014). *Acta Cryst.* **A70**, C279.
 Hansen, N. K. & Coppens, P. (1978). *Acta Cryst.* **A34**, 909–921.
 Hoof, R. W. W., Sander, C. & Vriend, G. (1996). *Proteins*, **26**, 363–376.
 Huang, L., Massa, L. & Karle, J. (2005). *Biochemistry*, **44**, 16747–16752.
 Jelsch, C., Guillot, B., Lagoutte, A. & Lecomte, C. (2005). *J. Appl. Cryst.* **38**, 38–54.
 Jones, H. W. (1993). *Int. J. Quantum Chem.* **45**, 21–30.
 Keesom, N. (1921). *Z. Phys.* **22**, 129–141.
 Koga, T., Kanayama, K., Watanabe, T., Imai, T. & Thakkar, A. J. (2000). *Theor. Chem. Acc.* **104**, 411–413.
 Koga, T., Tatewaki, H. & Thakkar, A. J. (1993). *Phys. Rev. A*, **47**, 4510–4512.
 Kumar, P., Bojarowski, S. A., Jarzemska, K. N., Domagała, S., Vanommeslaeghe, K., MacKerell, A. D. Jr & Dominiak, P. M. (2014). *J. Chem. Theory Comput.* **10**, 1652–1664.
 Kumar, P., Gruza, B., Bojarowski, S. A. & Dominiak, P. M. (2019). *Acta Cryst.* **A75**, 398–408.
 Leduc, T. (2019). PhD thesis, Université de Lorraine, France. <https://hal.archives-ouvertes.fr/tel-02528924/>.
 Leduc, T., Aubert, E., Espinosa, E., Jelsch, C., Iordache, C. & Guillot, B. (2019). *J. Phys. Chem. A*, **123**, 7156–7170.
 Löwdin, P.-O. (1956). *Adv. Phys.* **5**, 1–171.
 MacLeod, J. M. & Rosei, F. (2014). *Small*, **10**, 1038–1049.
 Mandal, S. K., Saha, P., Munshi, P. & Sukumar, N. (2017). *Struct. Chem.* **28**, 1537–1552.
 Massa, L., Keith, T., Cheng, Y. & Matta, C. F. (2019). *Chem. Phys. Lett.* **734**, 136650.
 Michael, J. R. & Volkov, A. (2015). *Acta Cryst.* **A71**, 245–249.
 Nguyen, D., Kisiel, Z. & Volkov, A. (2018). *Acta Cryst.* **A74**, 524–536.
 Nguyen, D., Macchi, P. & Volkov, A. (2020). *Acta Cryst.* **A76**, 630–651.
 Nguyen, D. & Volkov, A. (2019). *Acta Cryst.* **A75**, 448–464.
 Pantoliano, M. W., Petrella, E. C., Kwasnoski, J. D., Lobanov, V. S., Myslik, J., Graf, E., Carver, T., Asel, E., Springer, B. A., Lane, P. & Salemme, F. R. (2001). *J. Biomol. Screen.* **6**, 429–440.
 Schwartz, M., Perrot, T., Aubert, E., Dumarçay, S., Favier, F., Gérardin, P., Morel-Rouhier, M., Mulliert, G., Saiag, F., Didierjean, C. & Gelhaye, E. (2018). *Sci. Rep.* **8**, 8472.
 Thapa, B. & Raghavachari, K. (2019). *J. Chem. Inf. Model.* **59**, 3474–3484.
 Volkov, A., Koritsánszky, T. & Coppens, P. (2004). *Chem. Phys. Lett.* **391**, 170–175.
 Xing, J., Zhang, R., Jiang, X., Hu, T., Wang, X., Qiao, G., Wang, J., Yang, F., Luo, X., Chen, K., Shen, J., Luo, C., Jiang, H. & Zheng, M. (2019). *Eur. J. Med. Chem.* **163**, 281–294.