STRUCTURAL BIOLOGY

Volume 77 (2021)

Supporting information for article:

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

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

S1. Fine points of *Charger* calculations

S1.1. Theoretical background

We use the Hansen-Coppens model, (Hansen & Coppens, 1978) which is the cornerstone of experimental charge density modelling. It expresses electron density around an atomic nucleus as a sum of four components (we reproduce it here for clarity):

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

The first term describes a point-charge nucleus at fixed coordinates ($r = 0$). The next two terms describe spherical core and valence densities ($\rho_{\mathrm{core}}(r)$ and $\rho_{\mathrm{valence}}(\kappa, r)$, with population coefficients N_{core} and P_{val}). Finally, the third contains density-normalised (Michael & Volkov, 2015) spherical harmonic functions $d_{lm}(\theta, \phi)$. Such terms are called 'multipoles' when discussing charge distributions, so the Hansen-Coppens model is also well-known as the multipolar model. The function $R_l(\kappa', r)$ represents the radial part of the multipole, while P_{lm} is the multipolar population coefficient.

The coefficients κ, κ', P_{val} and P_{lm} (there are $(l_{max} + 1)^2$ of the last ones) describe the electron density surrounding one atom beyond the standard approximation (the "independent atom model", IAM, *i.e.* atoms as independent spheres). The terms $\rho_{core}(r)$ and $\rho_{valence}(\kappa, r)$ consist of a sum of Slater-type exponential functions:

$$R(\kappa, r) = N r^n \, \mathrm{e}^{-\zeta \kappa r} \quad (10),$$

with the normalization factor:

$$N = \kappa^3 (2\zeta)^n \sqrt{\frac{2\zeta}{(2n)!}} c \quad (11).$$

The term $R_l(\kappa', r)$ looks the same as $f(\kappa, r)$, but n depends on the multipolar level: $n = n_l$. The coefficients n (n_l) and ζ are Slater-type parameters, and c being a coefficient arising from the linear combination of Slater-type orbital functions. One can determine these coefficients from atomic wave-function calculations. Literature on atomic wave functions typically contains all three values (Clementi & Roetti, 1974; Koga *et al*., 1993; Koga *et al*., 2000). It is common to see the κ^3 term explicitly in the Hansen-Coppens equation throughout literature (like in the main body of the article), instead of in the normalisation constant.

The following integral determines the total electrostatic interaction energy between two atoms A and B represented by their total charge densities $\rho_A(\mathbf{r}_A)$ and $\rho_B(\mathbf{r}_B)$ (reproduced from the main text):

$$E_{\mathrm{Total}} = \iint \frac{\rho_A(\mathbf{r}_A)\rho_B(\mathbf{r}_B)}{|\mathbf{r}_A - \mathbf{r}_B|} \mathrm{d}\mathbf{r}_{\mathrm{A}} \mathrm{d}\mathbf{r}_{\mathrm{B}} \quad (2).$$

Many different terms come out of inputting the Hansen-Coppens model into this integral: the nucleus-nucleus term, three nucleus-electron terms (nucleus-core, nucleus-valence, nucleus-multipoles), and another nine different electron-electron energy terms. Regrouping them into six contributions simplifies the total energy expression (reproduced from the main text):

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

The index "n" denotes nuclei, "s" denotes spherical (core and valence) parts of electron clouds (the two middle terms of the multipolar model, eq. 1), and "m" denotes the multipolar distribution part from electron clouds (last term of the multipolar model, eq. 1).

The first term of the energy sum is the simplest (all equations use atomic units):

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

with z_A and z_B representing nuclear charges of the two atoms, and d the distance between them. The second term (E_{ns}, nucleus-spherical) is as simple as a multiplication of nuclear charge and the electrostatic potential:

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

The spherical part of the electron density generates the electrostatic potential Φ_B, which we evaluate at the inter-atomic distance d. Jones (1993) introduced the radial electrostatic potential of an electron cloud described as a Slater-type function (eq. 10):

$$\Phi_B(r) = k\,[\Phi_1(r) + \Phi_2(r) + \Phi_3(r)] \quad (12),$$

with:

$$k = \frac{4\pi}{2l+1} \quad (13),$$

$$\Phi_1(r) = \frac{\nu!}{\zeta^{\nu+1}}(\kappa r)^{-l-1} \quad (14),$$

$$\Phi_2(r) = e^{-\zeta\kappa r} \sum_{k=0}^{\nu} \frac{\nu!}{(\nu-k)!\zeta^{k+1}}(\kappa r)^{\nu-k-l-1} \quad (15),$$

$$\Phi_3(r) = e^{-\zeta\kappa r} \sum_{k=0}^{\mu} \frac{\mu!}{(\mu-k)!\zeta^{k+1}}(\kappa r)^{\mu-k+l} \quad (16).$$

The values ν and μ are related to the radial exponent n from the Slater function: $\nu = n + l + 2$ and $\mu = n - l + 1$. This potential is valid for spherical electron density by setting l=0 in the equations above, as well as for the radial part of multipolar terms.

The third term of the energy sum (E_{nm}, nucleus-multipole) for one Slater function is in essence similar to the second, with a small addition:

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

Φ_B remains the radial electrostatic potential, this time coming from the radial part of the multipolar density distribution. The new term $d_{lm}(\theta_A, \varphi_A)$ is the value of a density-normalized (Michael & Volkov, 2015) spherical harmonic. The polar angle θ_A and azimuthal angle φ_A are the spherical coordinates of nucleus A as seen from nucleus B.

Both E_{ns} and E_{nm} are asymmetrical terms. A full interaction energy calculation requires calculating this term twice, with indices A and B exchanged.

The electron-electron interaction terms E_{ss}, E_{sm} and E_{mm} all come from the potential described for a single Slater function (eqs. 12-16), hence they look similar to it:

$$E_{\text{electron-electron}} = k_0\,(I_1 + I_2 + I_3) \quad (17),$$

Complete expressions necessary to calculate E_{mm} follow. The other two terms come from these same formulas: one must set $l_A = l_B = m = 0$ to obtain E_{ss}, and either $l_A = m = 0$ or $l_B = m = 0$ to get E_{sm}. As we discuss below, these integrals are only legitimate when $m_A = m_B = m$.

$$E_{mm} = k_0 \, (I_1 + I_2 + I_3) \qquad (18),$$

$$k_0 = \frac{2\pi}{d^{l_B+l_A+1}\zeta_B^{n_B+l_B+2l_A+2}} \sqrt{\frac{2l_B+1}{2l_A+1}\frac{(l_B+m)!}{(l_B-m)!}\frac{(l_A-m)!}{(l_A+m)!}} \qquad (19),$$

$$I_1(r) = \sum_{i=0}^{i'}\sum_{j=0}^{j'} C_{l_A}^{n_B' l_B m_B}(i,j)\,(\kappa_B\zeta_B)^{i+j}\, d^i \frac{\nu!}{(\kappa_A\zeta_A)^{\nu+1}} \Big\{(-1)^j e^{-\kappa_B\zeta_B d}\int_0^d r^{j-2l_A}\, e^{\kappa_B\zeta_B r} dr$$

$$+(-1)^i e^{\kappa_B\zeta_B d}\int_d^\infty r^{j-2l_A}\, e^{-\kappa_B\zeta_B r} dr - e^{-\kappa_B\zeta_B d}\int_0^\infty r^{j-2l_A}\, e^{-\kappa_B\zeta_B r} dr\Big\}, \qquad (20),$$

$$I_2(r) = -\sum_{i=0}^{i'}\sum_{j=0}^{j'}\sum_{k=0}^{\nu} C_{l_A}^{n_B' l_B m_B}(i,j)\,(\kappa_B\zeta_B)^{i+j}\, d^i \frac{\nu!}{(\nu-k)!}\frac{1}{(\kappa_A\zeta_A)^{k+1}}$$

$$\times\Big\{(-1)^j e^{-\kappa_B\zeta_B d}\int_0^d r^{j-2l_A+\nu-k} e^{-(\kappa_A\zeta_A-\kappa_B\zeta_B)r} dr + (-1)^i e^{\kappa_B\zeta_B d}\int_d^\infty r^{j-2l_A+\nu-k} e^{-(\kappa_A\zeta_A+\kappa_B\zeta_B)r} dr$$

$$-e^{-\kappa_B\zeta_B d}\int_0^\infty r^{j-2l_A+\nu-k} e^{-(\kappa_A\zeta_A+\kappa_B\zeta_B)r} dr\Big\}, \qquad (21),$$

$$I_3(r) = \sum_{i=0}^{i'}\sum_{j=0}^{j'}\sum_{k=0}^{\mu} C_{l_A}^{n_B' l_B m_B}(i,j)\,(\kappa_B\zeta_B)^{i+j}\, d^i \frac{\mu!}{(\mu-k)!}\frac{1}{(\kappa_A\zeta_A)^{k+1}}$$

$$\times\Big\{(-1)^j e^{-\kappa_B\zeta_B d}\int_0^d r^{j+\mu-k+1} e^{-(\kappa_A\zeta_A-\kappa_B\zeta_B)r} dr + (-1)^i e^{\kappa_B\zeta_B d}\int_d^\infty r^{j+\mu-k+1} e^{-(\kappa_A\zeta_A+\kappa_B\zeta_B)r} dr$$

$$-e^{-\kappa_B\zeta_B d}\int_0^\infty r^{j+\mu-k+1} e^{-(\kappa_A\zeta_A+\kappa_B\zeta_B)r} dr\Big\} \qquad (22).$$

The marks introduced before (d, κ, ζ, l, m, μ, ν) retain their meanings, and the indices A or B precise to which atom do they belong. The summation limits are $i' = n_B + l_B - |m| + l_A + 1$ and $j' = n_B + l_A + 1$. Furthermore, for notational convenience, $n_B' = n_B + 1$. We calculate these integrals by following the formulas 38-44 given in the Nguyen, Kisiel and Volkov paper (Nguyen *et al.*, 2018).

The coefficient $C_{l_A}^{n_B' l_B m_B}(i,j)$ in the previous equation comes from the following expression:

$$\sum_{i=0}\sum_{j=0} C_{l'}^{nlm}(i,j) z^i r^j = \sum_{p=0}^{\frac{l+m}{2}}\sum_{q=0}^{l+m-2p}\sum_{v=0}^{l+m-2p-q}\sum_{p'=0}^{\frac{l'+m}{2}}\sum_{q'=0}^{l'-m-2p'}\sum_{v'=0}^{l'-m-2p'-q'}\sum_{k=0}^{t}\sum_{k'=0}^{t-k} z^x r^y$$

$$\times\frac{(-1)^{l+p+v+p'+q'}}{4^{l+l'+p-p'}(l-p)!\,p!\,p'!\,q!\,q'!\,v!\,v'!}$$

$$\times\frac{(2l-2p)!(2l'-2p')!t!}{(l+m-2p-q-v)!(l'-p')!(l'-m-2p'-q'-v')!k'!(t-k-k')!} \qquad (23).$$

with $x = n + l + 2l' - 2v - 2p' - 2v' - k - k'$, $y = 2v + 2p' + 2v' + k'$ and $t = n - l + 2p + 2q + 2q'$. We created a code in Mathematica that generates these polynomials and extracted the coefficients $C_{l'}^{nlm}(i,j)$ using Python. We pre-calculated these matrices up to $n = 12$ and stored them in library files. They are available as a supplement file to this paper and with new MoProSuite distributions. When using typical MoPro parameters, matrices up to $n = 12$ permit calculations at least up to the hexadecapolar level and with any atom at least up to iodine.

A distinctive lack of a spherical harmonic d_{lm} multiplier in the integral formulas comes from their integration rules. An integral over a product of two orthonormal spherical harmonic functions equals to either zero or one, depending on the functions involved. We first introduce an important concept to help us orthonormalize these integrals.

We define an *inter-atomic* local coordinate system as any coordinate system centred on nucleus A for which the z-axis points from atom A to atom B (Jones, 1993). One can choose the remaining two axes at will, provided they form an orthonormal basis. The E_{sm} and E_{mm} terms are only legitimate when multipolar populations of both atoms are expressed in the inter-atomic system. The computation also becomes more efficient when expressed in these terms.

The integrals in E_{mm} (eqs. 19-22) vanish only for $m_{\mathrm{A}} \neq m_{\mathrm{B}}$ in this case because spherical harmonics are orthogonal. They do not vanish, however, when $l_{\mathrm{A}} \neq l_{\mathrm{B}}$ as long as $m_{\mathrm{A}} = m_{\mathrm{B}}$. As a consequence, multipole-multipole cross-terms – for instance the d_{10} (d_z) dipole and the q_{20} (q_{z^2}) quadrupole – can differ from zero, and so can any spherical-multipole cross-terms for which $l = 0$. The individual energies will differ based on how one selects the x- and y-axes of the interatomic axis system, but their sum across one multipolar level (same spherical harmonic order l) will remain the same no matter the choice.

The procedure outlined thus far describes treatment of individual Slater functions. A usual description of an atomic electron cloud found in literature consists of groupings called atomic orbitals, which are sums of Slater-type wave functions (Clementi & Roetti, 1974; Koga *et al*., 1993; Koga *et al*., 2000). When these sums are normalized, they describe the behaviour of an orbital containing exactly one electron. Consequently, one must scale energies calculated from individual Slater functions to match their actual number of electrons.

An atomic orbital contains N_{orb} electrons – its neutral population. Whenever an orbital enters into the calculation (either for potential or for energy integral), the correct orbital energy equals the one-electron orbital energy multiplied by N_{orb} (for core and valence orbitals alike).

The Hansen-Coppens model contains a parameter called P_{val} (eq. 1), the 'real' population of valence orbitals, as opposed to N_{val}, the neutral valence population (*i.e.* the sum of N_{orb} for all orbitals designated as valence). 'Real' electrons spread equitably between all valence orbitals in the model, so the energies must be further scaled by the quotient $P_{\mathrm{val}}/N_{\mathrm{val}}$.

The Hansen-Coppens model further contains a panoply of multipolar populations P_{lm} (eq. 1). They function the same way as N_{orb} factors for orbitals: one has to multiply the one-electron energy with them. Furthermore, because integrals vanish whenever $m_{\mathrm{A}} \neq m_{\mathrm{B}}$, one has to integrate over all pairs of Slater functions

in spherical orbitals (which all have m=0), only spherical-multipolar pairs for z-symmetric multipoles (for which m = 0), and all multipolar-multipolar terms for which $m_A = m_B$. The term E_{mm} consequently contains both 'normal' interactions like dipole-dipole, quadrupole-quadrupole and so on, and 'cross' interactions like dipole-quadrupole, dipole-octupole, quadrupole-octupole and so on, as long as $m_A = m_B$.

We were interested in saving time by simplifying the calculation at long distances. The Buckingham approximation is valid when atoms are sufficiently far away, but its implementation is time-consuming. We have therefore opted for an easier approach of re-using existing code. The total energy in the pseudo-Buckingham approximation comes in the form of a sum of three terms (reproduced from the main text):

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

These terms mirror those from the expression (3). The index 'c' denotes the point charge part: at long distance, a spherical charge density has the same effect as a point charge at its centre, and the index 'm' denotes the multipole part.

The charge-charge term E_{cc} substitutes $E_{nn} + E_{ns} + E_{ss}$ terms. It is calculated like the E_{nn} term:

$$E_{cc} = q_A \, q_B / d \qquad (8).$$

The charges q for the respective atoms are defined as $P_{val} - N_{val} + P_{00}$. The point-multipolar term E_{cm} replaces $E_{nm} + E_{sm}$ terms. It is calculated similarly to the E_{nm} term, after substituting z with q:

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

Like the E_{nm} term, it is also asymmetrical. The final E_{mm} term remains the same as in the previous calculation. We calculate it using simply the NKV analytical integration (aEP) method.

S1.2. Details on *Charger* implementation and MoProViewer integration

We used the method outlined thus far in the construction of a novel code library called *Charger*. It is fully written in C and it allows the calculation of electrostatic energy between two atoms and electrostatic potential generated by a multipolar atom. It conveniently defines an *Atom* struct, which facilitates integration with existing charge density analysis software. Integration entails filling the *Atom* struct in and calling the relevant *Charger* functions.

Nucleus-electron interaction energy functions ($E_{ns} = Z \, \Phi_s$ and $E_{nm} = Z \, \Phi_m \, d_{lm}$) to the total electrostatic interaction energy evaluate quickly. *Charger* contains separate functions for: spherical (Φ_e) and multipolar (Φ_m) electrostatic potential, spherical-spherical interaction energy (E_{ss}), spherical-multipolar interaction energy (E_{sm}) and multipolar-multipolar interaction energy (E_{mm}). This deliberate separation helps to optimize the spherical-spherical energy calculation – the one that takes up the lion's share of computation time.

Charger treats the terms to which P_{00} monopole contributes in the same way as other terms with spherical contributions. However, it attributes P_{00} energy contributions to multipolar E_{sm} and E_{mm} energies. There is a switch to skip these calculations at long distance: P_{00} then takes part in the point charge q.

The Hansen-Coppens model allows for multiple values of expansion/contraction coefficient for different multipolar orders: $\kappa' = \kappa'(l)$, but they are rarely encountered in practice. The calculation speed of the multipole-multipole E_{mm} term improves by assuming that κ' is not a function of l, thus calculating the relevant integrals only once. This gives an overall performance improvement of about 30%. *Charger* automatically switches to a slower version only when it encounters $\kappa' = \kappa'(l)$.

The spherical-multipolar and multipolar-multipolar functions assume how many integrals they have to calculate from the number of P_{lm} parameters. When atoms are limited to a certain multipole level (*e.g.* up to quadrupoles), *Charger* should not get any coefficients beyond that level (*e.g.* $P_{3m} = P_{4m} = ... = 0$) to avoid wasting time calculating useless integrals. *Charger*'s interface within the molecular visualisation software MoProViewer ensures that this does not happen.

The speed-up at long distance comes largely from not calculating the many integrals involved in the spherical-spherical E_{ss} term, but also from substituting the integration within spherical-multipolar E_{sm} term with a faster potential evaluation to obtain E_{cm}. Fully substituting the multipolar-multipolar E_{mm} term with a real Buckingham approximation would further improve the calculation speed. *Charger* library users determine when to switch between the short-distance full integration and the pseudo-Buckingham approximation.

We have further improved the calculation at short distances by parallelizing the computation of the spherical-spherical E_{ss} term. Calculations with OpenMP achieve a modest 10–15% speed-up on 4 cores.

Another speed and accuracy improvement comes from excluding high-order multipolar terms at long distances. The user defines a distance cut-off after which the interaction energy E_{mm} for multipoles with a certain $l_A + l_B$ is no longer calculated. *Charger* supports different cut-off distance values for $l_A + l_B = 2$ (dipole-dipole interactions), 3 (dipole-quadrupole interactions), 4 (dipole-octupole and quadrupole-quadrupole interactions) … up to $l_A + l_B = 12$. This is the current limit of our modeling software: the largest multipolar levels MoPro uses are 64-poles ($l = 6$). Imposing the same distance limit for $l_A + l_B > 12$ as for $l_A + l_B = 12$ ensures future compatibility. We recommend to avoid calling the E_{mm} function if the distance surpasses the dipole-dipole cut-off distance (or some other 'absolute' cut-off distance), as this seriously improves calculation speed.

Charger has the ability to complement natively normal charges and multipolar populations with induced charges and multipoles. One can easily switch the induced parts on and off. Similarly, one can include or exclude any atom constituents (the nucleus, core electrons, neutral valence electrons, all valence electrons and any level of multipoles) for either group of atoms involved in interaction. The MoProViewer interface supports this functionality.

S2. Electric field lines topography analysis on Arg171Gln mutant

We wish to give a rationale why the Arg171Gln mutation improves the binding of the 2,4,4'-trihydroxybenzophenone ligand from an electrostatic interaction energy perspective. The ligand being neutral, the loss of the arginine positive charge should not affect the binding by a large margin. Moreover, dipoles moments of the ligand and of the arginine 171 are nearly perpendicular.

Despite this, permanent electrostatic interaction between 2,4,4'-trihydroxybenzophenone and Arg171 is unfavourable. The polarisation energy compensates that penalty without drastic changes of molecular dipole moments (+4.5 kcal/mol without polarisation, +0.8 kcal/mol with polarisation). In the Arg171Gln variant, the electrostatic interaction energy between Gln171 favours the binding and adding polarisation terms further strengthens this interaction.

We plotted electric field lines in the active sites of 6f68 and of the Arg171Gln variant to pinpoint main electrostatic influences of different residues on the ligand. Ligand polarisation comes from this electrostatic potential, so this method gives a qualitative information about which residues mainly influence the bound ligand. The electrostatic zones of influence linked to residue 171 are relatively conserved when comparing A and B chains. These structures persist in both the 6f66 and 6f68 complexes. They contain two different benzophenone ligands interacting with the same wild-type protein structure, representing two distinct binding modes.

First, the field lines run in opposite directions on the two panels of Figure S6. A further point lies in changes of influence zones caused by the mutation. In both panels, the dark red bundle always links the position 171 to Phe168. In the wild type (Figure S6.A), the dark red bundle starts from the negatively charged backbone oxygen (Phe168) and converges to C_δ atom of Arg171. This bundle only covers the hydroxyl group (O1) and two carbon atoms from the ligand aromatic ring (C1 and C6): it is specifically involved in the strong hydrogen bond between Phe168 and 2,4,4'-trihydroxybenzophenone.

In the Arg171Gln mutant (Figure S6.B), this dark red bundle extends much further – covering three carbon atoms of the ligand aromatic ring. It starts at the δ position of Gln171 and ends on the phenyl hydrogen atoms of Phe168. The field lines of the dark red bundle are close to perpendicular to the aromatic plane. In the mutated structure, this bundle is no longer involved in the hydrogen bond with Phe168 backbone oxygen atom. The main electrostatic interactors of the HBP 4'-hydroxyl group are now Phe168 and Phe123 (not shown on S4.B).

The green bundle starts from the π-electron cloud of Trp127 (wild-type, Figure S6.B). It covers most of the ligand aromatic ring, from the atoms covered by the dark red bundle to the keto group that bridges the phenol rings. Arg171 is no more the main electrostatic attractor beyond the keto group in the wild type. The green field lines explain the direction of induced dipole moments in a large portion of the ligand, and align reasonably with the ring plane – therefore with the polarizabilities of the aromatic carbon atoms (Figure S6.A). This may explain a reasonably favourable polarisation energy to compensate the unfavourable electrostatic interaction energy between Arg171 and the ligand.

In the Arg171Gln mutant, Trp127 is not involved in field line bundles with Gln171. The dark blue bundle and the green bundle in the Arg171Gln variant now cover most of the atoms covered by the green bundle in the wild type. The green field lines are close to perpendicular to the ligand ring planes (Figure S6.B) and converge to Arg124 (ligand positions C3 and C2 are covered). This bundle links together stronger charge concentrations in the mutant compared to the wild type. The dark blue bundle links the Gln171 O_{δ} atom with the Phe128 phenyl hydrogen atoms. This bundle covers half of the closer ligand aromatic ring, the keto group and the branching carbon of the other aromatic ring. Here the field lines align a little more with the aromatic plane compared to the Arg171Gln green bundle, underlining favourable polarisation contributions.

The residues Phe14, Pro16 and Tyr175 are in close proximity to the mutated residue 171. Bundles of field lines link these residues to the residue 171 in the wild type or in the mutated protein, but the considered bundles do no cover ligand atomic positions (and are omitted from Figure S6).

In the wild type variant, the alignment of field lines with the nearest aromatic plane suggest the cause of favourable polarisation energy. The influence zone of the mutant Gln171 is much larger than the influence zone of Arg171. The field lines in the mutant also align quite well with the aromatic planes, and cover more atomic positions, suggesting an even more favourable polarisation interaction. The only exception is the bundle between Phe168 and Gln171. Steeper electric fields between larger charge concentrations are also a beneficial influence to the polarisation energy in the mutated complex. These findings are coherent with interpreting the effect of this mutation as an electrostatic obstacle removal.

Table S1 Electrostatic interaction energies (in kcal·mol^{-1}) for interactions in benchmark systems used by Nguyen *et al*. (2018).

Dimer designation	Mononmer A	Monomer B	Charger		VMoPro	NKV paper**	δ_{NKV} / % $[(E_{Charger\ p.} - E_{NKV}) / E_{NKV}]$
			precise calculation*	rapid calculation*			
Gly1	Glycine	Glycine	–27.9	–27.9	–27.8	–27.7	0.7
Gly2	Glycine	Glycine	–7.2	–7.2	–7.3	–7.0	2.9
Gly3	Glycine	Glycine	–19.8	–19.8	–19.8	–20.9	–5.3
Gly4	Glycine	Glycine	–35.7	–35.7	–35.9	–39.1	–8.7
Gly5	Glycine	Glycine	10.1	10.1	10.1	11.1	–9.0
Gly6	Glycine	Glycine	–5.5	–5.5	–5.6	–5.6	–1.8
Lenk1	Leu-enkephalin	L-enk	–64.4†	–64.4†	–64.5†	–74.8†	–†
Lenk2	L-enk	L-enk				–37.5†	
Lenk3	L-enk	L-enk	4.5	4.5	4.5	–3.8	–218
Lenk4	L-enk	Water	–13.3	–13.3	–13.3	–15.3	–13.1
Lenk5	L-enk	Water	–14.4	–14.4	–14.4	–14.9	–3.4
Lenk6	L-enk	Water	–14.7	–14.7	–14.8	–10.7	37.4
Lenk7	L-enk	Water	–14.4	–14.4	–14.4	–12.9	11.6
Lenk8	L-enk	Water	–8.6	–8.6	–8.6	–8.5	1.9
Lenk9	L-enk	Water	–9.1	–9.1	–9.1	–5.7	59.7
Lenk10	L-enk	Water	–5.4	–5.4	–5.5	–7.1	–24.0
Lenk11	Water	Water	–10.6	–10.6	–10.5	–10.3	2.9
Non3††	Nonap.†††	DMA††	1.6	1.6	1.6	–9.6	–116
Dec1	Decap.†††	Decap.††	–36.2	–36.2	–36.4	–47.4	–23.6
Dec2	Decap.†††	Methanol	–15.6	–15.6	–15.6	–14.0	11.4
Dec3	Decap.†††	Methanol	–8.8	–8.8	–8.8	–7.0	25.7
Dec4	Decap.†††	Water	–10.4	–10.4	–10.6	–10.7	–2.8
Dec5	Decap.†††	Water	–2.3	–2.3	–2.3	–3.0	–23.3
Dod1	Dodecap.†††	Dodecap.††	–43.5	–43.1	–43.5	–48.4	–10.1
Dod2	Dodecap.†††	2-butanone	–8.8	–8.8	–8.9	–3.5	151.4
Dod3	Dodecap.†††	2-butanone	–1.7	–1.8	–1.8	–3.1	–45.2

* Precise and rapid cut-off criteria are reported in Table 1.

[**] Refers to values as given in (Nguyen *et al.*, 2018) and converted from kJ·mol^{-1} to kcal·mol^{-1}.The values are not the same because the models are not completely equal, specifically the multipolar parameters.

[***] Electrostatic interaction energy calculation by *Charger* based on an AMBER point-charge model. Excluded results had problems with model transfer

[†] The file Lenk_dimer_1.xyz in the Supplementary Materials of (Nguyen *et al.*, 2018) does not contain a Leu-enkephalin dimer. Instead, it contains the same glycine dimer as the file Gly_dimer_6.xyz. We were therefore only able to obtain results for Lenk2, which are listed in the table. We do not give the relative differences, as we are not sure whether the file Lenk_dimer_2.xyz contains the dimer listed as Lenk1 or as Lenk2.

[††] Nonapeptide – dimethylacetamide dimer. The files for the other two dimers (Non1 and Non2) contained three molecules, so they were not included.

[†††] Nonapeptide: Boc-(l-Leu-l-Leu-Aib)2-l-Leu-d-Leu-Aib-OMe; decapeptide: Boc-(l-Leu-Aib)5-OMe; dodecapeptide: Boc-l-Leu-l-Leu-Aib-(d-Leu-d-Leu-Aib)2-l-Leu-l-Leu-Aib-OMe; Boc = *tert*-butoxycarbonyl – (CH3)3**C**O**OC**–, Leu = leucine, Aib = 2-methylalanine. There are no parameters in the ELMAM2 database for the quaternary carbon and the carbonyl group in Boc (in bold), nor for the quaternary carbon C_α of 2-methylalanine.

We did not calculate a polarised electron density model for these calculations.

Table S2 Electrostatic interaction energies (in kcal·mol^{-1}) for interactions in benchmark systems used by Nguyen *et al*. (2018).

Comparison with the AMBER point-charge model.

Dimer designation	Mononmer A	Monomer B	Charger*	AMBER**	$\delta^{***}_{\mathrm{AMBER}}$ / %	AMBER P_{val}**	$\delta^{***}_{\mathrm{Amber}\ P\mathrm{val}}$ / %
Gly1	Glycine	Glycine	–27.9	–12.0	133	–23.2	20
Gly2	Glycine	Glycine	–7.2	–0.4	1700	–8.9	–19
Gly3	Glycine	Glycine	–19.8	–11.7	69	–16.7	19
Gly4	Glycine	Glycine	–35.7	–18.5	93	–30.7	16
Gly5	Glycine	Glycine	10.1	6.7	51	3.8	166
Gly6	Glycine	Glycine	–5.5	–2.3	139	–4.5	22
Lenk2†	Leu-enkephalin	L-enk	–64.4	–39.0	65	–51.9	24
Lenk3	L-enk	L-enk	4.5	2.7	67	–2.7	–267
Lenk4	L-enk	Water††	–13.3	–8.8	51	–18.6	–28
Lenk5	L-enk	Water††	–14.4	–10.4	38	–16.9	–15
Lenk6	L-enk	Water††	–14.7	–7.1	107	–13.1	12
Lenk7	L-enk	Water††	–14.4	–8.5	69	–13.7	5
Lenk8	L-enk	Water††	–8.6	–3.6	139	–8.6	0
Lenk9	L-enk	Water††	–9.1	–3.9	133	–11.6	–22
Lenk10	L-enk	Water††	–5.4	–5.9	–8	–9.1	–41
Lenk11	Water††	Water††	–10.6	–7.5	41	–12.6	–16

* Precise calculation, as reported in table S1.

** Electrostatic interaction energy calculation by *Charger* based on an AMBER point-charge model. The AMBER column contains results for models where AMBER point charges are interpreted as point charges. The AMBER P_{val} column contains results for models in which AMBER point charges are converted to P_{val} values from the Hansen-Coppens model, and full spherical integration is performed for all atoms closer than 5 Å apart.

*** $\delta_{\mathrm{AMBER}} = (E_{\mathrm{Charger}} - E_{\mathrm{AMBER}}) / E_{\mathrm{AMBER}}$, $\delta_{\mathrm{AMBER}\ P\mathrm{val}} = (E_{\mathrm{Charger}} - E_{\mathrm{AMBER}\ P\mathrm{val}}) / E_{\mathrm{AMBER}\ P\mathrm{val}}$.

† See note † from table S1 to explain why Lenk1 is missing.

†† Point charges of a TIP3P water molecule used.

Dimers excluded from this table (compared to Table S1) have parameters missing.

Table S3 Calculation times (in seconds) for electrostatic interaction energies in benchmark systems.

Dimer designation	Monomer A	Monomer B	Charger		VMoPro	NKV paper[**]
			precise calculation[*]	rapid calculation[*]		
Gly1	Glycine	Glycine	0.085	0.080	15	0.3
Gly2	Glycine	Glycine	0.094	0.087	16	0.4
Gly3	Glycine	Glycine	0.13	0.11	22	0.6
Gly4	Glycine	Glycine	0.17	0.16	30	0.8
Gly5	Glycine	Glycine	0.13	0.13	23	0.7
Gly6	Glycine	Glycine	0.087	0.086	16	0.4
Lenk1	Leu-enkephalin	L-enk	_[***]	_[***]	_[***]	3.3
Lenk2	L-enk	L-enk	0.60	0.56	51	1.6
Lenk3	L-enk	L-enk	0.63	0.57	58	1.7
Lenk4	L-enk	Water	0.077	0.070	8.2	0.3
Lenk5	L-enk	Water	0.058	0.056	4.4	0.2
Lenk6	L-enk	Water	0.078	0.076	14	0.3
Lenk7	L-enk	Water	0.069	0.068	7.5	0.3
Lenk8	L-enk	Water	0.068	0.067	6.2	0.3
Lenk9	L-enk	Water	0.014	0.013	21	0.6
Lenk10	L-enk	Water	0.089	0.087	13	0.4
Lenk11	Water	Water	0.017	0.016	2.7	0.1
Non3[***]	Nonap.[***††]	DMA[***]	0.23	0.21	24	1.2
Dec1	Decap.[††]	Decap.[††]	2.1	1.4	119	3.4
Dec2	Decap.[††]	Methanol	0.15	0.15	19	0.5
Dec3	Decap.[††]	Methanol	0.22	0.22	36	0.8
Dec4	Decap.[††]	Water	0.16	0.16	29	0.7
Dec5	Decap.[††]	Water	0.091	0.090	15	0.3
Dod1	Dodecap.[††]	Dodecap.[††]	2.8	1.5	125	3.6
Dod2	Dodecap.[††]	2-butanone	0.39	0.35	55	1.0
Dod3	Dodecap.[††]	2-butanone	0.38	0.35	52	1.2

For our calculations (Charger & VMoPro) we used Intel® Core™ i7-8700 @ 3.20 GHz, Passmark (processor benchmark): 13.106.

[*] Precise and rapid cut-off criteria are reported in Table 1.

[**] Refers to values as given in (Nguyen *et al.*, 2018) for an aEP/MM calculation. They used AMD Opteron 6348 @ 2.80 GHz, Passmark: 9.151.

[***] Nonapeptide – dimethylacetamide dimer. The files for the other two dimers (Non1 and Non2) contained three molecules, so they were not included.

[†] The file Lenk_dimer_1 in the Supplementary Materials of (Nguyen *et al.*, 2018) contain a glycine dimer instead of a Leu-enkephalin dimer. The energy value reproduced here serves mainly for comparison with Lenk2.

[††] Nonapeptide: Boc-(l-Leu-l-Leu-Aib)2-l-Leu-d-Leu-Aib-OMe; decapeptide: Boc-(l-Leu-Aib)5-OMe; dodecapeptide: Boc-l-Leu-l-Leu-Aib-(d-Leu-d-Leu-Aib)2-l-Leu-l-Leu-Aib-OMe; Boc = *tert*-butoxycarbonyl – (CH3)3**C**O**OC**–, Leu = leucine, Aib = 2-methylalanine. There are no parameters in the ELMAM2 database for the quaternary carbon and the carbonyl group in Boc (in bold), nor for the quaternary carbon C_α of 2-methylalanine.

Table S4 Electrostatic interaction energies (in kcal·mol^{-1}, electrostatic without polarisation) for interactions of glutathione transferase with its benzophenone and glutathione ligands (Schwartz *et al*., 2018).

			Monomer A			Monomer B			Dimer*		
			Charger		VMo-Pro	Charger		VMo-Pro	Charger		VMo-Pro
Ligand	PDB code	Active site	precise**	rapid**		precise**	rapid**		precise**	rapid**	
2,4,4'-trihydroxy-benzophenone	6f68	A	–20.3	–20.4	–20.2	–0.733	–0.718	–0.703	–52.2	–52.3	–52.3
		B	–0.353	–0.335	–0.334	–30.8	–30.9	–31.0			
3,4-dihydroxy-benzophenone	6f67	A	–25.5	–25.7	–26.6	–1.77	–1.74	–1.75	–55.7	–56.2	–55.8
		B	–2.37	–2.34	–2.35	–25.9	–26.2	–25.9			
2,4-dihydroxy-benzophenone	6f66	A	–37.6	–37.7	–37.4	0.432	0.471	0.444	–83.2	–83.2	–82.6
		B	0.0973	0.148	0.0998	–46.1	–46.2	–45.8			
2,3,4-trihydroxy-benzophenone	6f69	A	–35.0	–35.3	–34.8	1.06	1.08	1.08	–82.4	–82.6	–82.0
		B	0.953	0.954	0.971	–49.4	–49.4	–49.3			
glutathione	6f68	A	–130.6	–130.6	–130.8	34.0	34.0	33.9	–187.9	–187.6	–188.2
		B	32.2	32.3	32.2	–96.5	–96.3	–96.5			
	6f66	A	–142.2	–142.1	–142.5	19.0	19.2	19.0	–186.6	–186.3	–187.1
		B	42.1	42.2	42.1	–105.6	–105.6	–105.7			

* Values represent interaction energies of a protein dimer with both ligands.

** precise and rapid cut-off criteria are reported in Table 1.

Table S5 Electrostatic interaction energies (in kcal·mol^{-1}, electrostatic+polarisation) for interactions of glutathione transferase with its benzophenone and glutathione ligands (Schwartz *et al*., 2018).

			Monomer A			Monomer B			Dimer*		
			Charger			Charger			Charger		
Ligand	PDB code	Active site	precise**	rapid**	VMo-Pro	precise**	rapid**	VMo-Pro	precise**	rapid**	VMo-Pro
2,4,4'-trihydroxy-benzophenone	6f68	A	–36.0	–36.2	–35.9	–0.79	–0.74	–0.70	–84.0	–84.3	–84.0
		B	–0.55	–0.52	–0.50	–46.7	–46.8	–46.8			
3,4-dihydroxy-benzophenone	6f67	A	–42.8	–43.1	–42.9	0.44	0.52	0.47	–87.2	–87.9	–87.2
		B	0.26	0.32	0.28	–45.1	–45.6	–45.0			
2,4-dihydroxy-benzophenone	6f66	A	–56.7	–57.0	–56.3	0.12	0.21	0.01	–125.4	–125.5	–124.5
		B	0.32	0.40	0.34	–69.1	–69.2	–68.7			
2,3,4-trihydroxy-benzophenone	6f69	A	–55.5	–55.7	–55.2	1.78	1.76	1.81	–127.3	–127.7	–126.9
		B	1.77	1.77	1.79	–75.4	–75.5	–75.2			

* Values represent interaction energies of a protein dimer with both ligands.

** precise and rapid cut-off criteria are reported in Table 1.

Table S6 Calculation times (in seconds) for electrostatic interaction energy of glutathione transferase with its benzophenone and glutathione ligands (Schwartz *et al*., 2018).

Corresponding energies (electrostatic without polarisation) are given in table S4. Calculation of electrostatic+polarisation energies takes about the same amount of time. The difference between polarised and non-polarised density models is only in a handful of dipole population values. These are only used as multipliers and do not affect the time-consuming integral evaluations.

			Monomer A			Monomer B			Dimer*		
			Charger			Charger		VMo-Pro	Charger		VMo-Pro
Ligand	PDB code	Active site	precise**	rapid**	VMoPro	precise**	rapid**		precise*	rapid**	
2,4,4'-trihydroxy-benzophenone	6f68	A	6.2	3.9	309	3.7	1.3	0.42	19	10	602
		B	3.8	1.3	0.44	6.1	3.9	299			
3,4-dihydroxy-benzophenone	6f67	A	6.0	3.8	286	3.7	1.2	0.42	19	9.8	565
		B	3.7	1.3	0.42	5.9	3.7	282			
2,4-dihydroxy-benzophenone	6f66	A	6.1	3.8	305	3.7	1.2	0.44	19	9.8	613
		B	3.8	1.2	0.44	6.0	3.8	298			
2,3,4-trihydroxy-benzophenone	6f69	A	6.1	3.8	294	3.9	1.3	0.42	20	10	612
		B	3.9	1.3	0.44	6.2	3.9	317			
Glutathione	6f68	A	7.6	4.5	383	5.2	1.9	24	25	13	814
		B	5.2.	1.9	26	7.6	4.5	373			
	6f66	A	7.6	4.6	379	5.3	2.0	22	25	13	806
		B	5.2	1.9	24	7.8	4.7	393			

* Values represent calculation times for a protein dimer with both ligands.

** precise and rapid cut-off criteria are reported in Table 1.

Table S7 Comparison of experimental thermal shifts with calculated interaction energies.

Glutathione transferase thermal shifts with different benzophenone ligands (Table 2), alongside with *Charger* electrostatic interaction energies (precise calculation, non-polarised | polarised).

Ligand	ΔT_d / K *	E_A / kcal mol^{-1} **	E_B / kcal mol^{-1} **	E_{mean} / kcal mol^{-1}	E_{dimer} / kcal mol^{-1} ***
2,4,4'-trihydroxy-benzophenone	2.87	–20.3 | –36.0	–30.8 | –46.7	–25.6 | –41.4	–52.2 | –84.0
3,4-dihydroxy-benzophenone	4.36	–25.7 | –42.8	–25.9 | –45.1	–25.8 | –44. 0	–55.7 | –87.2
2,4-dihydroxy-benzophenone	4.96	–37.6 | –56.7	–46.1 | –69.1	–41.8 | –62.9	–83.2 | –125.4
2,3,4-trihydroxy-benzophenone	5.69	–35.0 | –55.5	–49.4 | –75.4	–42.2 | –65.5	–82.4 | –127.3

* The relative scale uses apo-GST denaturation temperature as zero.

** Interaction energy between a GST monomer and the ligand bound in its active site.

*** Interaction energy between the GST dimer and both ligands.

In the 6f68 crystal structure, the 4'-hydroxyl-ligand has been refined with an occupancy of 0.8 and B_{mean} 7 Å^2 higher than the protein B_{mean}. Residual electron density could indicate a second conformation with a low occupancy. This second conformation was not considered in the context of this study.

Correlation coefficients (non-polarised energies):

(%)	E_A	E_B	E_{mean}	E_{dimer}
Pearson (E with ΔT_d)	–96	–69	–83	–87
Spearmann (E with rank ΔT_d)	98	78	89	92
Spearmann (rank E with rank ΔT_d)	–100	–60	–80	–100

Correlation coefficients (polarised energies):

(%)	E_A	E_B	E_{mean}	E_{dimer}
Pearson (E with ΔT_d)	–95	–75	–84	–84
Spearmann (E with rank ΔT_d)	96	81	89	90
Spearmann (rank E with rank ΔT_d)	–100	–60	–80	–80

Note: the minus sign in correlation coefficients means that the corresponding trend line has a downward slope.

Table S8 Correlation between electrostatic interaction energy contributions (non-polarised) of individual glutathione transferase residues (all from one monomer) with benzophenone ligands.

The underlined values on the diagonal represent correlations between monomers A and B of the same structure (which hold the same ligand). The darker blue values in the upper triangle represent correlations between A-monomers with different ligands, while the lighter green values in the lower triangle represent correlations between B-monomers with different ligands.

R^2 (%)	6f68	6f67	6f69	6f66
6f68	90	30	26	11
6f67	38	99.6	80	75
6f69	39	78	87	87
6f66	30	76	61	88

Ligands are present in the following structures (PDB codes): 2,4,4'-trihydroxybenzophenone in 6f68, 3,4-dihydroxybenzophenone in 6f67, 2,4-dihydroxybenzophenone in 6f69, 2,3,4-dihydroxybenzophenone in 6f69.

Table S9 Dipole moment parameters and interaction energy (from *Charger*, precise calculation, with polarisation) between the residue 171 (wild type: arginine, mutant: glutamine) and the ligand present in investigated structures.

		Dipole moments	Energy / (kcal/mol)		
Contact	4'-group	/ D*	Dipole angle / °	Monomer**	Average***
Arg171…6f68A§†	hydroxyl	18.5 | 6.6	63.0	2.77	0.76
Arg171…6f68B§		18.7 | 6.0	87.6	–1.25	
Gln171…6f68A§†		9.3 | 8.9	128.9	–2.04	–2.15
Gln171…6f68B§		9.5 | 6.9	80.4	–2.24	
Arg171…6f67A§	hydrogen	18.9 | 4.5	168.5	–9.38	–10.18
Arg171…6f67B§		18.9 | 4.9	169.5	–10.97	
Arg171…6f69A§†		18.9 | 12.3	168.1	–7.32	–5.68
Arg171…6f69B§		18.8 | 12.3	160.7	–4.03	
Arg171…6f66A§		19.1 | 10.4	147.0	–8.04	–7.86
Arg171…6f66B§		18.9 | 11.4	144.8	–7.68	

* The first number is dipole moment of the residue 171, and the second that of the ligand.

** Electrostatic interaction energy between the GST monomer (A or B) and the ligand bound within.

*** Average between the two electrostatic interaction energies in the previous column.

§ Refers to the ligand present in the structure with the given PDB codes: 2,4,4'-trihydroxybenzophenone in 6f68, 3,4-dihydroxybenzophenone in 6f67, 2,4-dihydroxybenzophenone in 6f69, 2,3,4-dihydroxybenzophenone in 6f69.

† These rows of the table correspond to the three sub-figures in figure 7.

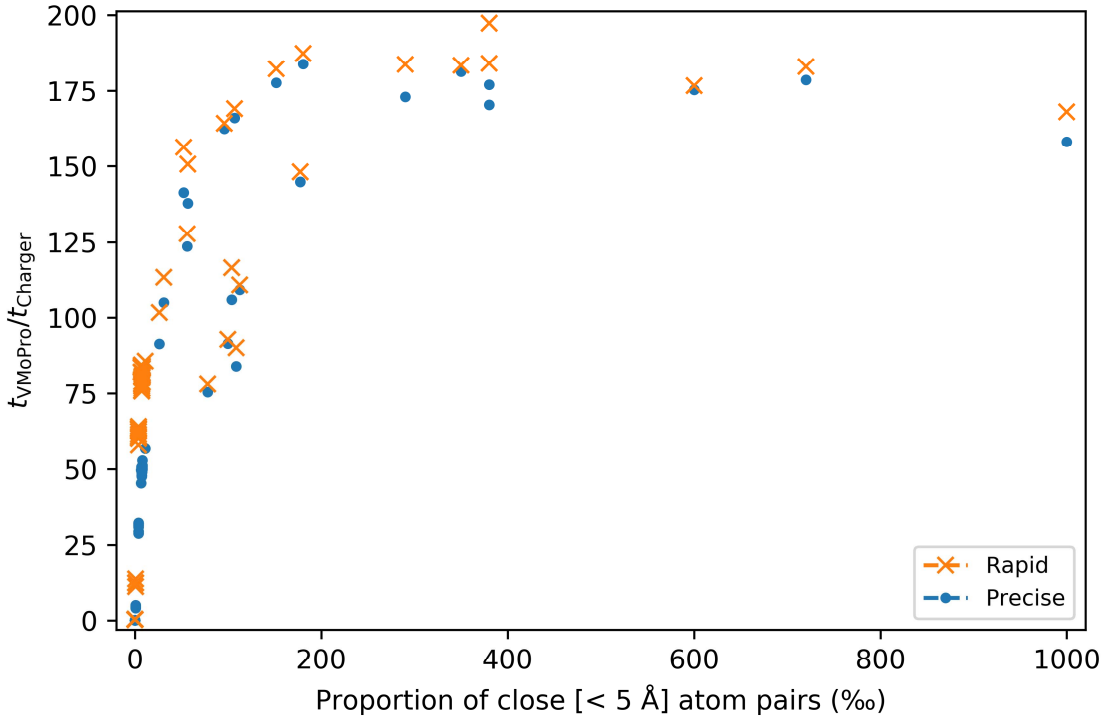

Figure S1 Computation speed enhancement. Improvement in *Charger* calculation speed as compared to VMoPro-EP/MM. The amount of close atom pairs (in promille, ‰) represents the number of atoms closer than 5 Å apart divided by the number of all atom pairs. Precise and rapid cut-off criteria are noted in Table 1.

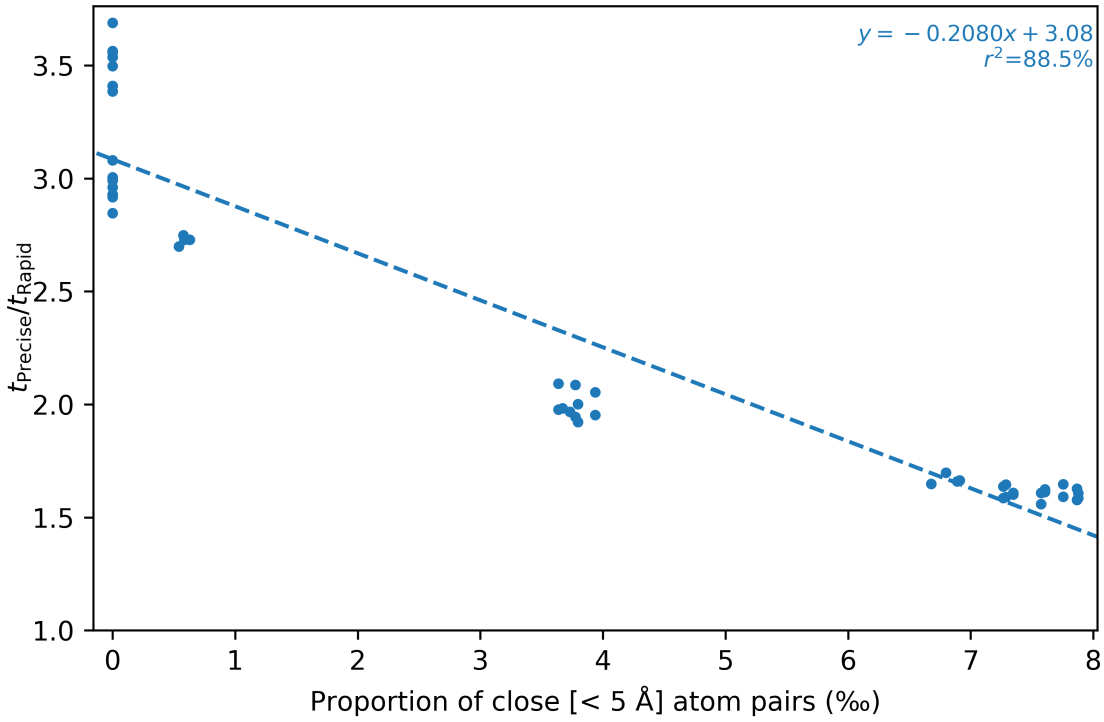

Figure S2 Comparison of different cut-off performance (for precise and rapid calculations) within *Charger*. The graph presents only data from the protein/ligand systems. Proportion of close atom pairs (in promille, ‰) represent atoms closer than 5 Å apart divided by all atom pairs. Precise and rapid cut-off criteria are reported in Table 1.

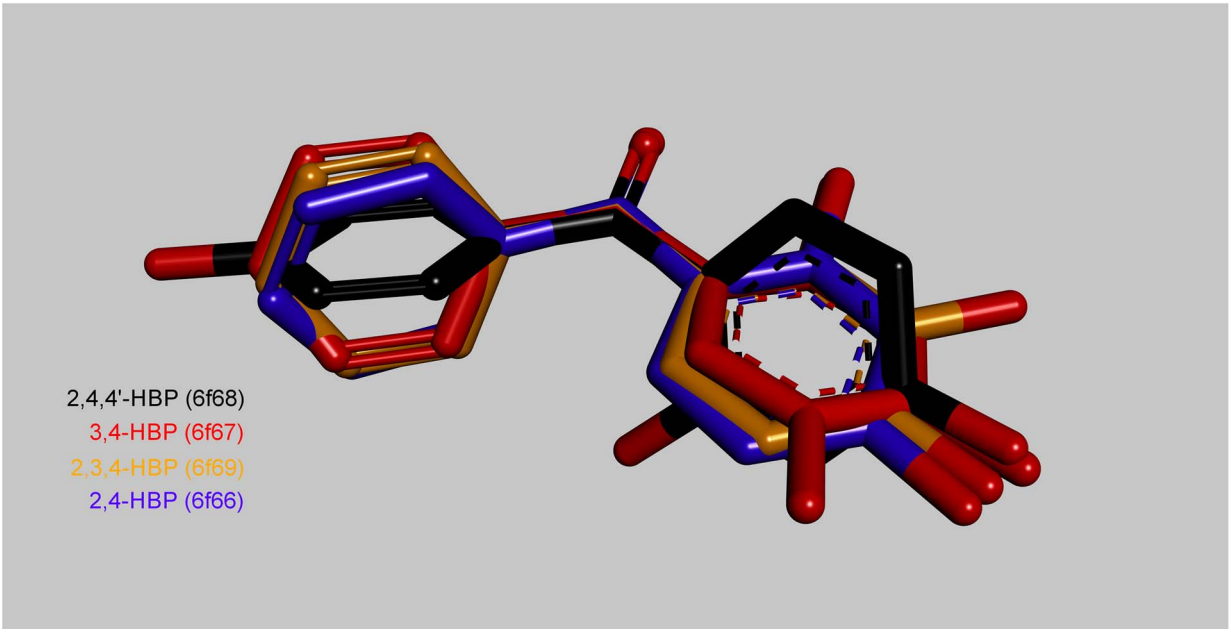

Figure S3 Overlay of HBP ligands. The superposition of the four different ligands in the GST active site. The configuration of 2,4,4'-HBP is visibly different from the other three ligands.

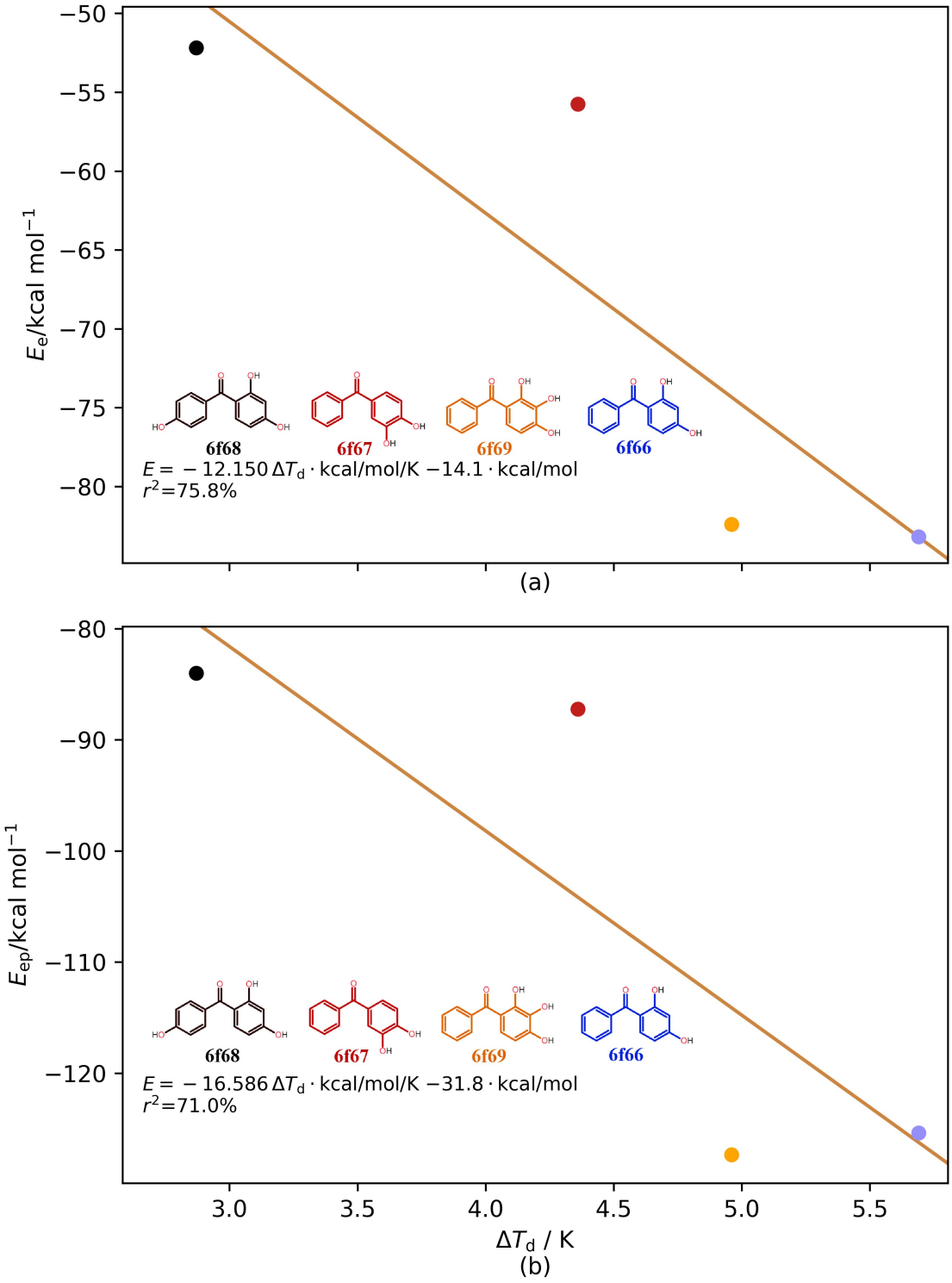

Figure S4 Relationship between experimental thermal shifts and computed energies for GST dimers. Relationship between computed electrostatic interaction energies (a: non-polarised, b: polarised) from *Charger* calculations and experimental denaturation temperatures for glutathione transferase and four ligands. Energies shown are the interactions between the whole protein (a dimer) and two ligands in their binding sites. Values given in table S7.

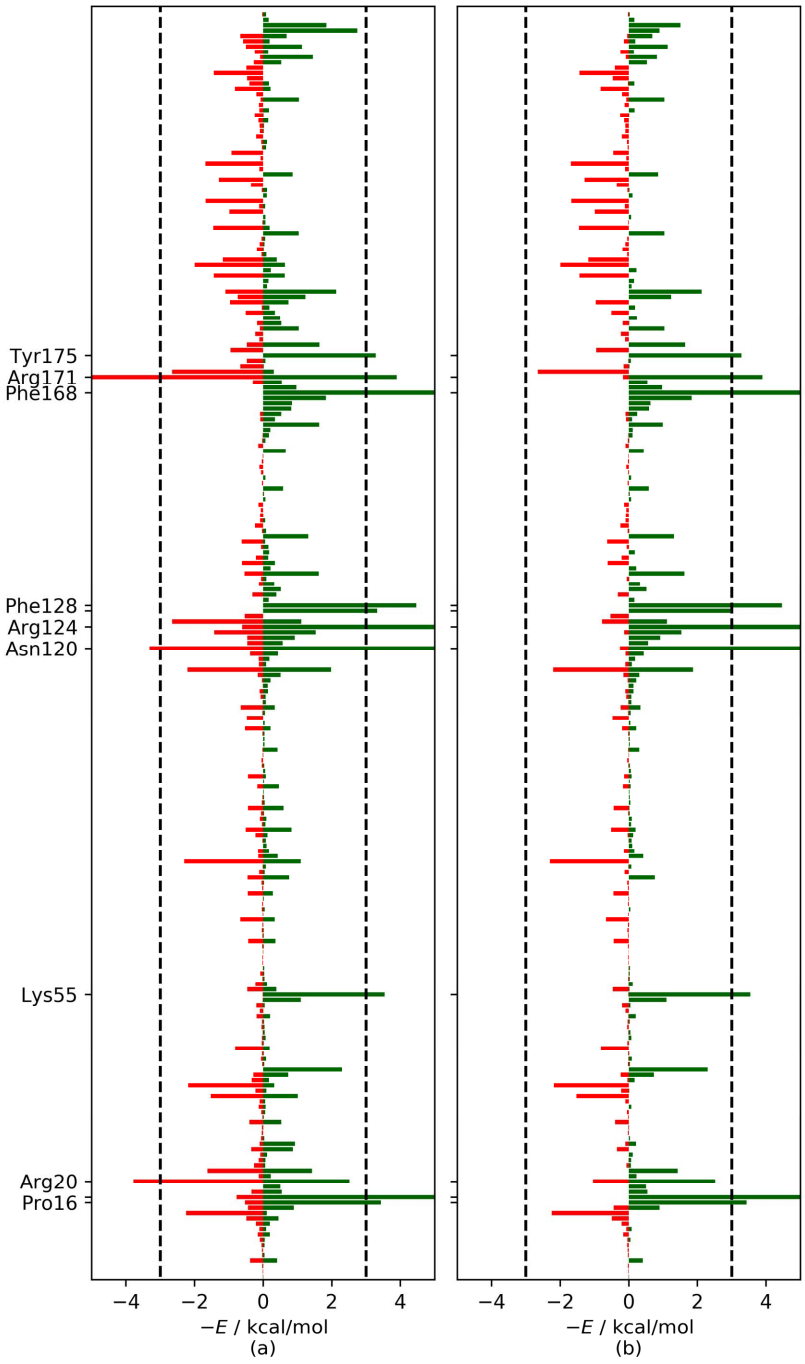

Figure S5 Largest favourable (green) and unfavourable electrostatic interaction energy contributions for individual GST residues and the benzophenone ligands. Energies have been capped at ±5 kcal/mol. Ticks on the *y*-axis represent the residues with at least one interaction energy contribution beyond ±3 kcal/mol (black lines). The omitted labels are Tyr17 and Trp127. (a) All four benzophenone ligands are considered. (b) Only the three benzophenone ligands with a similar binding mode are considered. Taken together, these graphs show that many amino acids contribute a small amount to binding from the electrostatic point of view, but their contributions roughly balance out. The energy contributions of residues outside the ±3 kcal/mol are highlighted on figure 6.

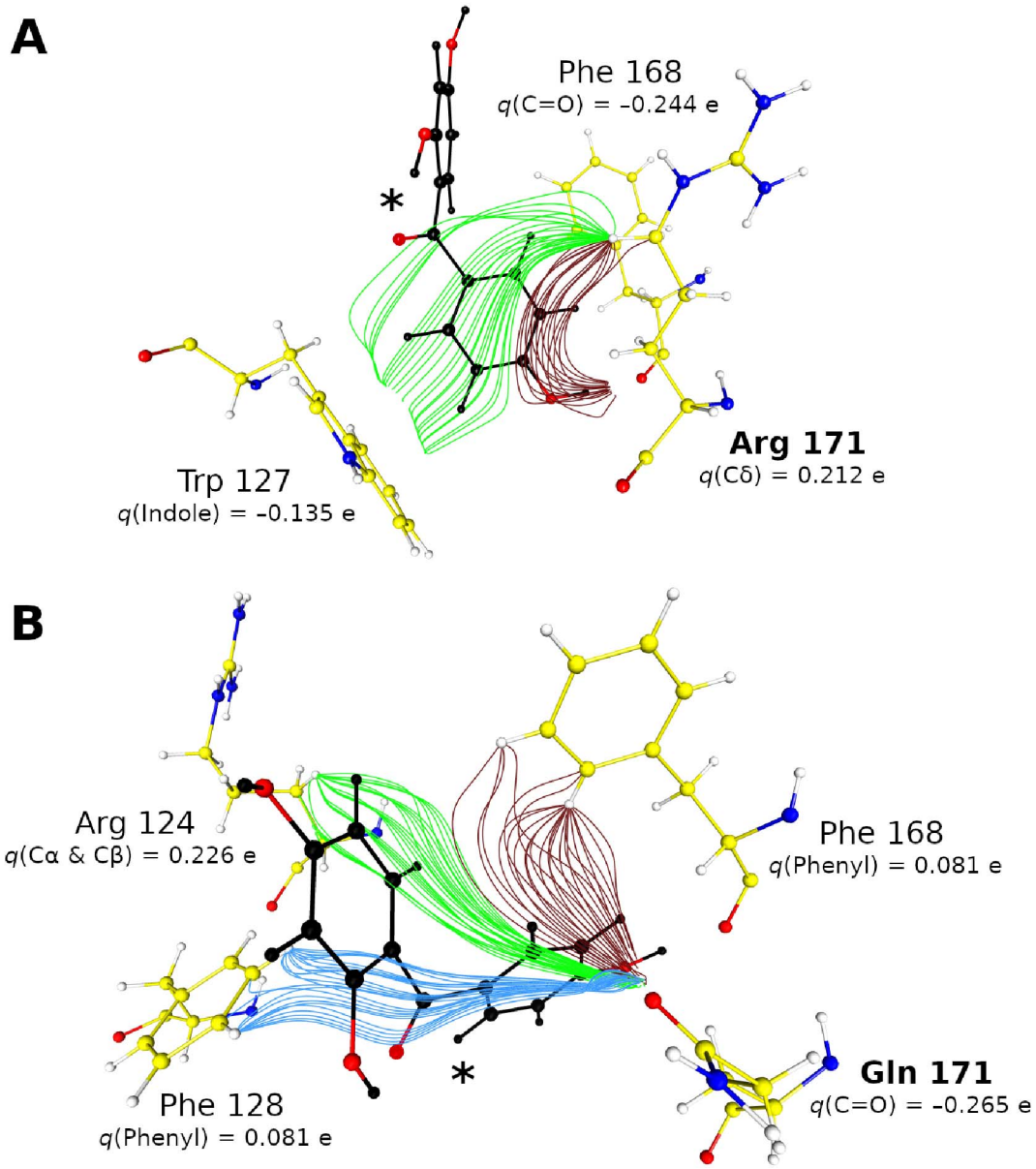

Figure S6 Electrostatic field lines in the active site. **(a)** Electric field lines converging to Arg171 in the active site of the wild-type GST (chain B, PDB code: 6f68) bound to 2,4,4'-trihydroxybenzophenone ligand (marked by an asterisk). The dark red bundle originates from the lone electron pairs of the backbone oxygen of Phe168 and terminates on the delta methylene of Arg171. The green bundle starts from the π-electrons of Trp127 and end on the delta methylene of Arg171. Charges are displayed for involved functional groups to give a sense of how steep the electric field is in a given bundle. (**B)** Electric field lines originating from lone pairs of the Gln171 delta oxygen, in the active site of the Arg171Gln GST variant (chain B), bound to the 2,4,4'-trihydroxybenzophenone ligand (marked by an asterisk). The dark red bundle ends on the phenyl hydrogen atoms of Phe168. The green bundle ends on the alpha and beta positions of Arg124, covering atoms from both benzophenone cycles. The dark blue bundle ends on phenyl hydrogens from Phe128.

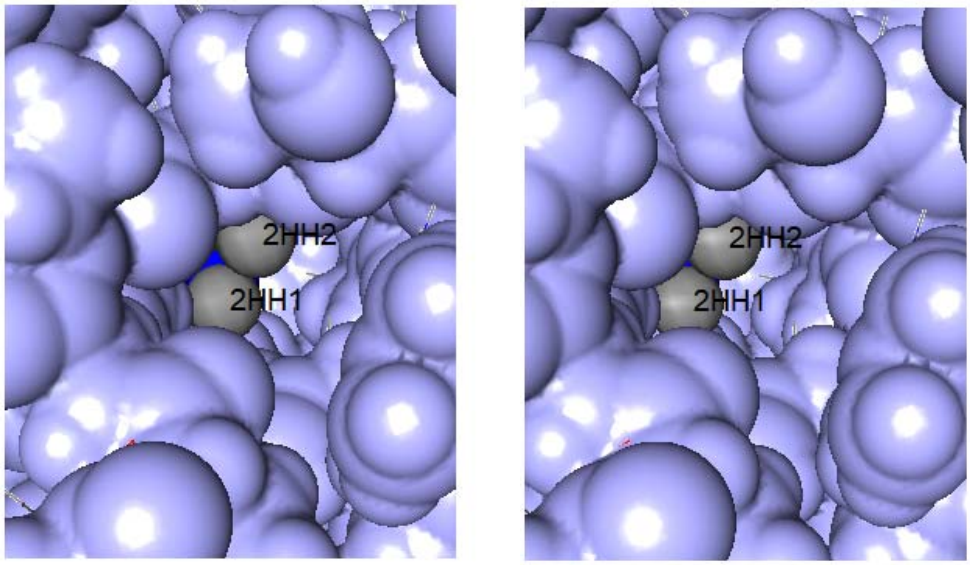

Figure S7 Stereographic view van der Waals surface of the cavity around the Arg171 side chain extremity.