Electronic Supplementary Material (ESI) for Physical Chemistry Chemical Physics.

Supporting Information:

On the Polarization of Ligands by Proteins

Soohaeng Yoo Willow,[†] Bing Xie,[†] Jason Lawrence,[‡] Robert S. Eisenberg,[¶] and David D. L. Minh[*,†]

†*Department of Chemistry, Illinois Institute of Technology, Chicago, Illinois, 60616*

‡*Department of Computer Science, Illinois Institute of Technology, Chicago, Illinois, 60616*

¶*Department of Applied Mathematics, Illinois Institute of Technology, Chicago, Illinois, 60616*

E-mail: dminh@iit.edu

Histograms

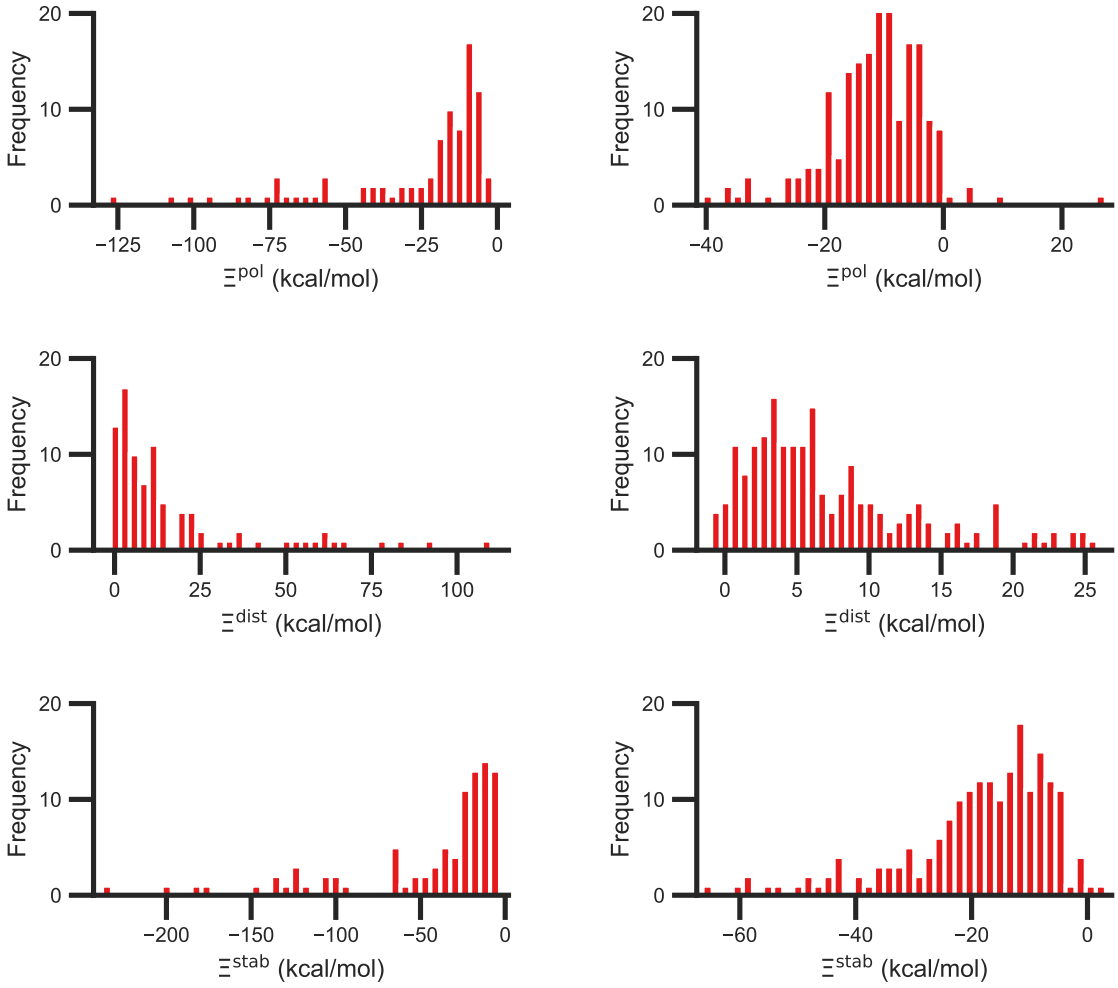


Figure S1: Histograms of the ligand polarization (top, Ξ^{pol}), distortion (middle, Ξ^{dist}), and stabilization (bottom, Ξ^{stab}) energies in the PDBBind Core Set for systems with (left) and without (right) cations. The three quantities are related by $\Xi^{\mathrm{pol}} = \Xi^{\mathrm{dist}} + \Xi^{\mathrm{stab}}$.

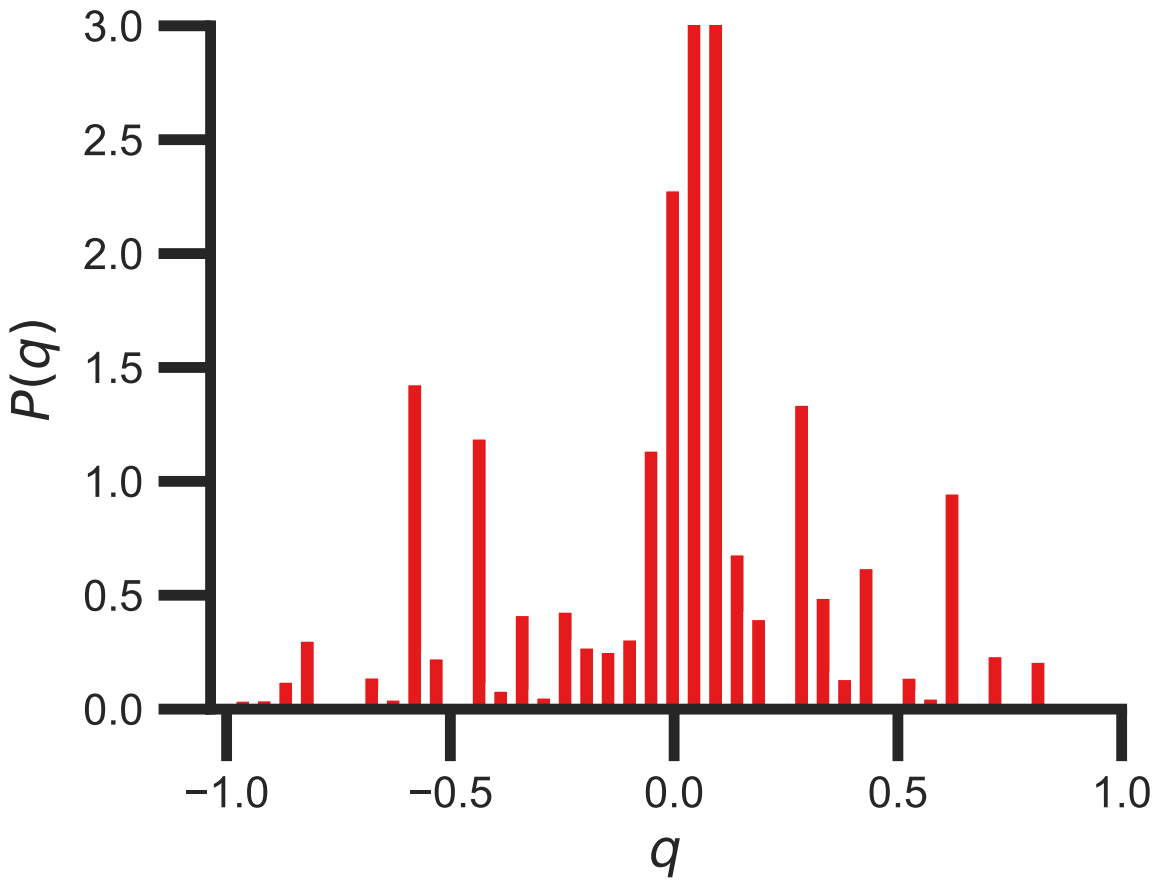


Figure S2: Normalized histogram of partial atomic charges for protein atoms in the data set.

Correlations

The percentage of atoms in a protein that are highly charged does not appear to be a significant factor in the ligand polarization energy (Fig. S3). In all the systems, only a small percentage of atoms (less than 8%) have atomic charges of $|q| \geq 0.6$.

The number density of charged atoms is more related with the ligand polarization energy. However, we only observed weak correlation, with Pearson's R being -0.34 for all complexes and -0.34 for complexes for which $\Xi^{\text{pol}} > -50$ kcal/mol (Fig. S3).

It would be reasonable to think that the polarization energy is related to the Coulomb interaction energy. However, the correlation is also weak, with Pearson's R being 0.36 for all complexes and 0.25 for complexes for which $\Xi^{\text{pol}} > -50$ kcal/mol (Fig. S3).

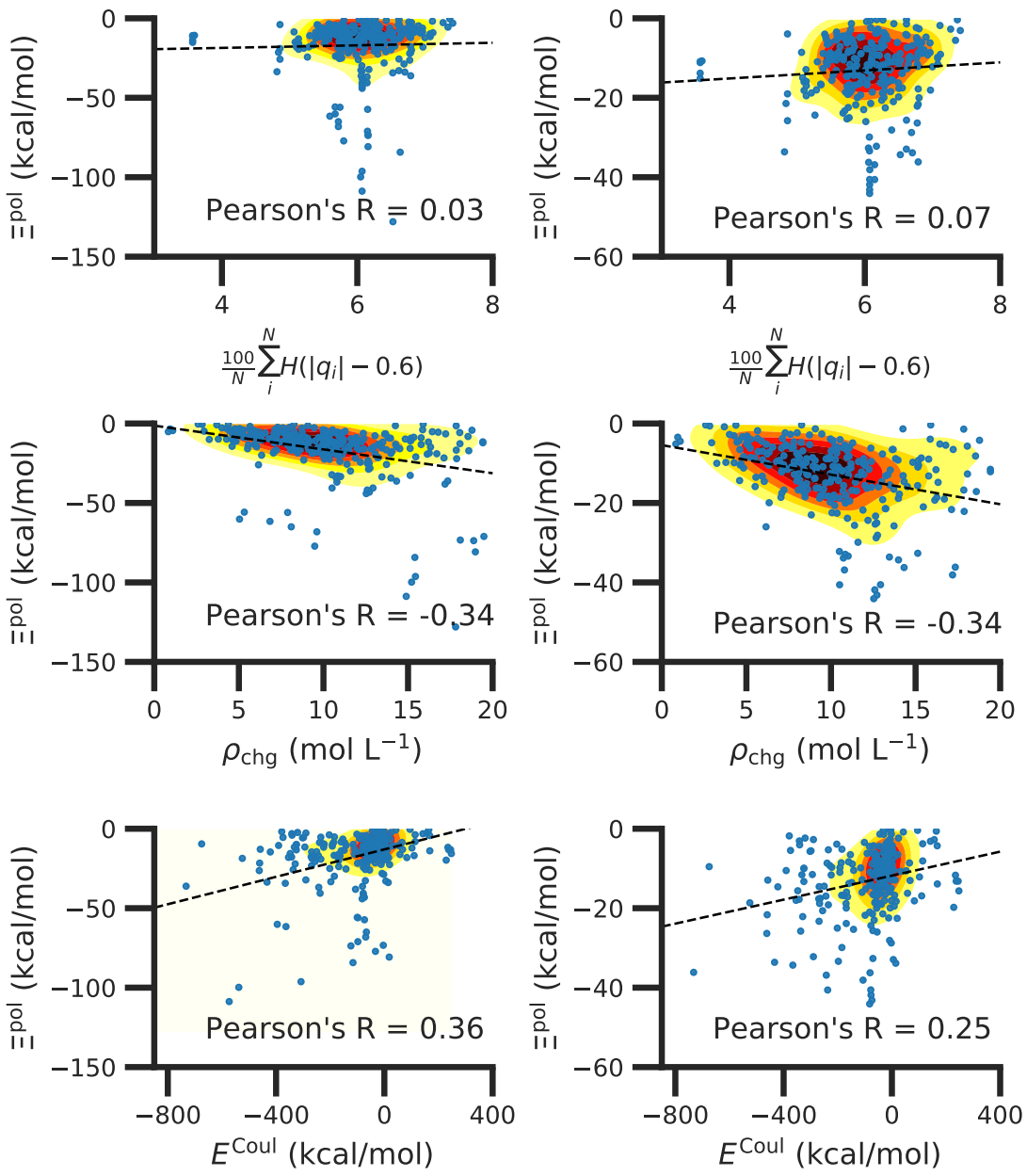


Figure S3: The polarization energy Ξ^{pol} as a function of the percentage of charged atoms (top), the number density of highly charged atoms (middle), and Coulomb energy E^{Coul}. Data are included for complexes with $\Xi^{pol} < 0$ kcal/mol (left) or only for complexes with -50 kcal/mol $< \Xi^{pol} < 0$ kcal/mol (right). The number density of charged atoms in a binding site is defined as the number of charged atoms with $|q| > 0.6$ divided by the volume of the site. The volume of the site is the region within 6 Å of any ligand atom.

The molecular polarizability scalar of ligand molecules (α_L) has a strong linear correlation with the number of electrons in the system (Fig. S4). This observation is reminiscent of one of the properties of halide anions, whose polarizabilities are observed in the following order: $F^- < Cl^- < Br^- < I^-$. However, there is no clear relationship between the molecular polarizability scalar and the ligand polarization energy.

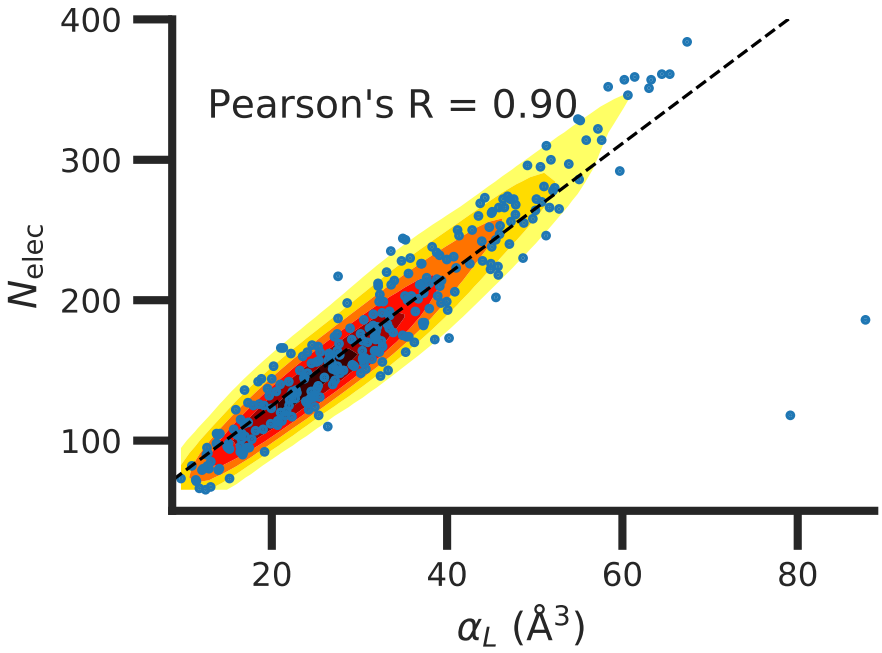


Figure S4: Polarizability of the ligand (α_L) versus the the electron number (N_{elec}) in ligands from the protein-ligand complexes.

In contrast with the aforementioned properties, there is a much clearer relationship between the ligand polarization energy, Ξ^{pol}, and the magnitude of the electric field (Fig. S5). The linear correlation is strong with the magnitude of the electric field on the ligand center of mass, $|\mathbf{E}_L^0|$, and even stronger with the magnitude of the total electric field vector active on all ligand atoms, $\left|\sum_{A\in L} \mathbf{E}_A^0\right|$. Intriguingly, in both cases, there appear to be two distinct trends relating the electric field to the magnitude of the electric field; a linear correlation exists in systems where Ξ^{pol} < -50 kcal/mol, but the slope is distinct from in systems where -50 kcal/mol < Ξ^{pol} < 0 kcal/mol. The two measures of the electric field are also correlated with each other, with a Pearson's R of 0.54 (Fig. S6). In general, the magnitude of $\left|\sum_{A\in L} \mathbf{E}_A^0\right|$ is greater than the magnitude of Ξ^{pol}, suggesting that electric field vectors on individual atoms generally point in a similar direction.

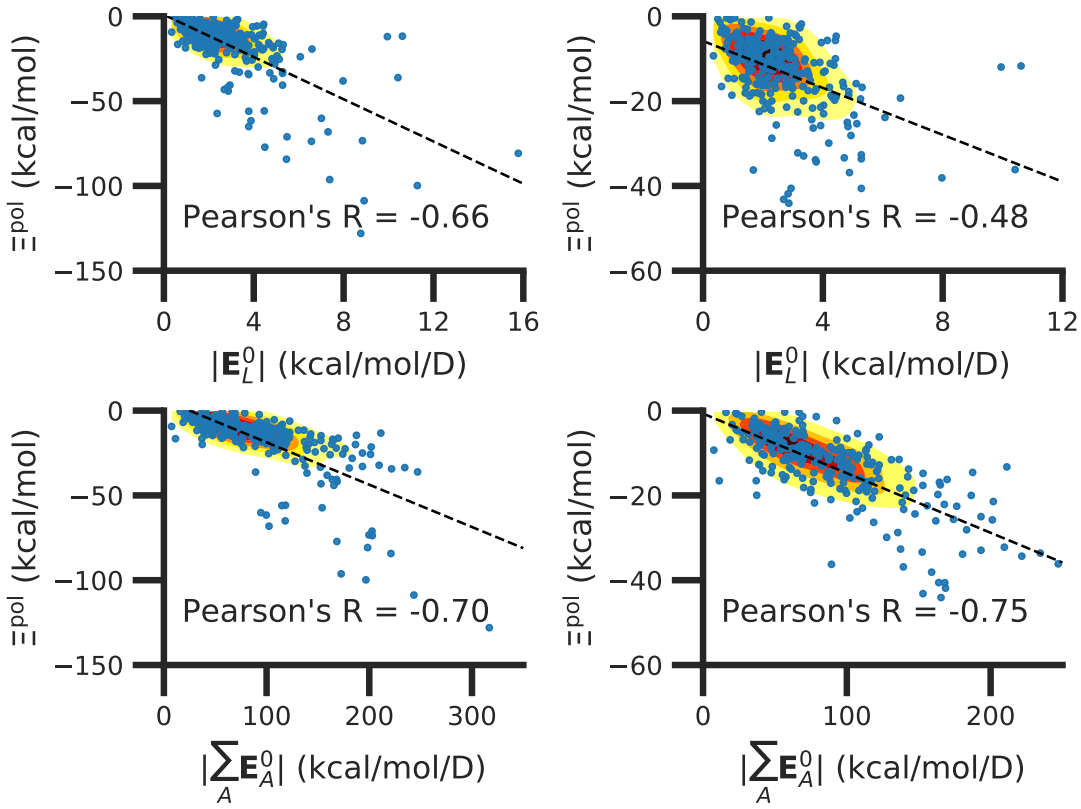


Figure S5: The ligand polarization energy, Ξ^{pol}, as a function of the magnitude of the electric field. The electric field vector is either on the ligand center of mass, $|\mathbf{E}^0_L|$, where $\mathbf{E}^0_L$ is from Eq. 22 (top) or the sum of vectors on the ligand atom sites, $\left|\sum_{A\in L} \mathbf{E}^0_A\right|$, where $\mathbf{E}^0_A$ is from Eq. 29 (bottom). The range of Ξ^{pol} is either $\Xi^{\text{pol}} < 0$ kcal/mol (left) or -50 kcal/mol $< \Xi^{\text{pol}} < 0$ kcal/mol (right).

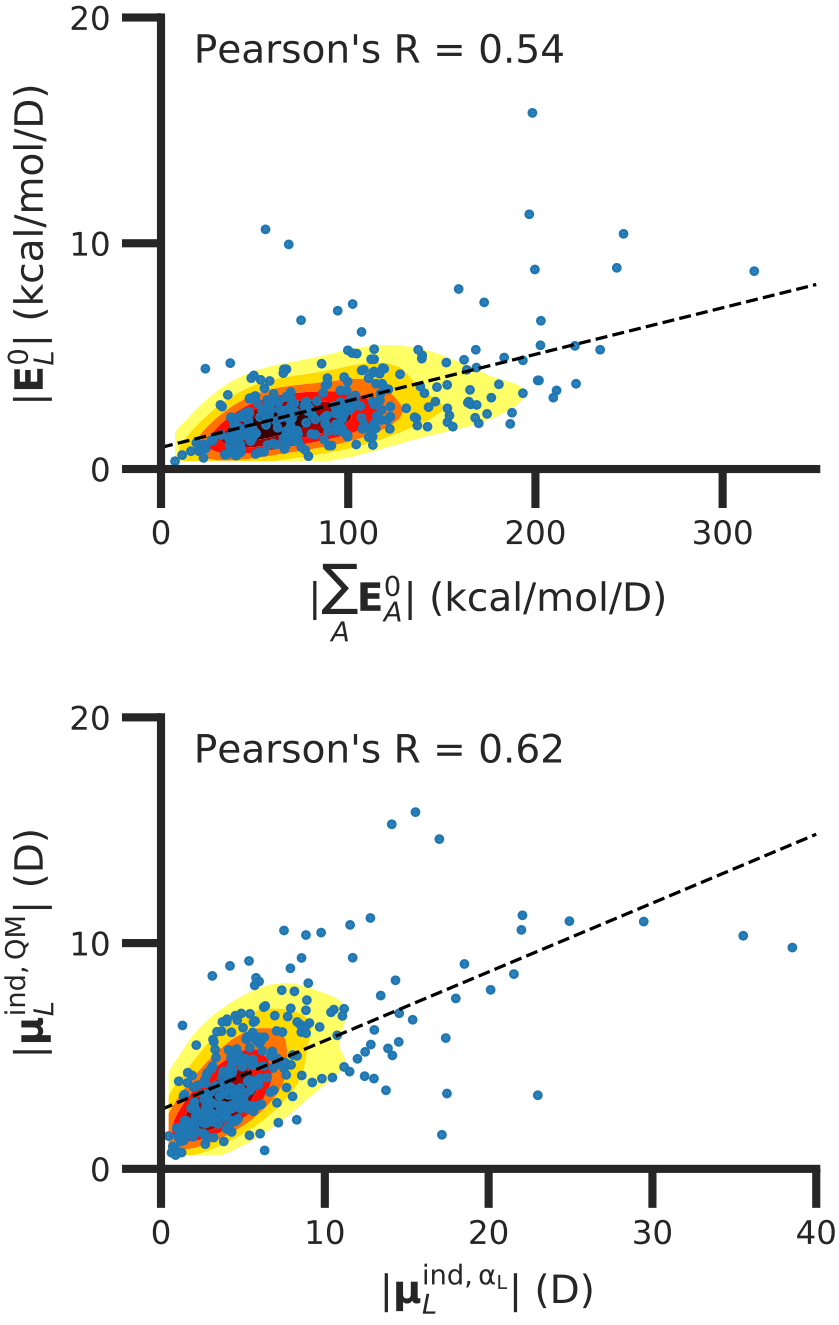


Figure S6: **Comparison of electric field estimates.** The magnitude of the electric field on the ligand, $|\mathbf{E}_L^0|$, versus of the vector sum of the electric field on all ligand atoms, $|\sum_{A\in L} \mathbf{E}_A^0|$ (top). The magnitude of the induced dipole based on the molecular polarizability tensor, $|\mu_L^{\text{ind},\alpha_L}|$ versus of the induced dipole based on the dipole operator, $|\mu_L^{\text{ind,QM}}|$ (bottom).

Similarly, the ligand polarization energy Ξ^{pol} is also correlated with the magnitude of the induced dipole moment on the ligand. There is a stronger correlation between the ligand polarization energy Ξ^{pol} and the induced dipole based on the wave functions $\boldsymbol{\mu}_L^{\text{ind,QM}}$ (Eq. 23) than the induced dipole based on the molecular polarizability tensor $\boldsymbol{\mu}_L^{\text{ind},\alpha_{\text{L}}}$ (Fig. S7). The latter quantity, $\boldsymbol{\mu}_L^{\text{ind},\alpha_{\text{L}}}$, which is ultimately based on three pairs of point charges, does not perfectly recapitulate polarizability of the more complex embedding field; Pearson's R is 0.62 (Fig. S6.)

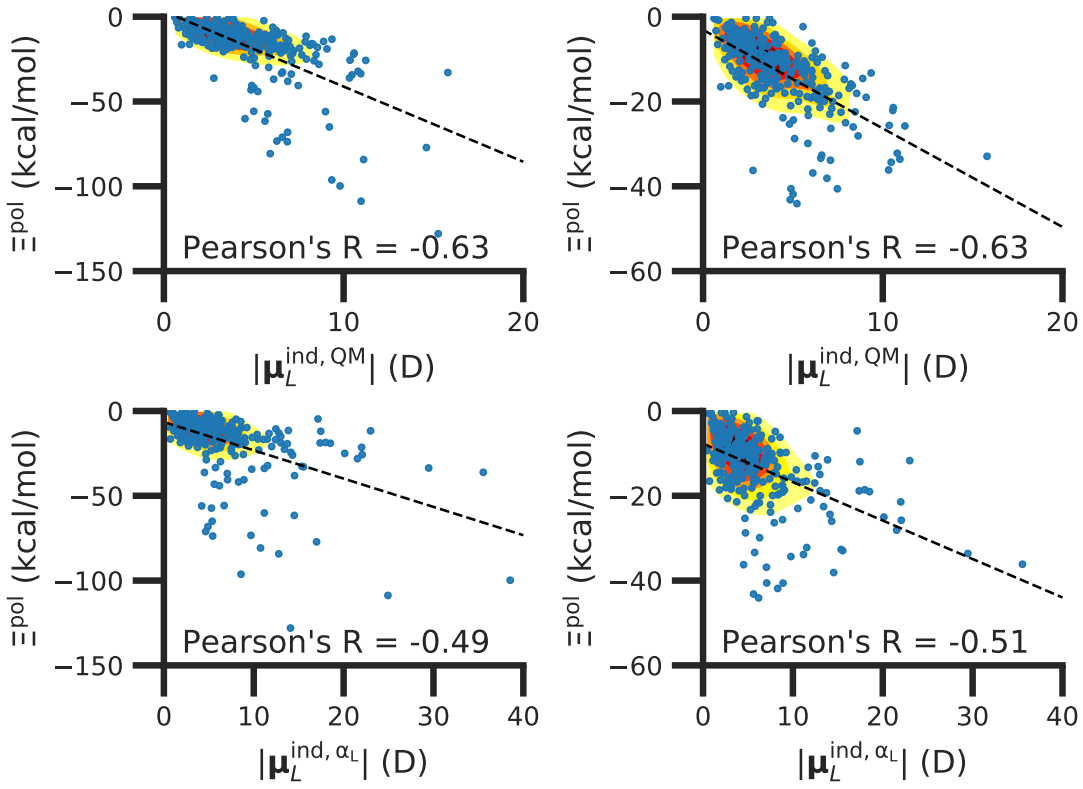


Figure S7: The ligand polarization energy, Ξ^{pol}, as a function of the magnitude of the induced dipole moment. The induced dipole moment is either based on wave functions, $|\boldsymbol{\mu}_L^{\text{ind,QM}}|$, where $\boldsymbol{\mu}_L^{\text{ind,QM}}$ is from Eq. 23 (top) or the molecular polarizability tensor, $|\boldsymbol{\mu}_L^{\text{ind},\alpha_{\text{L}}}|$, where $\boldsymbol{\mu}_L^{ind,\alpha_L}$ is from Eq. 24 (bottom). The range of Ξ^{pol} is either $\Xi^{\text{pol}} < 0$ kcal/mol (left) or -50 kcal/mol $< \Xi^{\text{pol}} < 0$ kcal/mol (right).

In addition to the strong relationship between the ligand polarization energy Ξ^{pol} and both the magnitude of the electric field and the induced dipole, there is also a clear correspondence between the ligand polarization energy Ξ^{pol} and the classical polarization energy $\Xi^{\mathrm{pol,cL}}$ (Fig. S8). The clear correlation between the two quantities suggests that the classical model of the ligand as a single dipole in an electric field is a reasonable explanation for the quantum behavior. In contrast, the correlation between Ξ^{pol} and $\Xi^{\mathrm{pol,cA}}$ is much weaker, which indicates that the classical model of the ligand as a set of atom-centered dipoles is a poor description of the quantum phenomenon. The correlation is stronger between Ξ^{pol} and $\Xi^{\mathrm{pol,cL}}$ than between Ξ^{pol} and $\Xi^{\mathrm{pol,cL},\alpha_{\mathrm{L}}}$ because the molecular polarizability model does not perfectly capture the induced dipole moment (Fig. S6).

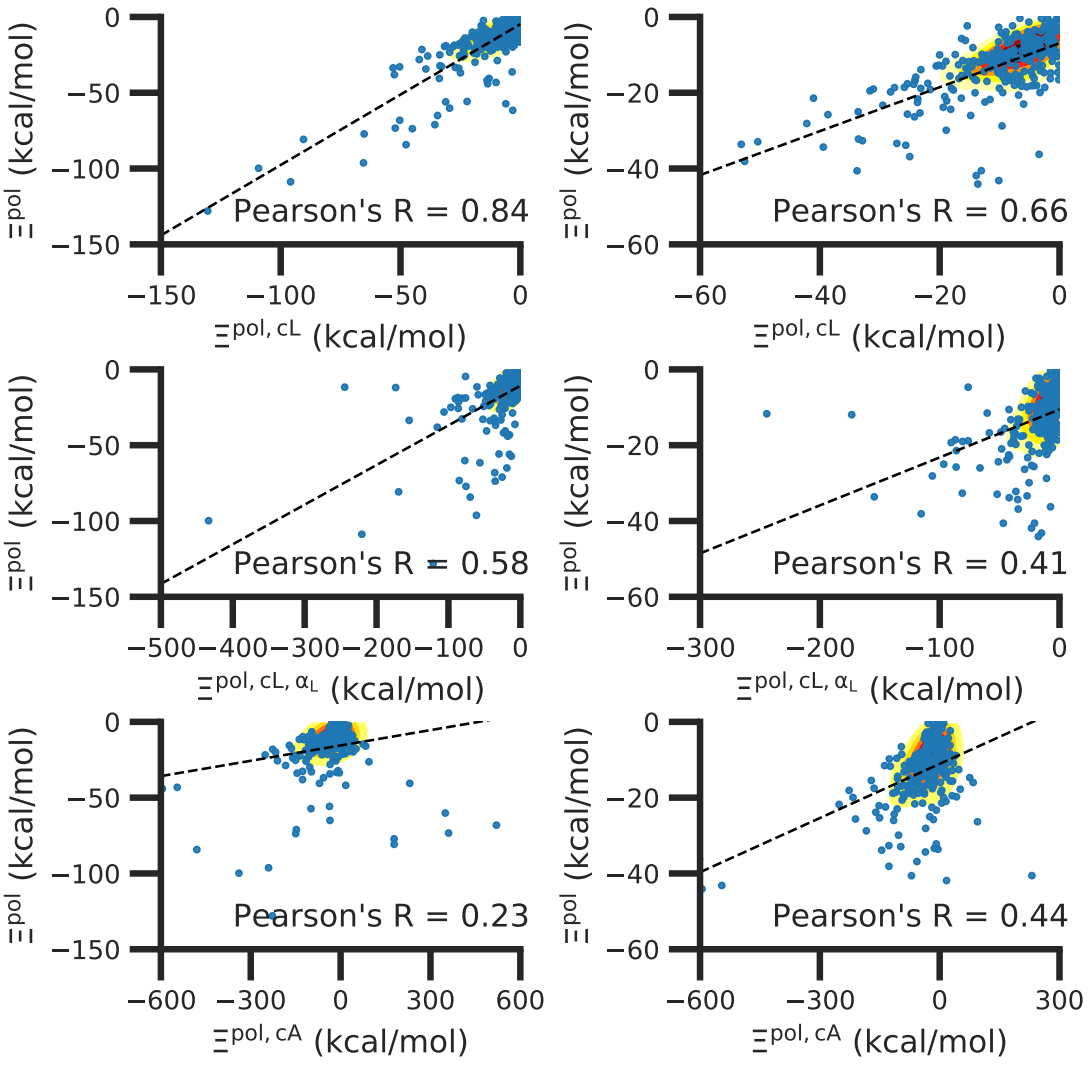


Figure S8: The ligand polarization energy, Ξ^{pol}, as a function of the classical polarization energy. The classical polarization energy is either $\Xi^{\text{pol,cL}}$ (Eq. 22), using Eq. 23 for the induced dipole moment (top), $\Xi^{\text{pol,cL,}\alpha_{\text{L}}}$ (Eq. 22), using Eq. 24 for the induced dipole moment (middle), or $\Xi^{\text{pol,cA}}$ (Eq. 27). The range of Ξ^{pol} is either $\Xi^{\text{pol}} < 0$ kcal/mol (left) or -50 kcal/mol $< \Xi^{\text{pol}} < 0$ kcal/mol (right).

Limitations of molecular polarizability model

In many cases, the failure of the molecular polarizability to recapitulate the induced dipole is due to the location of the ligand center of mass. For most complexes, the magnitude of the induced dipole moment based on the molecular polarizability tensor, $|\mathbf{E}^{\mathrm{ind},\alpha_L}_{L\boldsymbol{\mu}_L}|$ is comparable to the magnitude of the induced dipole from the quantum mechanical operator, $|\boldsymbol{\mu}^{\mathrm{ind,QM}}_L|$. However, in nearly 8% of complexes, $|\mathbf{E}^{\mathrm{ind},\alpha_L}_{L\boldsymbol{\mu}_L}|$ is much larger than $|\boldsymbol{\mu}^{\mathrm{ind,QM}}_L|$. In many of these cases, such as 3tsk (Fig. S9), the ligand center of mass is within the protein (Fig. S10). Because the ligand center of mass is within the protein, it is very close to embedding field charges and the magnitude of the electric field is particularly strong.

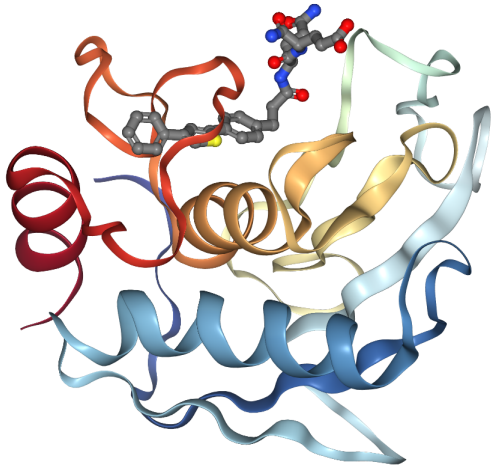


Figure S9: The structure of the complex 3tsk of human matrix metalloprotease-12 (MMP12) in complex with L-glutamate motif inhibitor. The center of mass of the ligand is placed inside the protein.

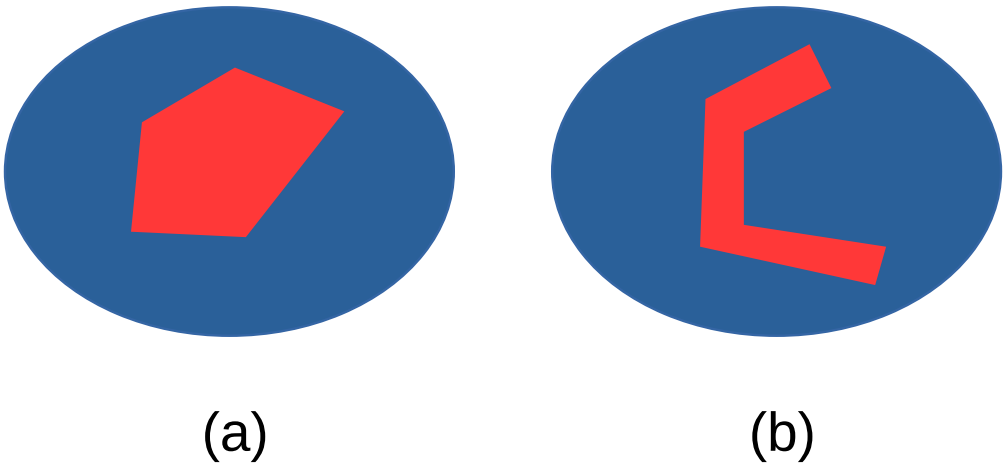


Figure S10: Schematic of protein-ligand complexes in which the ligand center of mass is inside the ligand or the protein. The ligand is colored red and protein blue. In (a), the center of mass of the ligand is placed inside the ligand, whereas in (b) the center of mass of the ligand is inside the protein.

Other figures

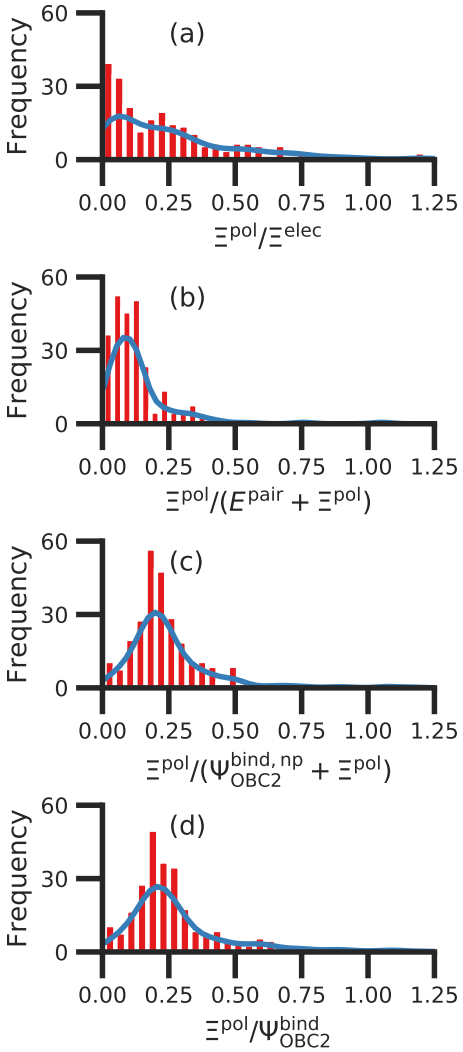


Figure S11: Histograms of ratio of the polarization energy of the ligand to (a) the electrostatic interaction ($\Xi^{elec} = E^{Coul} + \Xi^{pol}$), (b) the intermolecular pairwise potential energy with the ligand polarization energy ($E^{pair} + \Xi^{pol}$), (c) the binding energy without considering ligand polarization in the solvation free energy ($\Psi^{bind,np}_{OBC2} + \Xi^{pol}$), and (d) the binding energy with considering ligand polarization in the solvation free energy (Ψ^{bind}_{OBC2}). Data are from all complexes where $\Xi^{pol} < 0$ kcal/mol. The histograms are truncated at a ratio of 1.25.

Solvent Effect on Ligand Polarization Energy

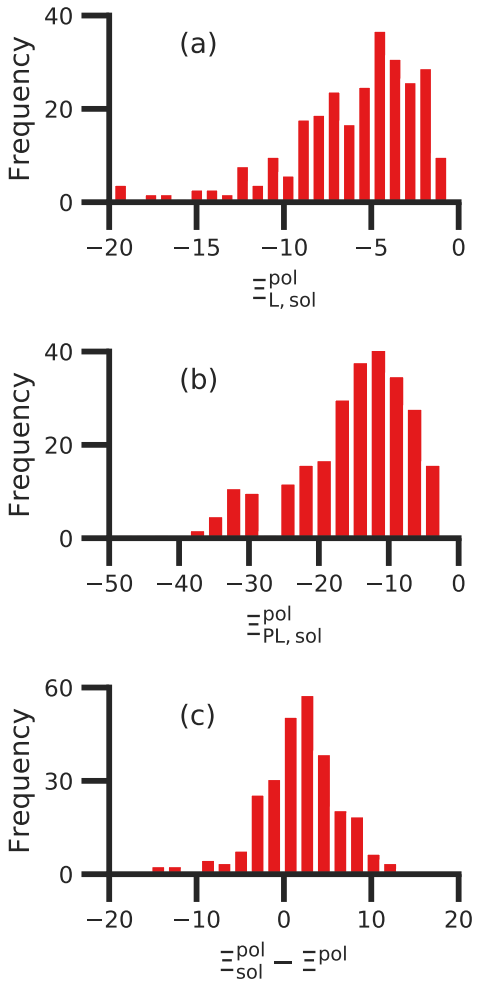


Figure S12: Histograms of ligand polarization energies of (a) ligands and (b) protein-ligand complexes in the ddCOSMO implicit solvent. (c) The solvent effect on the ligand polarization energy in terms of the change in the ligand polarization energy due to the ddCOSMO implicit solvent.

The effect of ddCOSMO implicit solvent on the ligand polarization energies is in the range from -20 kcal/mol $< \Xi^{pol}_{L,sol} < 0$ kcal/mol. Most values are near -5 kcal/mol. In case of the protein-ligand complexes, the ligand polarization energies are much lower: -40 kcal/mol $< \Xi^{pol}_{PL,sol} < 0$ kcal/mol, owing to the strong electric field from the atomic charges $\{q_F\}$ of the protein. Figure S12 (c) shows the solvent effect on the ligand polarization energy: $\Xi^{pol}_{sol} - \Xi^{pol}$. Only a small percentage (7%) belongs to $|\Xi^{pol}_{sol} - \Xi^{pol}| > 10$ kcal/mol. For

most complexes (70%), $|\Xi_{\rm sol}^{\rm pol} - \Xi^{\rm pol}| < 5$ kcal/mol. Since most ligands are embedded in their receptors (proteins), the solvent accessible areas on ligands are smaller. This is why the solvent effect on the ligand polarization energy is small compared to the contribution of atomic charges of their receptors.

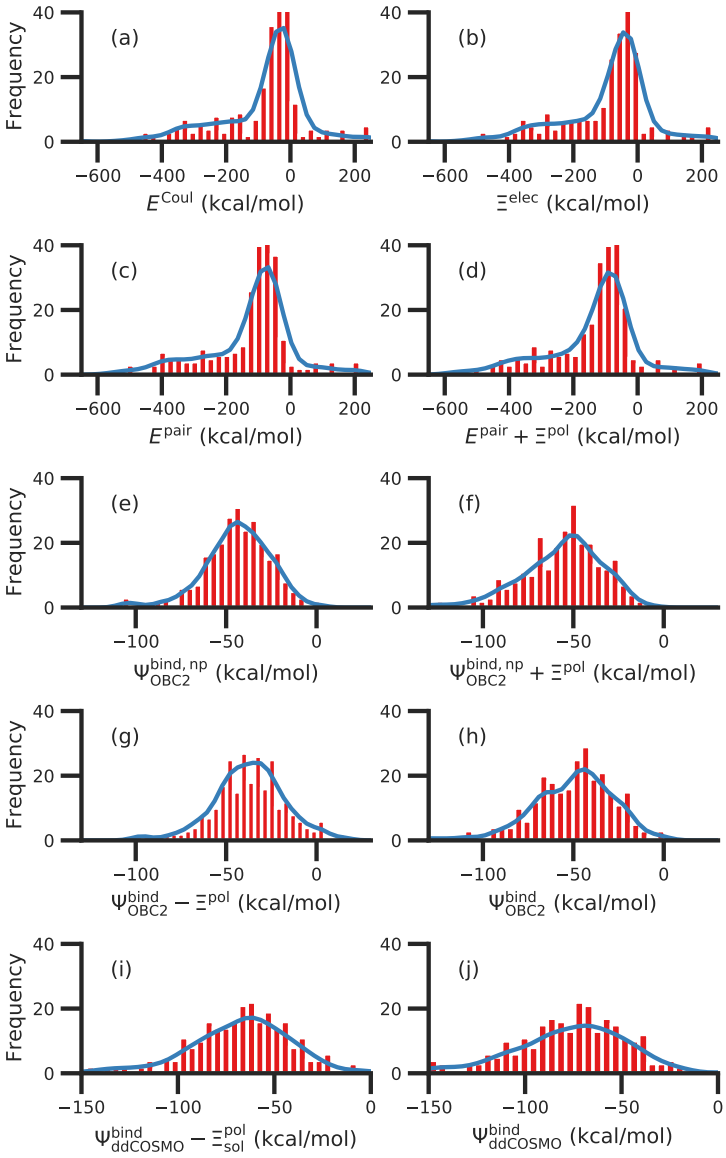
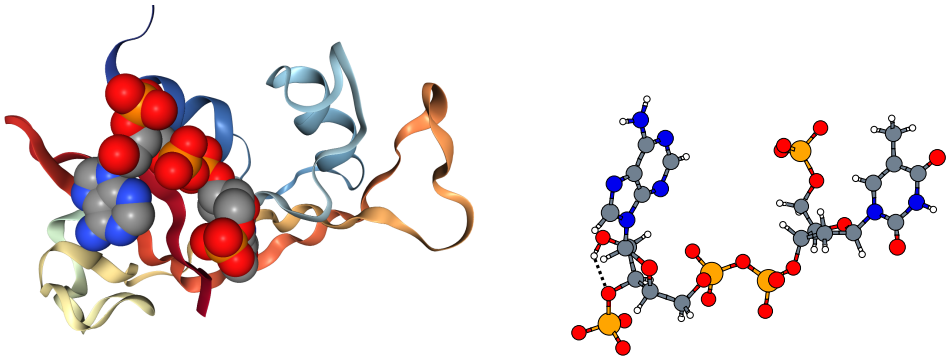


Figure S13: Histograms of intermolecular potential energies and binding energies. The intermolecular potential energies are (a) the permanent Coulomb interaction (E^{Coul}), (b) the electrostatic interaction ($\Xi^{\mathrm{elec}} = E^{\mathrm{Coul}} + \Xi^{\mathrm{pol}}$), (c) the intermolecular pairwise potential energy ($E^{\mathrm{pair}} = E^{\mathrm{vdW}} + E^{\mathrm{Coul}}$), and (d) the intermolecular pairwise potential energy with the polarization energy of the ligand ($E^{\mathrm{pair}} + \Xi^{\mathrm{pol}}$) in the gas phase. The OBC2 binding energies are (e) without considering ligand polarization at all, $\Psi^{\mathrm{bind,np}}_{\mathrm{OBC2}}$, (f) considering ligand polarization for electrostatic interactions but not in the solvation free energy, $\Psi^{\mathrm{bind,np}}_{\mathrm{OBC2}} + \Xi^{\mathrm{pol}}$, (g) considering ligand polarization in the solvation free energy but not for electrostatic interactions, $\Psi^{\mathrm{bind}}_{\mathrm{OBC2}} - \Xi^{\mathrm{pol}}$, (h) considering ligand polarization both in the electrostatic interactions and the solvation free energy. The ddCOSMO binding energies are (i) without and (j) with the ligand polarization energy. A similar plot that includes systems containing cations is available in the main text.



(a) (b)

Figure S14: (a) The structure of the complex 1u1b of bovine pancreatic Ribonuclease A with the ligand (3′-phosphothymidine (3′-5′)-pyrophosphate adenosine 3′-phosphate) and (b) the molecular structure of the ligand.

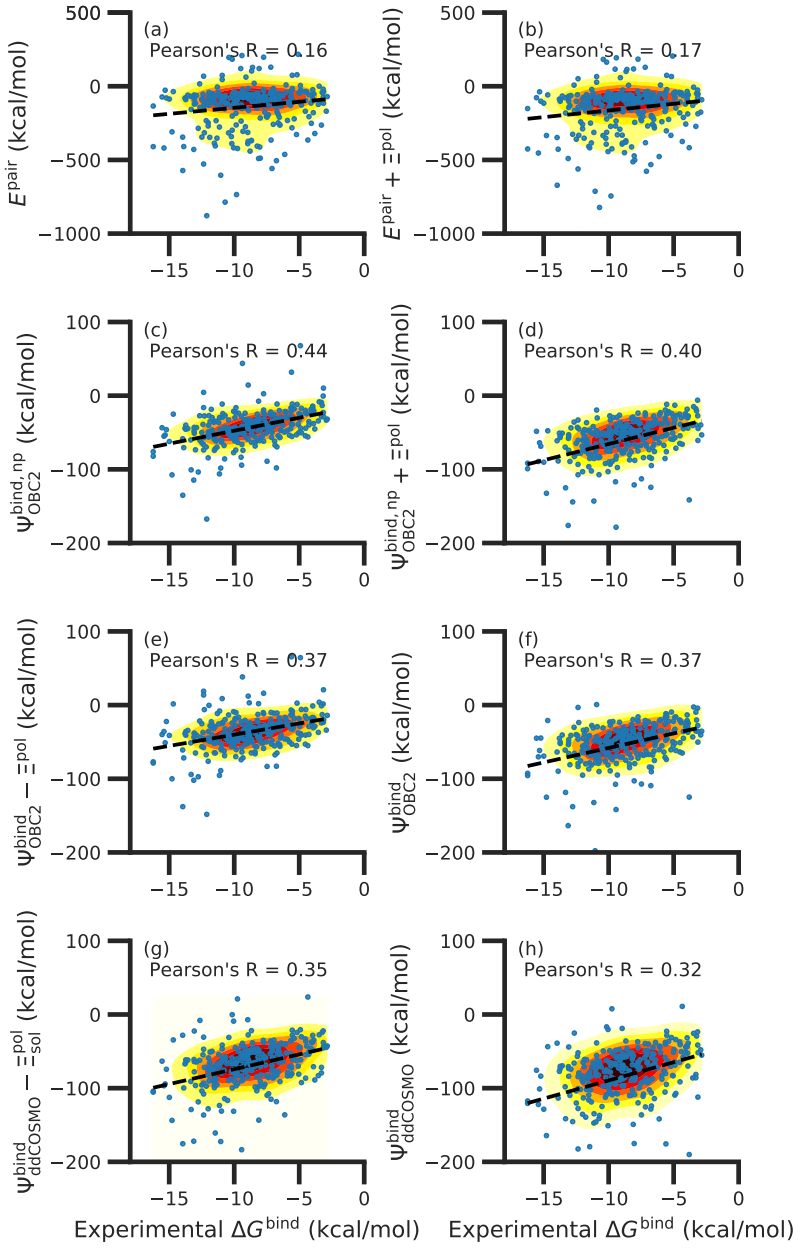


Figure S15: Comparison of interaction energies to experimentally measured binding free energies for all complexes where $\Xi^{pol} < 0$ kcal/mol. Interaction energies are according to (a) the intermolecular pairwise potential energy ($E^{pair} = E^{vdW} + E^{Coul}$) and (b) the intermolecular pairwise potential energy with the polarization energy of the ligand ($E^{pair} + \Xi^{pol}$) in the gas phase. Panels (c-h) are binding energies, with (c-f) based on the OBC2 and (g-h) based on the ddCOSMO implicit solvent models. The OBC2-based binding energies are: (c) without considering ligand polarization at all, $\Psi^{bind,np}_{OBC2}$; (d) considering ligand polarization for electrostatic interactions but not in the solvation free energy, $\Psi^{bind,np} + \Xi^{pol}$; (e) considering ligand polarization in the solvation free energy but not for electrostatic interactions, $\Psi^{bind} - \Xi^{pol}$; or (f) considering ligand polarization both in the electrostatic interactions and the solvation free energy. The ddCOSMO-based binding energies are (g) without and (h) with considering the ligand polarization energy.

Estimated overpolarization

When cations are close to ligands, the extent of ligand polarization is likely overestimated by the QM/MM scheme used in this paper. To assess the extent of overpolarization, we performed some calculations in which cations were included in the QM region. In this modified scheme, there is no direct way to isolate the ligand polarization energy from energy of the complex. Instead, we estimated the induced dipole of the ligand based on RESP atomic charges,

$$\mu_L^{\mathrm{ind,RESP}} = \sum_{A \in L} (q_A^{\mathrm{QM:Q_L}} - q_A^{\mathrm{QM}}) \boldsymbol{R}_A. \tag{1}$$

The ligand polarization energy is then computed by,

$$\Xi^{\mathrm{pol,RESP}} \quad = \quad -\boldsymbol{\mu}_L^{\mathrm{ind,RESP}} \cdot \mathbf{E}_L^0, \tag{2}$$

where $\mathbf{E}_L^0$ is the electric field acting on the center of mass of the ligand.

In the selected systems where cations are very close to ligand atoms, the ligand polarization energy estimated with the main QM/MM scheme in this paper is likely too low (Table S1). When the only ligand is in the QM region, the estimated ligand polarization energy is fairly consistent; $\Xi^{\mathrm{pol}}(L) \sim \Xi^{\mathrm{pol,cL}}(L) \sim \Xi^{\mathrm{pol,RESP}}(L)$. When the QM region is expanded to include cations, the ligand polarization energy based on RESP atomic charges is significantly higher. For 3dx1 and 3dx2, it is about 20 kcal/mol higher. For 2zcq, where Ξ^{pol} is especially low, $\Xi^{\mathrm{pol,RESP}}(LC)$ has the opposite sign!

Table S1: Dependence of the ligand polarization energy on the QM region. $\Xi^{\mathrm{pol}}(X)$, $\Xi^{\mathrm{pol,cL}}(X)$, and $E^{\mathrm{pol,RESP}}(X)$ are from Eqs. 6, 22, and S2, respectively, with the QM region of X. Here, either the ligand (L) or the ligand and cations (LC) are included in the QM region. The unit of the polarization energy is in kcal/mol.

PDB ID	$\Xi^{\mathrm{pol}}(L)$	$\Xi^{\mathrm{pol,cL}}(L)$	$\Xi^{\mathrm{pol,RESP}}(L)$	$\Xi^{\mathrm{pol,RESP}}(LC)$
3dx1	-80.77	-92.20	-87.55	-69.11
3dx2	-73.34	-51.71	-49.51	-30.77
2zcq	-128.01	-132.86	-128.67	109.02

Table S2: Complexes with ratios outside of the range of Fig. 7.

PDB ID	Ξ^{pol}	Ξ^{elec}	$\Xi^{\text{pol}}/\Xi^{\text{elec}}$
1mq6	-10.590	48.843	-0.217
1o3f	-18.158	56.933	-0.319
1oyt	-18.273	-3.716	4.917
3gc5	-28.756	-12.884	2.232
3ge7	-33.854	2.873	-11.785
3gy4	-19.494	-12.887	1.513
3ui7	-16.902	93.232	-0.181
3uuo	-8.104	108.339	-0.075
4abg	-12.907	-4.975	2.594
4ea2	-32.932	21.277	-1.548
4llx	-12.174	-6.190	1.967
4mme	-10.732	21.803	-0.492
4msc	-18.087	6.736	-2.685
4msn	-7.198	8.481	-0.849
5c1w	-11.798	5.292	-2.229
5c28	-10.404	15.595	-0.667
5c2h	-22.453	-2.850	7.878

PDB ID	Ξ^{pol}	$E^{\text{pair}} + \Xi^{\text{pol}}$	$\Xi^{\text{pol}}/(E^{\text{pair}} + \Xi^{\text{pol}})$
1o3f	-18.158	27.304	-0.665
3ui7	-16.902	51.054	-0.331
3uuo	-8.104	65.857	-0.123

PDB ID	Ξ^{pol}	$\Psi^{\text{bind,np}} + \Xi^{\text{pol}}$	$\Xi^{\text{pol}}/(\Psi^{\text{bind,np}} + \Xi^{\text{pol}})$
2weg	-68.146	-53.759	1.268
3dx1	-80.769	-12.953	6.235
3dx2	-73.340	-29.526	2.484
3kwa	-77.169	-45.308	1.703