Journal of Molecular Graphics and Modelling 27 (2009) 846–851

Contents lists available at ScienceDirect

Journal of Molecular Graphics and Modelling

journal homepage: www.elsevier.com/locate/JMGM

Short communication

# Coulomb energies of protein–protein complexes with monopole-free charge distributions

Madhurima Das [b], Gautam Basu [a,*]

[a] Department of Biophysics, Bose Institute, P-1/12 CIT Scheme VIIM, Kolkata 700054, India
[b] Bioinformatics Centre, Bose Institute, P-1/12 CIT Scheme VIIM, Kolkata 700054, India

ARTICLE INFO

*Article history:*
Received 21 August 2008
Received in revised form 11 December 2008
Accepted 12 December 2008
Available online 24 December 2008

*Keywords:*
Protein–protein interactions
Electrostatics
Monopole
Dipole

ABSTRACT

From an analysis of Coulomb energy distributions of a large set of protein–protein complexes we show that the positive tail in the energy distribution disappears when the monopole–monopole term, the only energy term independent of inter-subunit orientations, is removed. This indicates that unfavorable Coulomb energies associated with subunit orientations are excluded in protein–protein complexes. The overall result remained unchanged when solvent effects were included. Our results have important bearing on the restriction of subunit orientations in protein–protein complexes and complement a recent work [K. Brock, K. Talley, K. Coley, P. Kundrotas, E. Alexov, Optimization of electrostatic interactions in protein–protein complexes, Biophys. J. 93 (2007) 3340–3352.] which showed that Coulomb energy of interaction in protein–protein complexes is sequence-optimized.



## 1. Introduction

Understanding the energetic basis that drives protein–protein interactions is extremely important since such complexes modulate a variety of biological functions. This is evident from the recent surge of interest in the field from experimentalists as well from computational biologists [1,2]. The availability of large number of structures of protein–protein complexes in the protein data bank (pdb) [3] has allowed a detailed statistical and energetic analysis of protein–protein interfaces yielding important empirical rules [4,5]. However, a thorough understanding of the diversity and subtlety of protein–protein complexes [6] is yet to be completely captured. In addition to local interactions, active at the protein–protein interfaces, interactions that depend on global properties of proteins, like shape, size or spatial charge distribution, also can contribute to protein–protein interactions. If the spatial charge distributions of two protein subunits are known then the electrostatic component of binding energies can be estimated by numerically solving the Poisson–Boltzmann equation under a suitable piecewise continuous dielectric model [7]. Using such an approach and from calculations on a large set of protein–protein complexes with known structures, Brock et al. [8] recently demonstrated that the electrostatic components of the association free energies are mostly positive, implying that electrostatics opposes protein–protein association. However, the associated Coulomb energies were found to favor (dominantly negative) protein association and were shown to be sequence-optimized for wild type protein–protein complexes.

Whether electrostatics favors or disfavors protein–protein association needs more scrutiny since electrostatic binding energies, calculated using a piecewise dielectric continuum model, depend on the exact definition of the dielectric boundary [9]. However, the observed trend in Coulomb energies is artifact free. An important feature of the distribution of Coulomb energies was a positive energy tail, the origin of which was not addressed by Brock et al. [8]. In this report we show that the positive tail largely disappears if the monopole–monopole term is subtracted from the total Coulomb energy. This finding has important implications in electrostatic optimization of inter-subunit orientations in protein–protein complexes since of all cross-terms in the expression of Coulomb energy, only the monopole–monopole term is orientation independent.

## 2. Materials and methods

Coulomb interaction energies, $E_{AB}$, were calculated as:

$$E_{AB} = \sum_i \sum_{j>i} \frac{q_{Ai} q_{Bj}}{r_{ij}} \quad (1)$$

where $q_{Ai}$ is the partial charge on the $i$-th atom on subunit A, $q_{Bj}$ is the partial charge on the $j$-th atom on subunit B and $r_{ij}$ is the interatomic distance (see Fig. 1a). After adding hydrogen atoms with PDB2PQR [10], all calculations were performed with two sets of AMBER charges [11], one with neutral (HIS-set) and the other with

* Corresponding author. Fax: +91 33 2355 3886.
*E-mail addresses:* gautam@boseinst.ernet.in, gautamda@gmail.com (G. Basu).

1093-3263/$ – see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmgm.2008.12.002

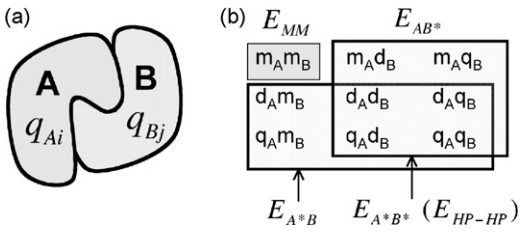


**Fig. 1.** (a) A cartoon representation of two interacting subunits in a dimeric (subunits A and B) protein complex. The subunits are comprised of $N_A$ and $N_B$ atoms, each carrying a partial charge ($q_{Ai}$ and $q_{Bi}$). The overall charge distribution of each subunit in panel (a) can be considered to be a superposition of monopole and higher order terms giving rise to a number of distinct interaction energy cross-terms. A select set of cross-terms arising from the interaction of monopoles ($m$), dipoles ($d$) and quadrupoles ($q$) of subunits A and B are shown in panel (b). $E_{MM}$ is the total monopole–monopole energy, $E_{A^*B}$ is the total energy between monopole-devoid charge distribution of subunit A and subunit B, $E_{AB^*}$ is the total energy between monopole-devoid charge distribution of subunit B and subunit A and $E_{A^*B^*}$ is the total energy between monopole-devoid charge distribution of subunit A and monopole-devoid charge distribution of subunit B.

charged (+1) histidine (HIP-set). The monopole–monopole energy term ($m_Am_B$ in Fig. 1b) was calculated as $E_{MM}$ (using total charge on the subunits) and $E_{MM0}$ (using mean charge on the subunits):

$$E_{MM} = \frac{Q_A Q_B}{R_{AB}}$$
$$E_{MM0} = \langle q_{Ai}\rangle \left\langle q_{Bj}\right\rangle \sum_i \sum_{j>i} \frac{1}{r_{ij}} \quad (2)$$

where $Q$ is the total charge on a subunit and $R_{AB}$ is the inter-subunit distance (between geometric centers). Dipole–dipole interaction energies were calculated as:

$$E_{DP-DP} = |\vec{\mu}_A||\vec{\mu}_B| \frac{[3(\hat{\mu}_A \cdot \hat{R}_{AB})(\hat{\mu}_B \cdot \hat{R}_{AB}) - \hat{\mu}_A \cdot \hat{\mu}_B]}{|\vec{R}_{AB}|^3} \quad (3)$$

where $\mu$ (=$\sum q_i r_i$; with the origin set to the geometric center of a subunit) is the dipole moment of a subunit. Coulomb energies with monopole-devoid charge distributions, on either ($E_{A^*B}$ and $E_{AB^*}$) or both ($E_{A^*B^*}$) subunits, were calculated as:

$$E_{A^*B} = \sum_i \sum_{j>i} \frac{(q_{Ai} - \langle q_{Ai}\rangle) q_{Bj}}{r_{ij}}$$
$$E_{AB^*} = \sum_i \sum_{j>i} \frac{q_{Ai}\left(q_{Bj} - \left\langle q_{Bj}\right\rangle\right)}{r_{ij}} \quad (4)$$
$$E_{A^*B^*} = \sum_i \sum_{j>i} \frac{(q_{Ai} - \langle q_{Ai}\rangle)\left(q_{Bj} - \left\langle q_{Bj}\right\rangle\right)}{r_{ij}}$$

Solvent screening of Coulomb energies were estimated by adding reaction field energies, calculated by numerically solving the linearized Poisson–Boltzmann equation using the program Delphi [12]. Parameters used in the calculation were: $\varepsilon_{in}$: 4.0; $\varepsilon_{out}$: 80.0; partial charges and radii: AMBER [11]; percent fill 95%; five stage focusing until 2 grid points/Å; solvent probe radius: 1.4 Å; ionic strength: 0.0 and 0.145 M. Reaction field and total electrostatic free energies of binding were calculated by taking appropriate difference between the bound and the unbound states [7].

The three letter pdb [3] codes of representative set of protein–protein complexes used in this work, along with the total charge on each subunit, shown within parenthesis (HIS-set; HIP-set), are given below. The homo-dimers are identical to the set of protein–protein complexes used by Rodier et al. [13]. The hetero-dimers and protein domains were extracted from the ProtCom database (June 2006) (http://www.ces.clemson.edu/compbio/protcom/), selected at 40% sequence identity level (subset: DPPC40 database).

**Homo-dimers (102):** 12as(−13; 0), 1a3c(−5; −3), 1a4i(−3; 3), 1a4u(1; 8), 1aa7(2; 3), 1ad3(−4; 8), 1ade(−10; −4), 1afw(−1; 2), 1ajs(2; 10), 1amk(2; 8), 1aor(−11; 0), 1aq6(−7; −5), 1auo(−9; −3), 1b3a(5; 6), 1b5e(−4; 1), 1b67(1; 2), 1bbh(−4; −3), 1bd0(−4; 12), 1bif(−6; 5), 1biq(−23; −15), 1bis(−2; 5), 1bjw(−5; 1), 1bkp(−9; 1), 1bmd(−6; −2), 1brw(−5; 2), 1bsl(−16; −4), 1bsr(9; 13), 1buo(−8; −3), 1bxg(−17; −8), 1bxk(−11; 2), 1cdc(0; 1), 1chm(−14; 1), 1cmb(−4; −1), 1cnz(−10; −3), 1coz(−4; 1), 1csh(0; 12), 1ctt(−9; 0), 1cvu(−2; 13), 1czj(−5; 3), 1daa(−4; 6), 1dor(−10; −5), 1dpg(−22; −16), 1dqs(1; 8), 1dxg(−4; −4), 1ebh(−5; 6), 1f13(−13; 1), 1fip(3; 3), 1gvp(3; 4), 1hxp(−9; 6), 1icw(5; 7), 1imb(−8; −4), 1isa(−6; 1), 1jhg(−3; −1), 1jsg(−6; −3), 1kba(2; 3), 1kpf(−2; 6), 1m6p(−2; 1), 1mkb(−1; 1), 1nox(3; 7), 1nse(−4; 9), 1nsy(−10; −6), 1oac(−21; −1), 1pgt(−3; −1), 1qfh(−8; −6), 1qhi(1; 11), 1qr2(−4; 2), 1r2f(−17; −12), 1reg(3; 6), 1rpo(−7; −5), 1ses(−4; 4), 1slt(−3; −1), 1smn(−2; 3), 1sox(−13; 0), 1tc1(1; 6), 1tox(−10; 6), 1trk(−4; 13), 1uby(−4; 2), 1utg(−1; 0), 1vfr(−8; 3), 1vok(12; 14), 1wtl(−1; −1), 1xso(−7; 1), 2arc(−4; 4), 2ccy(2; 3), 2hdh(3; 11), 2lig(−3; 0), 2mcg(−1; 2), 2nac(−11; 5), 2ohx(4; 11), 2spc(−5; −1), 2sqc(−14; −4), 2tct(−7; 0), 2tgi(1; 3), 3dap(−15; −2), 3grs(1; 17), 3sdh(4; 6), 3ssi(−5; −3), 4cha(3; 5), 5csm(−1; 3), 5rub(−10; 3), 8prk(−8; −2), 9wga(2; 4).

**Hetero-dimers (60):** 1ai4(−4/−5; −2/6), 1all(−3/1; −3/1), 1bpl(−6/−7; 4/7), 1bun(1/7; 4/10), 1bvn(−5/−5; 4/−3), 1cp9(−10/−11; −8/4), 1dj7(−7/−2; −4/2), 1efv(0/2; 5/4), 1euc(9/−9; 19/−7), 1f2t(6/−3; 9/−1), 1f3v(−4/2; 0/4), 1f60(12/−10; 23/−8), 1fm0(−7/−6; −5/−2), 1fm2(−9/−11; −5/−1), 1frf(−9/1; −1/20), 1gh6(−10/5; −7/13), 1gka(2/−5; 4/−1), 1gl4(−7/3; 5/5), 1h2a(−5/2; 3/24), 1hx1(−5/−4; 7/−3), 1jat(−4/−1; −2/3), 1ksg(−3/−2; 0/−1), 1lw6(−2/1; 10/1), 1lzw(−5/−8; −4/−2), 1okk(−1/2; 0/5), 1oxb(−11/0; −9/1), 1qav(4/1; 5/3), 1rke(−10/6; −1/8), 1sko(−1/−1; 9/5), 1smp(−27/1; −15/3), 1spp(1/3; 3/3), 1svx(−14/−7; −1/2), 1ta3(2/−14; 13/−7), 1tnr(0/0; 7/7), 1u7e(5/−1; 13/0), 1ujw(−17/2; −6/5), 1vra(1/−17; 13/−13), 1wmh(−10/2; −8/6), 1wyw(0/−5; 7/−2), 1 × 3w(−3/−5; 4/−4), 1xb2(−6/−4; 8/10), 1xg2(2/−8; 4/−6), 1yc0(6/2; 17/9), 1ycs(6/−15; 13/−10), 1yvb(−10/1; −6/2), 1z3e(1/−2; 4/−2), 1zhh(−9/−12; −2/−4), 2a1j(−1/−3; 2/4), 2a73(−1/−21; 9/−3), 2b59(−5/−8; 1/0), 2blf(−1/−5; 3/−3), 2bov(0/15; 1/17), 2cg5(−4/−5; 10/−4), 2co6(1/0; 1/4), 2d74(0/1; 15/12), 2f2l(2/5; 4/12), 2fcw(−6/−8; 2/−7), 2ftx(−6/−1; −1/0), 2gy7(−2/−2; 2/13), 3ygs(−3/−1; 1/0).

**Protein domains (52):** 172l(−2/10; −1/10), 1a8l(−2/−8; −1/−6), 1b1b(−4/−2; −3/1), 1bk7(2/7; 6/8), 1bxo(−11/−12; −8/−12), 1c1k(6/4; 7/5), 1cfb(−3/−6; 0/−4), 1cid(−1/−2; −1/−2), 1g1t(−2/−2; −1/−1), 1ghr(7/−2; 9/3), 1h8l(−14/0; −6/1), 1ij9(−3/−6; −2/−3), 1inp(−8/−4; −7/3), 1iqq(4/3; 8/6), 1ko2(−9/−5; −5/0), 1lbu(0/0; 0/7), 1lj8(−17/−4; −4/−1), 1lq2(−1/−4; 6/−1), 1mb8(−5/7; −3/14), 1n81(−2/2; −1/3), 1nhy(7/−2; 7/−1), 1oha(−5/−1; −1/0), 1pii(−4/−7; 3/−3), 1qts(−1/0; −1/2), 1ra6(−10/−1; −2/3), 1ral(−4/6; 0/7), 1rc9(−7/2; −2/2), 1t2x(−2/3; 1/10), 1tje(−8/−1; −6/10), 1tzv(2/−3; 3/−2), 1u09(−12/0; 1/5), 1uv5(4/2; 5/8), 1vkw(1/0; 2/3), 1wd3(−26/−10; −22/−6), 1wmd(4/−5; 11/−5), 1x0x(2/−7; 4/−1), 1x8h(2/−1; 5/3), 1xbn(−4/7; −4/8), 1xeu(−1/−5; 1/−3), 1xt0(−2/−1; 0/0), 1y6i(−4/1; −3/5), 1yqe(−13/2; −7/5), 1yvj(1/−2; 4/2), 1yw5(2/−1; 3/4), 2bbc(8/3; 19/5), 2bh0(−3/6; −2/9), 2c9a(−8/6; −3/7), 2czr(−11/1; −5/4), 2fbo(−1/−5; 4/−4), 2gwl(−4/2; −4/5), 2plc(5/5; 8/6), 3lck(−1/−5; 1/−3).

## 3. Results and discussions

### 3.1. Total coulomb energies

Coulomb energies $E_{AB}$ between subunits were calculated for 102 homo-dimers, 60 hetero-dimers and 52 protein domains for two sets of charge sets, HIS-set and the HIP-set. Distributions of the

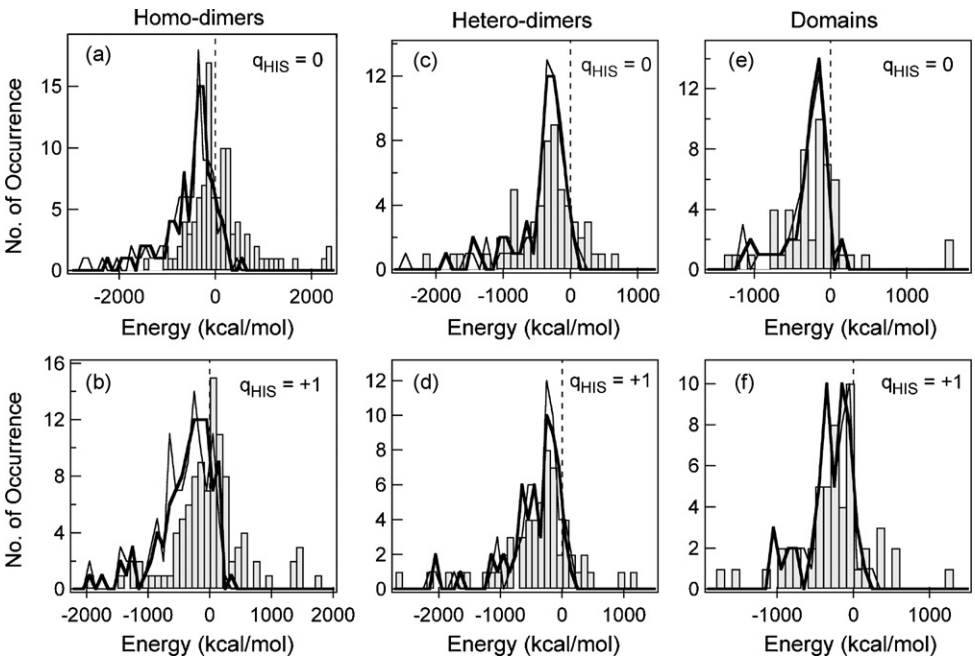


**Fig. 2.** Distributions (grey bars) of total Coulomb interaction energy $E_{AB}$ for (a) homo-dimers with neutral His, (b) homo-dimers with charged His, (c) hetero-dimers with neutral His, (d) hetero-dimers with positively charged His, (e) domains with neutral His, and (f) domains with positively charged His. The corresponding distributions of electrostatic interaction energies without the monopole–monopole term, are also shown: thin solid lines: ($E_{AB} - E_{MM}$), thick solid lines: ($E_{AB} - E_{MM0}$).

total Coulomb interaction energies, $E_{AB}$, between the two subunits, are shown in Fig. 2. For all the three categories, the homo-dimers, the hetero-dimers and the protein domains, and for both the charge sets, HIS-set and HIP-set, the energy distributions show a clear preference for negative energies, as was observed earlier [8]. However, an important aspect for all the three cases is the appearance of a positive energy tail. This is most evident for homo-dimers, followed by hetero-dimers and protein domains.

### 3.2. Monopole–monopole energies

Although subunit charge distributions are inherently complex, the distribution can be approximated to arise from a linear combination of simple charge distributions under multipole expansion–monopole (mp), dipole (dp), quadrupole (qp) and higher order terms. To understand if the positive tail in $E_{AB}$ distribution originates systematically from any one of these underlying terms, we separately estimated the contributions of the corresponding energy cross-terms to the total energy $E_{AB}$. Under multipole expansion, $E_{AB}$ can be approximated to be originating from a large number of inter-subunit cross-terms as shown (up to the quadrupole terms) schematically in Fig. 1b. Of all the cross-terms, only the monopole–monopole term, $E_{MM}$, is independent of the relative orientation of subunits in the dimeric-complex. The magnitude of the monopole depends on the net balance between acidic and basic residues in the sequence and the pH. In our calculation Lys and Arg always carry a net positive charge, Asp and Glu carry a net negative charge (pH 7), while His is neutral for the HIS-set and positive for the HIP-set. This gives rise to a distribution of net protein charges, both positive and negative, for the subunits for the two charge sets.

By definition, $E_{MM}$ is always positive for homo-dimers and so it is expected that without the monopole–monopole term, electrostatic energies will decrease. As seen in Fig. 2a (HIS-set) and Fig. 2b (HIP-set), the distribution of interaction energies shifts towards negative values when $E_{MM}$ is subtracted from the total Coulomb energy for homo-dimers. For hetero-dimers, the sign of $E_{MM}$ will depend on the polarity of total subunit charges (present dataset: opposite polarity: HIS-set: 23, HIP-set: 32; like polarity: HIS-set: 30, HIP-set: 21; zero charge on either subunit: HIS-set: 7, HIP-set: 7). Distribution of interaction energies for hetero-dimers, $E_{AB}$, without $E_{MM}$, also showed a clear shift towards negative energies with the disappearance of the positive tail (HIS-set: Fig. 2c and HIP-set: Fig. 2d). A similar shift was observed for protein domain data set as well (HIS-set: Fig. 2e and HIP-set: Fig. 2f). Strictly speaking, $E_{MM}$ represents the monopole–monopole term when the two subunits are far from each other. A more general expression for the monopole–monopole term is given by $E_{MM0}$. When $E_{MM0}$ was subtracted from $E_{AB}$, energy distributions without the new monopole–monopole term ($E_{AB} - E_{MM0}$), as indicated by the thick solid line in Fig. 2, were not much different from distributions of ($E_{AB} - E_{MM}$), the latter indicated by the thin solid line. Especially, the characteristic disappearance of the positive tail remained unchanged. Therefore, disappearance of the positive energy tail is a common feature of protein–protein complexes and is independent of the protonation state of histidine.

### 3.3. Interaction energies with monopole-free charge distribution

As shown in Fig. 1b, all possible cross-terms contribute to $E_{AB}$. Of these, only $E_{MM}$ is independent of inter-subunit orientation. All other terms depend on the relative subunit orientations. The contributions of terms other than the monopole term were estimated. These terms can be broadly classified into two groups: (1) MP–HP terms, arising from interaction between monopole of one subunit and dipoles and higher poles of the other subunit, and, (2) HP–HP terms, arising from interaction between dipoles and higher poles of one subunit and dipoles and higher poles of the other subunit. The energies associated with these two groups, $E_{MP-HP}$ and $E_{HP-HP}$, add up to yield the total Coulomb energy without the monopole–monopole energy: ($E_{MP-HP} + E_{HP-HP}$) = ($E_{AB} - E_{MM0}$). In order to estimate $E_{MP-HP}$ and $E_{HP-HP}$, we first calculated energies with monopole-devoid charge distributions on either or both subunits and obtained the corresponding energies $E_{A^*B}$, $E_{AB^*}$ and $E_{A^*B^*}$, where the cross-terms contributing to $E_{A^*B}$, $E_{AB^*}$ and $E_{A^*B^*}$ are shown in Fig. 1b (only up to quadrupole terms). As evident from

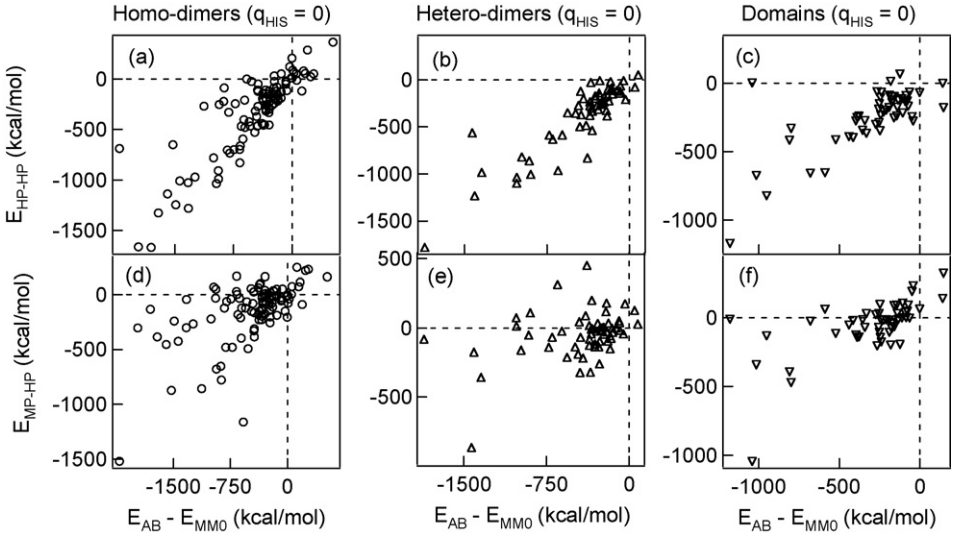


**Fig. 3.** Correlation between $E_{HP} - E_{HP}$ and $E_{AB} - E_{MM0}$ (neutral histidine charge set) for: (a) homo-dimers, (b) hetero-dimers, and, (c) domains. Correlation between $E_{MP} - E_{HP}$ and $E_{AB} - E_{MM0}$ (neutral histidine charge set) for: (a) homo-dimers, (b) hetero-dimers, and, (c) domains.

Fig. 1b, $E_{A^*B^*}$ is equal to $E_{HP-HP}$. $E_{MP-HP}$, on the other hand, can be expressed as $E_{MP-HP} = E_{A^*B} + E_{AB^*} - 2E_{A^*B^*}$ (see Fig. 1b).

Without the monopole–monopole term, $(E_{AB} - E_{MM0})$ was mostly negative for all the three sets (Fig. 2). As shown in Figs. 3(a–c) (HIS-set) and 4(a–c) (HIP-set), a very similar trend was observed for $E_{HP-HP}$ as well, especially for hetero-dimers and protein domains. In fact, $E_{HP-HP}$ and $(E_{AB} - E_{MM0})$ show a very strong correlation. However, unlike $E_{HP-HP}$, $E_{MP-HP}$ values were not consistently negative for the data set and the correlation with $(E_{AB} - E_{MM0})$ was also very weak, as shown in Figs. 3(d–f) (HIS-set) and 4(d–f) (HIP-set). Data presented in Figs. 3 and 4 clearly indicated: (1) inter-subunit interaction between dipole/higher poles in each subunit exhibit favorable electrostatic interactions, (2) $E_{HP-HP}$ is the dominant contributor in electrostatically favoring the native subunit orientations in dimeric-complexes, as indicated by the strong correlation of $E_{HP-HP}$ with $(E_{AB} - E_{MM0})$, (3) monopoles on a subunit may or may not exhibit favorable electrostatic interactions with the dipoles and higher poles of the other subunit. In fact when $E_{HP-HP}$ was plotted against $E_{MP-HP}$ (data not shown), no clear correlation was observed indicating that optimization of $E_{HP-HP}$ is mostly independent of $E_{MP-HP}$. We also estimated dipole–dipole interaction energies $E_{DP-DP}$. $E_{DP-DP}$ was almost equally distributed between positive and negative values and showed a rather weak correlation with $E_{HP-HP}$ (data not shown). This indicated that dipole–dipole terms do not consistently contribute towards optimizing $E_{HP-HP}$. The analyses presented here clearly show that $E_{HP-HP}$, in totality, is the primary contributor to negative Coulomb energies and plays a major role in optimizing native inter-subunit optimization.

### 3.4. Effect of solvent screened monopole–monopole electrostatic energies

The effect of removing the monopole–monopole energy term from Coulomb energy of interaction (Fig. 2) is by itself an important result and the main focus of this paper. However, it is also true that in solution Coulomb energies are screened by the solvent polarization through reaction field energies. Electrostatic energies were calculated by including the effect of solvent using a continuous dielectric model (no salt). The reaction field energies of complex formation ($\Delta\Delta G_{RF}$) for the hetero-dimeric data set were computed by numerically solving the Poisson–Boltzmann equa-

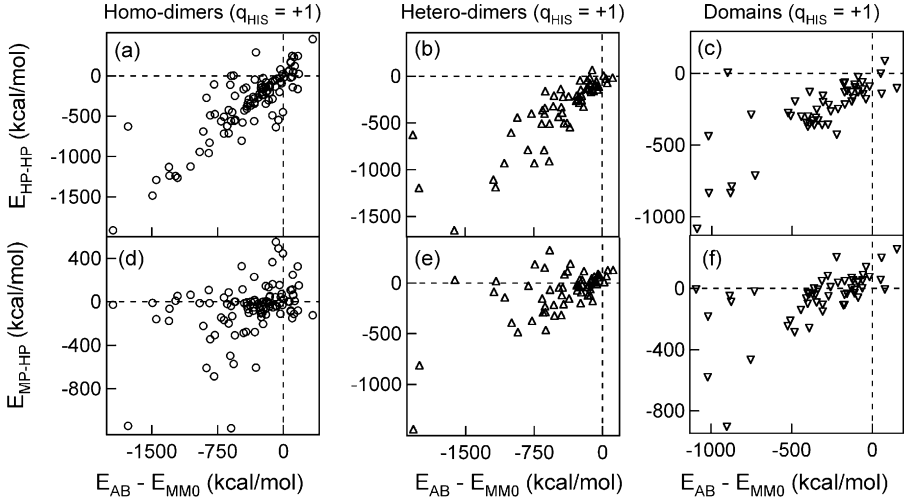


**Fig. 4.** Correlation between $E_{HP} - E_{HP}$ and $E_{AB} - E_{MM0}$ (positively charged histidine set) for: (a) homo-dimers, (b) hetero-dimers, and, (c) domains. Correlation between $E_{MP} - E_{HP}$ and $E_{AB} - E_{MM0}$ (positively charged histidine set) for: (a) homo-dimers, (b) hetero-dimers, and, (c) domains.

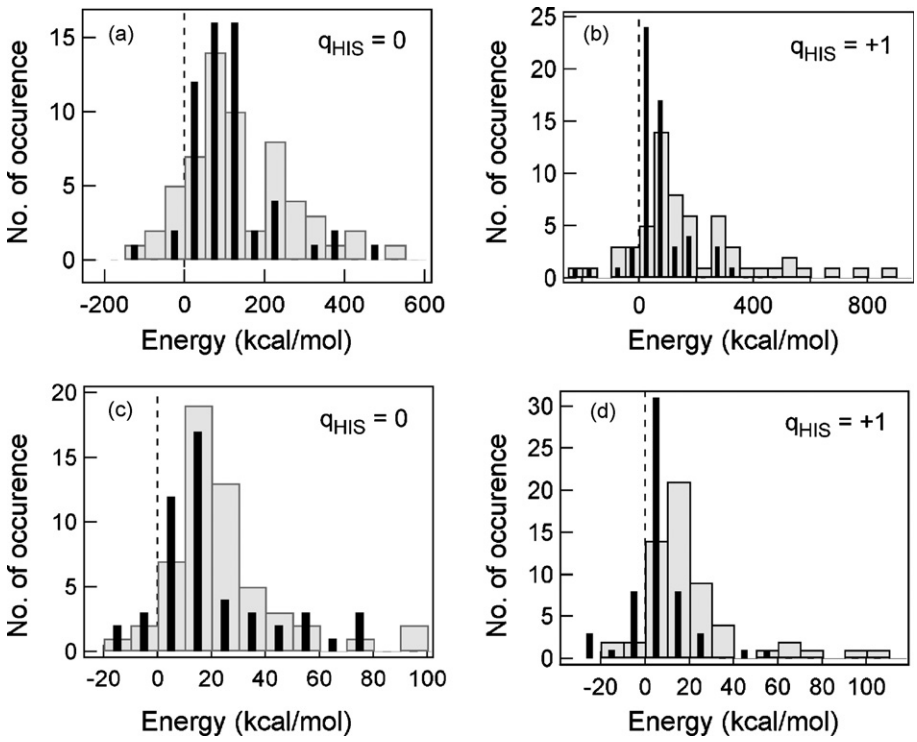


**Fig. 5.** Distribution of reaction field energies $\Delta\Delta G_{RF}$ for the hetero-dimeric set with neutral His (a) and charged His (b). Shaded histograms represent total reaction field energies $\Delta\Delta G_{RF}$ (total) while the black bars represent $\Delta\Delta G_{RF}$ (total) − $\Delta\Delta G_{RF}$ (mp–mp). Distribution of electrostatic component of binding free energies $\Delta\Delta G_{EL}$ (Coulomb plus $\Delta\Delta G_{RF}$) for the hetero-dimeric set with neutral His (c) and charged His (d). Shaded histograms represent the total electrostatic binding energies $\Delta\Delta G_{EL}$ (total) while the black bars represent $\Delta\Delta G_{EL}$ (total) − $\Delta\Delta G_{EL}$ (mp–mp).

tion [12]. The distribution of $\Delta\Delta G_{RF}$ for the hetero-dimeric data set in shown by the shaded histogram in Fig. 5a (HIS-set) and Fig. 5b (HIP-set). As expected for most cases $\Delta\Delta G_{RF}$ is positive, compensating the mostly negative Coulomb energies (Fig. 2). We also calculated the mp–mp component of $\Delta\Delta G_{RF}$ (mp–mp) by replacing all point charges on the protein by a single point charge placed at the geometric center of the subunits, with a magnitude equal to the total subunit charge. Similar to what was done for Coulomb energies, this mp–mp component was subtracted from $\Delta\Delta G_{RF}$ and the distribution of the resulting $\Delta\Delta G_{RF}$ difference is shown as filled bars in Fig. 5a (HIS-set) and Fig. 5b (HIP-set). The distributions in Fig. 5a and b showed that without the mp–mp term, the $\Delta\Delta G_{RF}$ distribution shifts towards lower energies. The total electrostatic components of binding energy, $\Delta\Delta G_{EL}$ ($\Delta\Delta G_{RF}$ + Coulomb energy), with and without the monopole contribution, were also calculated and their distributions are shown in Fig. 5c (HIS-set) and Fig. 5d (HIP-set). Without the monopole–monopole term, the distribution shifted towards lower energies. To assess the effect of salt, all energy calculations were repeated in the presence of 0.145 M mono-valent salt. Although increasing the salt concentration generally decreased the overall binding energies, as has been reported for a set of five proteins [14], the observed shift in $\Delta\Delta G_{RF}$ and $\Delta\Delta G_{EL}$ towards lower energies in the salt-free model remained unchanged in the presence of salt (data not shown).

The shift of electrostatic binding energies towards negative values upon removal of the monopole–monopole term is similar to what was observed for simple Coulomb energies. In other words, with the inclusion of solvent, using continuum dielectric models, the role of the monopole–monopole term remains qualitatively unaltered. However, it is worth noting that the physical interpretation of this effect, as arising from removal of orientation independent electrostatic energy terms, is not straightforward. The orientation dependence of the monopole–monopole screening term, in the presence of solvent, is different from the Coulomb monopole–monopole term. While the Coulomb monopole–monopole term is orientation independent, the monopole–monopole screening term is orientation independent only for spherically symmetric protein shapes like a perfect sphere. For realistic protein shapes, the geometric center can lie close to one side of the protein. If that side gets preferentially buried upon binding, there will be more desolvation penalty for the monopole than if the other side (change of orientation) gets buried. So, strictly speaking, removing the solvent screened monopole–monopole term also removes some orientation dependent effects from the total binding energy. Despite such complications, our results indicate that the qualitative effect of removing the monopole–monopole term from the total electrostatic energy remains unaltered even when solvation energies are included.

## 4. Summary and perspective

To summarize, we showed that the positive tail of the Coulomb energy distributions disappears when the monopole–monopole energy term is removed from the total Coulomb energy of homo-dimers, hetero-dimers and protein domains. Because the monopole–monopole term is the only term in the Coulomb energy expression that is independent of subunit orientation, this indicates that electrostatically unfavorable subunit orientations, that would have given rise to positive energies without the monopole–monopole term, are excluded in experimental protein–protein structures. Recently it was shown that the Coulomb energies were sequence-optimized [8], meaning that upon sequence permutation, the new Coulomb energies were always more positive in a statistically significant way. Our results show that Coulomb energies are orientation-optimized as well. Although

Coulomb energies include no solvent effect, calculations based on a piecewise continuous dielectric model showed that the main conclusion of this work, that electrostatically unfavorable subunit orientations are excluded in protein–protein complexes, remains unaltered with the inclusion of solvent using continuum dielectric models.

When two subunits come together to form a complex, the rate-limiting step is the formation of the encounter complex, leading to the transition state. Because the typical rates of formation of protein–protein complexes are much faster than what is expected from a model of encounter complex formation by random diffusion, there must be some built-in bias that discriminates against "wrong" orientations of the subunits in the encounter complex. It has been shown that electrostatic effects can enhance the diffusion-controlled protein–protein association rates for barstar–barnase interaction [15]. Assuming that the subunit orientations in crystal structures are very close to the corresponding transition states, our results show that the electrostatic restriction of subunit orientations in protein–protein encounter complexes [16], giving rise to the transition state, is a general feature of protein–protein complexes. The electrostatic control of choosing the correct over incorrect subunit orientations bear a similarity to electrostatic control of choosing a cognate ligand over a non-cognate ligand, when the cognate/non-cognate ligand pairs have very similar shapes but markedly different dipoles [16].

Orientational biases of the subunits, originating from electrostatics, can come from monopole–higher pole terms and higher pole–higher pole terms. We estimated both these contributions and showed that higher pole–higher pole terms were the dominant contributor. However, the higher pole–higher pole terms contributed as a whole. For example, the distributions of dipole–dipole interaction energies were spread over both positive and negative energies. This implies that development of algorithms for predicting native-like orientations of protein complexes using the overall charge distribution of protein subunits might be inadequate if only mutual orientations of subunit dipoles are considered. Rather, the mutual orientation of some composite quantity, representing dipoles and higher poles, may be more suitable.

## Acknowledgements

M.D. acknowledges financial support from DBT, India. G.B. acknowledges critical comments by Pinak Chakrabarti, Barun Chatterjee and Sibaji Raha.

## References

[1] B.A. Shoemaker, A.R. Panchenko, Deciphering protein–protein interactions. Part I. Experimental techniques and databases, PLoS Comput. Biol. 3 (2007) e42.
[2] B.A. Shoemaker, A.R. Panchenko, Deciphering protein–protein interactions. Part II. Computational methods to predict protein and domain interaction partners, PLoS Comput. Biol. 3 (2007) e43.
[3] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne, The protein data bank, Nucleic Acids Res. 28 (2000) 235–242.
[4] J. Janin, P. Rodier, P. Chakrabarti, R.P. Bahadur, Macromolecular recognition in the Protein Data Bank, Acta Crystallogr. D63 (2007) 1–8.
[5] D. Reichmann, O. Rahat, M. Cohen, H. Neuvirth, G. Schreiber, The molecular architecture of protein–protein binding sites, Curr. Opin. Struct. Biol. 17 (2007) 67–76.
[6] I.M.A. Nooren, J.M. Thornton, Diversity of protein–protein interactions, EMBO J. 22 (2003) 3486–3492.
[7] N. Froloff, A. Windemuth, B. Honig, On the calculation of binding free energies using continuum methods: application to MHC class I protein–peptide interactions, Protein Sci. 6 (1997) 1293–1301.
[8] K. Brock, K. Talley, K. Coley, P. Kundrotas, E. Alexov, Optimization of electrostatic interactions in protein–protein complexes, Biophys. J. 93 (2007) 3340–3352.
[9] F. Dong, H.-X. Zhou, Electrostatic contribution to the binding stability of protein–protein complexes, Proteins 65 (2006) 87–102.
[10] T.J. Dolinsky, J.E. Nielsen, J.A. McCammon, N.A. Baker, PDB2PQR: an automated pipeline for the setup of Poisson–Boltzmann electrostatics calculations, Nucleic Acids Res. 32 (2004) W665–W667.
[11] W.D. Cornell, P. Cieplak, C.I. Bayly, I.R. Gould, K.M. Merz Jr., D.M. Ferguson, D.C. Spellmeyer, T. Fox, J.W. Caldwell, P.A. Kollman, A second generation force field for the simulation of proteins, nucleic acids, and organic molecules, J. Am. Chem. Soc. 117 (1995) 5179–5197.
[12] A. Nicholls, A. Honig, A rapid finite difference algorithm, utilizing successive over-relaxation to solve the Poisson-Boltzmann equation, J. Comput. Chem. 12 (1991) 435–445.
[13] F. Rodier, R.P. Bahadur, P. Chakrabarti, J. Janin, Hydration of protein–protein interfaces, Proteins 60 (2005) 36–45.
[14] C. Bertonati, B. Honig, E. Alexov, Poisson-Boltzmann calculations of nonspecific salt effects on protein–protein binding free energies, Biophys. J. 92 (2007) 1891–1899.
[15] M. Vijayakumar, K.-W. Wong, G. Schreiber, A.R. Fersht, A. Szabo, H.-X. Zhou, Electrostatic enhancement of diffusion-controlled protein–protein association: comparison of theory and experiment on barnase and barstar, J. Mol. Biol. 278 (1998) 1015–1024.
[16] G. Basu, D. Sivanesan, T. Kawabata, N. Go, Electrostatic potential of nucleotide-free protein is sufficient for discrimination between adenine and guanine-specific binding sites, J. Mol. Biol. 342 (2004) 1053–1066.