Improving Electrostatics Description In Scoring Functions - Insights For Their Role For Drugs

Dmitry A. Shulga, Arslan R. Shaimardanov, Nikita N. Ivanov, Vladimir A. Palyulin

Department of Chemistry at Moscow State University, Moscow, Russia

XXIX Symposium on Bioinformatics and
Computer-Aided Drug Discovery

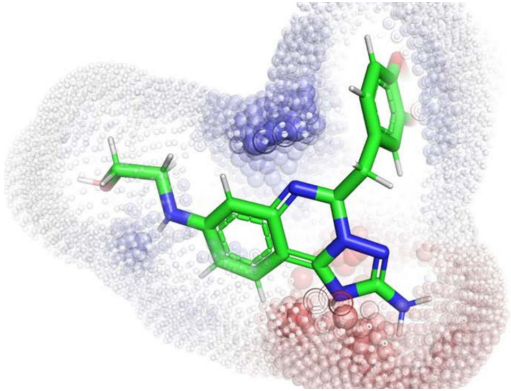
2023

On importance of electrostatic interactions

1. physics: electrostatic interactions are important
2. wide practical application within the force fields (FF) and scoring functions (SF) confirms it
3. so accurate electrostatics seem to be crucial for SFs…

DENR - a platform for partial charge calculation

1. we have developed charges for long time…
2. our current and perhaps the best project is empirical method based on the Dynamic ElectroNegativity Relaxation (DENR) principle (incomplete EN equalization) [1]
3. our target metrics is quantitative reproduction of the QC reference MEP by DENR charges, so the charges are widely applicable by design
4. practical benefits of DENR
    a. reasonable QC MEP reproduction
    b. empirical foundations - the model is rather “rough”, so the chances to overfit are low
    c. fast calculation (compared to semi-empirical and ab initio QC methods)
    d. graph based - no need for 3D geometry
5. our plan (was)
    a. apply our charges in SF and finally in docking
    b. we need to first check how it performs



1. Shulga, D. A. et al // *SAR and QSAR in Env. Res.*, **2008**, *19*(1-2), 153-165.

Experiment design: scoring function (SF)

AutoDock 4.2

1. force-field (FF) (or physics) based SF
2. widely used and freely accessible
3. already uses Gasteiger partial charges to describe electrostatic part of interactions
4. there are two papers in which better quality MEP reproductions charges were tested in combination with AD4.2
   a. RESP (gold standard), AM1-BCC (practical standard) [1]
   b. several known charge methods [2]
5. in common
   a. promising results reported for QC MEP related methods (RESP and AM1-BCC) [1]
   b. …but low throughput and high complexity are the price

1. Wang J. C. et al. // JCIM, **2011**, 51, 2528-2537.
2. Hou X. et al. // JCIM, **2013,** 53, 188-200.

Experiment design: ligand-receptor complex set

1. ligand-receptor complex set - core set from CASF 2016 Update [1] - 285 structures
2. obtain the best linear model for each charge set (ligand and receptor)
3. AutoDock4.2 score:
4. task: obtain new set of Wj for each charge scheme

$$\Delta G_{\text{bind}} = W_{\text{vdW}} \times \sum_{i,j} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} \right) + W_{\text{H-bond}} \times \sum_{i,j} E(t) \left(\frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}} \right) + W_{\text{estat}} \times \sum_{i,j} \frac{q_i q_j}{\varepsilon(r_{ij}) r_{ij}} + W_{\text{desol}} \times \sum_{i,j} \left(S_i V_j + S_j V_i \right) \mathrm{e}^{\left(- r_{ij}^2 / 2\sigma^2 \right)} + W_{\text{tor}} \times N_{\text{tors}}$$

NB: Desolvation values taken using Gasteiger charges for all other charges

1. Su M. et al. // JCIM, **2018**, 59, 895-913.

Experiment design: design

L-R complexes

CASF 2016 core set

ΔG_{ref} (K_i, K_d)

Receptors

Ligands

Calculate charges

1. **Gasteiger** (default for AD4.2)
2. Kollman
3. Formal charges (fc)

1. **Gasteiger** (default for AD4.2)
2. **DENR** (***main test***)
3. MMFF94
4. AM1-BCC
5. RESP
6. Formal charges (fc)
7. …

$$\Delta G_{\text{bind}} = W_{\text{vdW}} \times \sum_{i,j} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} \right) + W_{\text{H-bond}} \times \sum_{i,j} E(t) \left(\frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}} \right) + W_{\text{estat}} \times \sum_{i,j} \frac{q_i q_j}{\varepsilon(r_{ij}) r_{ij}} + W_{\text{desol}} \times \sum_{i,j} (S_i V_j + S_j V_i) \, e^{(-r_{ij}^2 / 2\sigma^2)} + W_{\text{tor}} \times N_{\text{tors}}$$

Final stats

Train Wj model

New values for Ej

Experiment aim in short

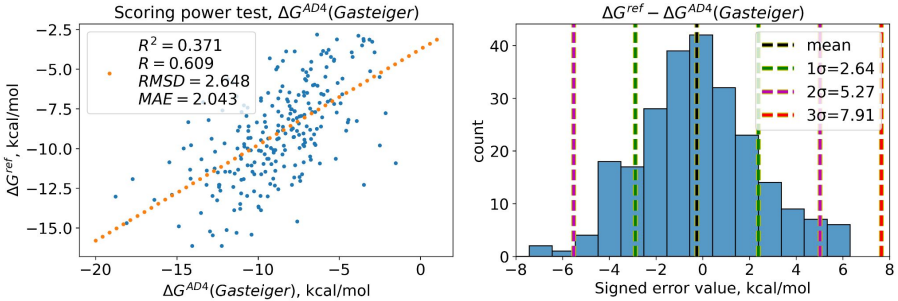
Study whether replacement of Gasteiger charges (default in AutoDock4.2) with our DENR charges (with parameters fit to QC MEP) will brings benefits (which we expect)

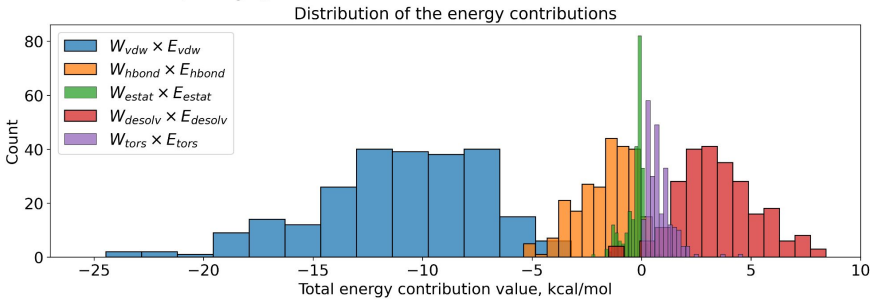
All in fair conditions!

Results

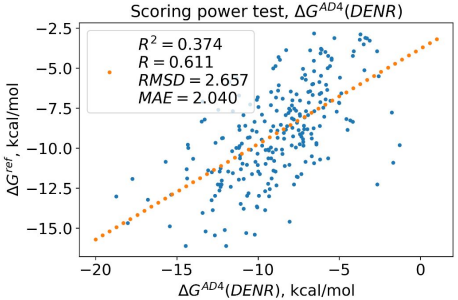
The main outcome

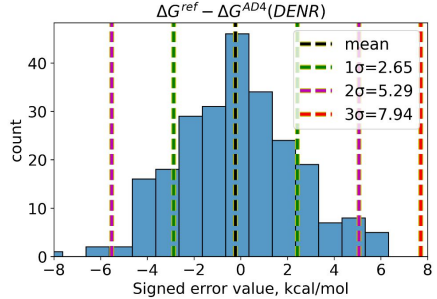
Gasteiger (L) - Gasteiger (R)

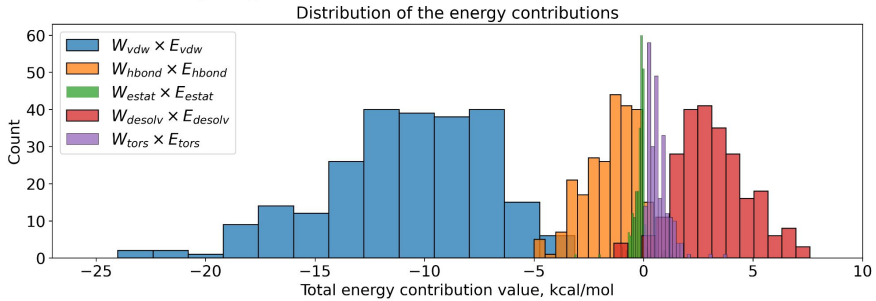




DENR (L) - DENR (R)



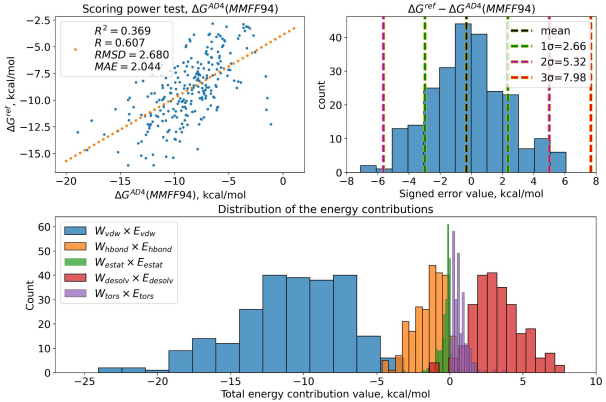




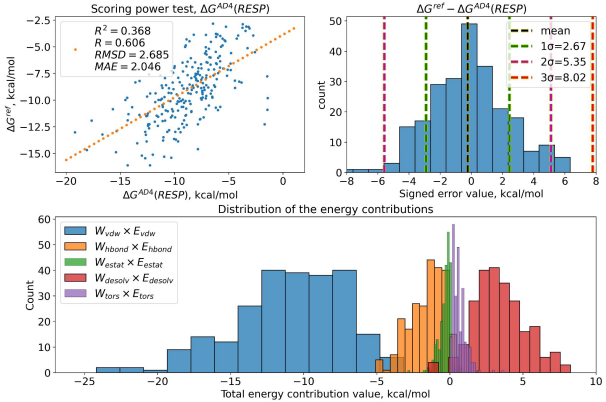
Improvement is negligible!

The main outcome

MMFF94 (L) - MMFF94 (R)

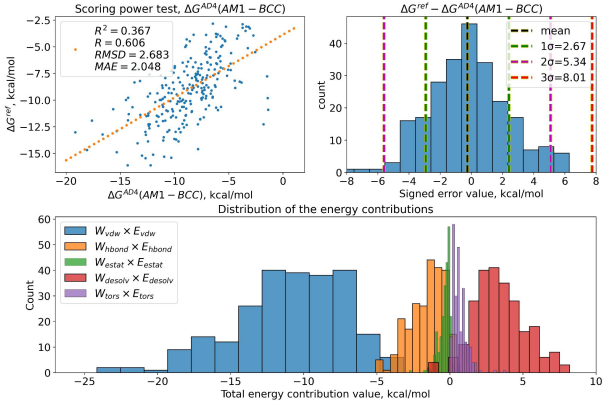


RESP (L) - Kollman (R)

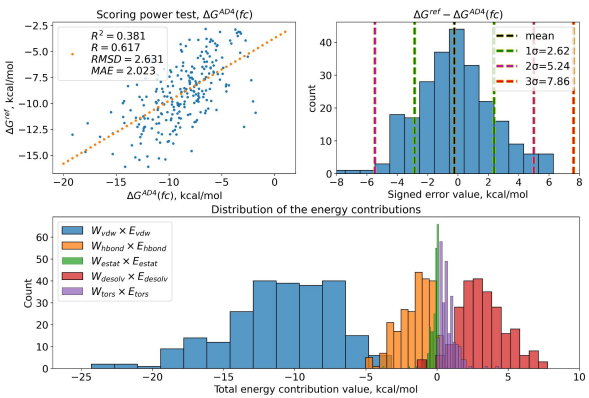


!!

AM1-BCC (L) - Kollman (R)



Formal q (L) - Formal q (R)



No difference

The questions arising

1. technical
    a. a purely empirical HB term inside physics (FF) based SF - possibly duplicating electrostatics
    b. desolvation is very crude - there’s room for improvement
2. conceptual
    a. is electrostatics important at all? when we consider small molecule - receptor interactions in solution?

Electrostatics in scoring ligand-receptor interactions

dilemma

a. on one hand,
    i. in (vacuum) physics E(electrostatic) is a dominating and long range interaction
    ii. it is omnipresent in practice - there must be firm reason for it!
b. on the other hand,
    i. E(electrostatic) contribution is negligible in FF-based score (in energy terms)
    ii. wide variation of charges - different E(electrostatic) quality - does not affect the outcome much

Electrostatics in scoring ligand-receptor interactions

further thoughts

a. main driving forces of interactions
   i. hydrophobic interactions
   ii. shape complementarity
b. decoration: HB as directed interactions
c. BUT!
   i. thermodynamic effect of HB is highly questionable on affinity!
   ii. L · *N*(H2O) + R · *M*(H2O) —> L-R (*K* HBs) + (*M*+*N*) · H2O, ***K* << *M*+*N*** (geometrically)

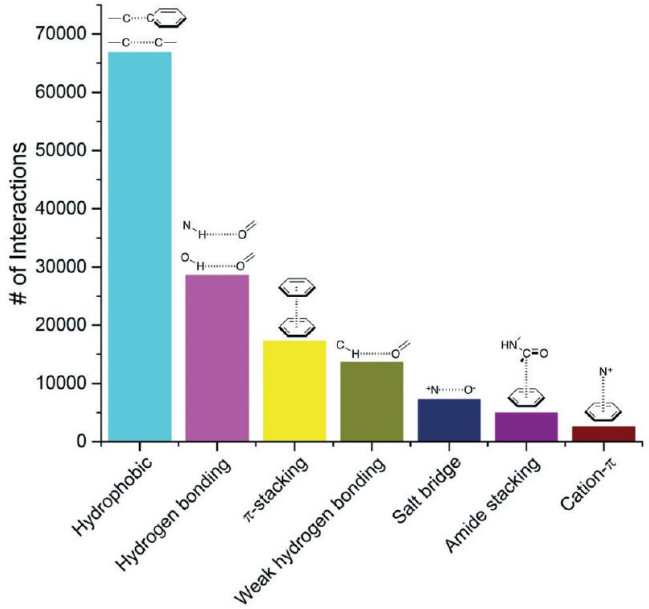


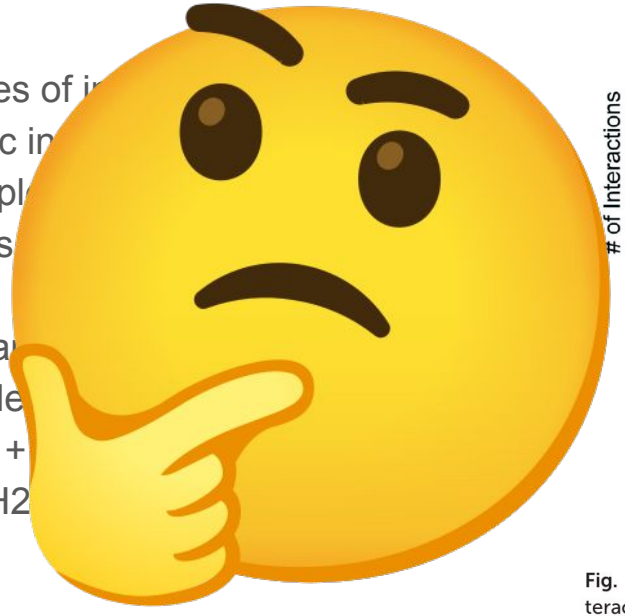
Fig. 1 Frequency distribution of the most common non-covalent interactions observed in protein–ligands extracted from the PDB.

Taken from Med. Chem. Commun., 2017, 8, 1970

Electrostatics in scoring ligand-receptor interactions

further thoughts

a. main driving forces of i
   i. hydrophobic in
   ii. shape compl
b. decoration: HB as
c. BUT!
   i. thermodyna
   questionable
   ii. L · *N*(H2O) +
   + (*M*+*N*) · H2



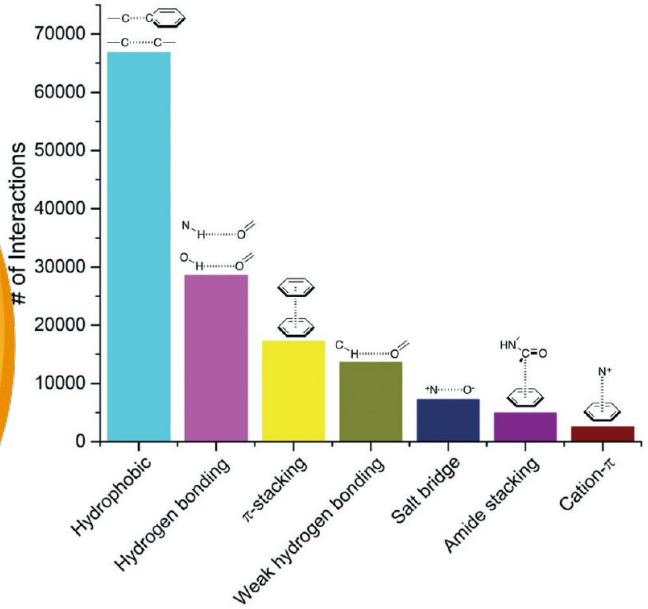
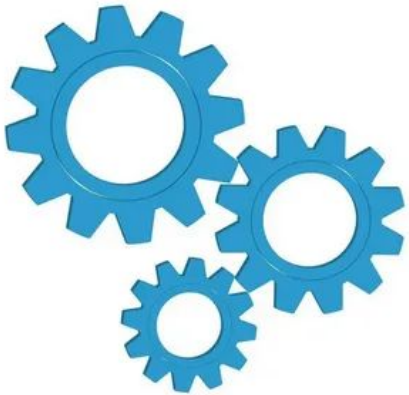


Fig. 1 Frequency distribution of the most common non-covalent interactions observed in protein–ligands extracted from the PDB.

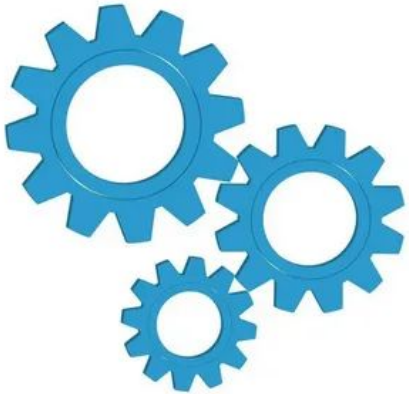
Taken from Med. Chem. Commun., 2017, 8, 1970

What we actually seek for?



Highly specialized tool
Ideal for *just* L-R affinity prediction!

What we actually seek for?



VS

Highly specialized tool
Ideal for *just* L-R affinity prediction!

Versatile tool
Good enough for prediction of *overall* success ligand -> drug!

Idea elaboration: HBs beyond affinity in SFs

Hypothesis

a. HBs and electrostatics serve not for the affinity reasons! But rather for overall “success”

Reasons

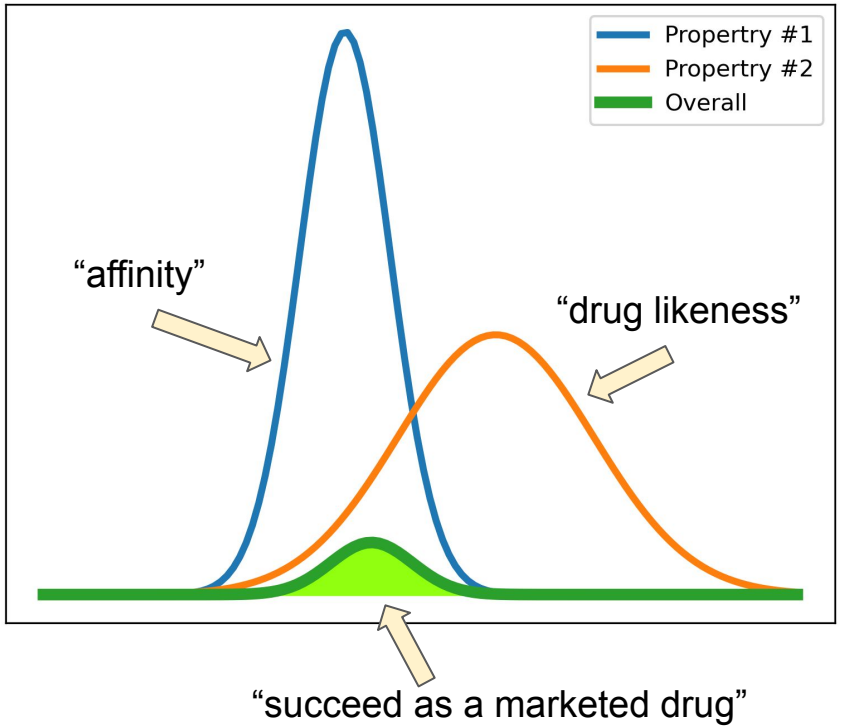
a. datasets (PDBBind, etc.) for SF development contain (almost) drugs as positive examples -> other properties crucial for drugs were taken indirectly into account = “Survivorship bias”
b. presence of HBs (HBD, HBA) in L-R complex
    i. directed interactions, demand for complementarity - off-target selectivity
    ii. if in desired range (e.g. Ro5) - favorable ADMET properties
    iii. improved synthesizability

Idea elaboration: HBs beyond affinity in SFs

Graphic illustration

(Maximum likelihood principle)

To get a drug one must obtain both high affinity and good drug-like properties…



Claims and consequences

Reasons

a. Claim: most contemporary scoring functions (SFs) measure not only pure **affinity**, but rather a more complex function estimating “**ligand developability**”

b. Corollary: these different contributions should be explicitly split and consciously dealt with
   i. sub-corollary: directed electrostatic interactions and hydrogen bonds (HBs) are more relevant to estimate the relevant ADMET properties, rather than the affinity per se (where it seems to be useless!)

Thank you!