

Universal short-range *ab initio* atom–atom potentials for interaction energy contributions with an optimal repulsion functional form

Jan K. Konieczny¹ · W. Andrzej Sokalski¹

Received: 10 May 2015 / Accepted: 8 June 2015 / Published online: 14 July 2015
© Springer-Verlag Berlin Heidelberg 2015

Abstract The repulsion term in conventional force fields constitutes a major source of error. Assuming that this could originate from a too simple analytical functional form, we analyzed various analytical functions using *ab initio* exchange component values as a reference and obtained $(\alpha + \beta R^{-1}) \exp(-\gamma R)$ as the optimal form to represent the repulsion term. Universal exchange, delocalization, and electrostatic penetration potentials approximating the corresponding interaction energy components defined within hybrid variation-perturbation theory (HVPT) were derived using as a reference a training set of 660 biomolecular complexes. The electrostatic multipole term was calculated using cumulative atomic multipole moments, whereas correlation contribution including dispersion term and first-order correlation correction was estimated from nonempirical D_{as} functions derived by Pernal et al. The resulting non-empirical atom–atom potentials (NEAAP) were tested for several urokinase–inhibitor complexes yielding improved docking results.

Keywords Intermolecular interactions · Nonempirical potentials · Exchange repulsion · Optimal analytical function

Introduction

Precise knowledge of intermolecular interactions between proteins and their ligands is of critical importance in rational drug or biocatalyst design. Experimental studies yield still very limited information in this respect, while rigorous *ab initio* calculations are prohibitively expensive, due to their steep scaling with basis set size. Therefore, simulations of large biomolecular systems are usually based on empirical force fields (FF), which exhibit limited transferability between molecular systems mainly due to the electrostatic term [1–3] and do not preserve any clear physical meaning of its particular non-bonded contributions. Historically, first conventional non-bonded force field parameters have been fitted to reproduce experimentally available data such as enthalpies of vaporization or sublimation, however they usually differ considerably between themselves. Alternative and more reliable source of data for FF fitting constitute accurate quantum chemical interaction energies, which may cover much wider range of intermolecular distances. Therefore, such FF could be called non-empirical. To our knowledge, only a few attempts have been made to derive universal [1, 4] FF nonbonded energy terms directly by fitting interaction energy components as defined by perturbation theory. Corresponding system-specific potential functions were much more frequent [5–8]. In this contribution, we focus on improving the functional form for the exchange repulsion term, which has been noted as a major

This paper belongs to Topical Collection 6th conference on Modeling & Design of Molecular Materials in Kudowa Zdrój (MDMM 2014)

✉ W. Andrzej Sokalski
sokalski@pwr.edu.pl

¹ Department of Chemistry K1/W3, Wrocław University of Technology, Wyb. Wyspiańskiego 27, 50-370 Wrocław, Poland

Table 1 Root mean square errors (RMSE) for various functional forms approximating the exchange interaction energy $E_{EX}^{(10)}$ (in [kcal/mol])

R/R_{eq}	0.70	0.90	0.95	1.00	1.05	1.10	1.25	TOT
αR^{-12}	24.51	7.87	5.84	4.30	3.15	2.32	0.90	11.13
$\alpha \exp(-\gamma R)$	9.01	2.36	1.92	1.57	1.32	1.15	0.70	3.93
$(\alpha + \beta R) \exp(-\gamma R)$	7.85	2.34	1.91	1.53	1.21	0.58	0.42	3.55
$(\alpha + \beta R^{-1}) \exp(-\gamma R)$	7.73	2.12	1.62	1.26	0.98	0.46	0.34	3.46
$(\alpha + \beta R^{-2}) \exp(-\gamma R)$	7.03	2.29	1.76	1.36	1.05	0.60	0.37	3.27
$(\alpha + \beta R^{-1} + \delta R) \exp(-\gamma R)$	7.18	2.18	1.70	1.35	1.09	0.66	0.47	3.28
$(\alpha + \beta R^{-1} + \delta R^{-2}) \exp(-\gamma R)$	6.80	2.23	1.74	1.39	1.13	0.74	0.49	3.19
$(\alpha + \beta R^{-1} + \delta R^{-2} + \kappa R^{-3}) \exp(-\gamma R)$	6.59	2.29	1.84	1.50	1.22	0.76	0.48	3.13
$(\alpha + \beta R^{-1} + \delta R + \kappa R^{-2}) \exp(-\gamma R)$	6.85	2.21	1.72	1.39	1.14	0.77	0.55	3.20
$(\alpha + \beta R^{-1} + \delta R + \kappa R^2) \exp(-\gamma R)$	6.91	2.27	1.82	1.49	1.21	0.81	0.56	3.19
$(\alpha + \beta R^{-1} + \delta R + \kappa R^2 + \omega R^{-2}) \exp(-\gamma R)$	6.59	2.28	1.86	1.55	1.29	0.93	0.63	3.10

Bold entries indicate minimal error values for particular distances illustrating the range of validity of the main conclusion i.e. choice of (a+b/R) expression

source of error in conventional force fields [9]. We have already encountered this issue when studying inhibitors docked into urokinase, where some inhibitors were docked by a force field 0.5 Å too deep into the active site compared to ab initio results [10]. One of the reasons for this could be an inappropriate functional form for the repulsion term, usually represented by αR^{-12} or $\alpha \exp(-\beta R)$, whereas the analytical formula for the first order exchange term involves squared overlap integrals, composed in turn as products of complex polynomials and exponential functions [11, 12].

With this in mind, we here explore in a systematic way the various possible analytical formulas for reproducing the first-order exchange component for large number of biomolecular complexes and determine their optimal parameters. To facilitate applications in conventional software, we limit ourselves to products of polynomial and

exponential functions resembling the above mentioned, since they can be easily incorporated into typical force fields.

This work is an extension of our earlier report [1], in which perhaps one of the first universal non-empirical atom-atom potentials were derived for main interaction energy components. These were calculated in a minimal valence basis set for 336 hydrogen-bonded dimers and tested on the packing of N_2 , CO_2 , and nitromethane crystals [13]. The present contribution is based on an extended aug-cc-pVDZ basis set and a much wider collection of 660 biomolecular complexes, including various types of interactions besides hydrogen bonds. We tested the resulting nonempirical atom-atom potentials on the mentioned urokinase-inhibitor complexes, which exhibited artefact structures when optimized with a standard Tripos 5.2 force field [10].

Table 2 Mean unsigned error (MUE) for various analytic functions approximating exchange interaction energies $E_{EX}^{(10)}$ (in [kcal/mol])

R/R_{eq}	0.70	0.90	0.95	1.00	1.05	1.10	1.25	TOT
αR^{-12}	17.91	7.25	5.32	3.79	2.66	1.85	0.60	7.15
$\alpha \exp(-\gamma R)$	6.89	1.85	1.52	1.24	1.02	0.85	0.46	2.39
$(\alpha + \beta R) \exp(-\gamma R)$	6.22	1.55	1.18	0.91	0.70	0.44	0.22	2.04
$(\alpha + \beta R^{-1}) \exp(-\gamma R)$	6.02	1.50	1.09	0.80	0.58	0.36	0.16	1.96
$(\alpha + \beta R^{-2}) \exp(-\gamma R)$	5.55	1.60	1.16	0.84	0.63	0.45	0.22	1.96
$(\alpha + \beta R^{-1} + \delta R) \exp(-\gamma R)$	5.65	1.52	1.13	0.89	0.72	0.53	0.33	1.96
$(\alpha + \beta R^{-1} + \delta R^{-2}) \exp(-\gamma R)$	5.31	1.56	1.18	0.94	0.76	0.59	0.36	1.96
$(\alpha + \beta R^{-1} + \delta R^{-2} + \kappa R^{-3}) \exp(-\gamma R)$	5.19	1.51	1.16	0.95	0.80	0.60	0.33	1.91
$(\alpha + \beta R^{-1} + \delta R + \kappa R^{-2}) \exp(-\gamma R)$	5.38	1.52	1.15	0.93	0.78	0.62	0.43	1.97
$(\alpha + \beta R^{-1} + \delta R + \kappa R^2) \exp(-\gamma R)$	5.46	1.60	1.28	1.04	0.83	0.63	0.40	1.97
$(\alpha + \beta R^{-1} + \delta R + \kappa R^2 + \omega R^{-2}) \exp(-\gamma R)$	5.27	1.60	1.29	1.07	0.90	0.73	0.48	1.97

Bold entries indicate minimal error values for particular distances illustrating the range of validity of the main conclusion i.e. choice of (a+b/R) expression

Table 3 Mean unsigned relative error (MURE) for various analytic functions approximating exchange interaction energies $E_{EX}^{(10)}$ (in [%])

R/R_{eq}	0.70	0.90	0.95	1.00	1.05	1.10	1.25	TOT
αR^{-12}	24.3	47.9	52.0	54.6	56.1	56.6	53.5	44.6
$\alpha \exp(-\gamma R)$	9.4	12.8	15.3	18.3	21.8	26.3	44.2	17.9
$(\alpha + \beta R) \exp(-\gamma R)$	8.9	10.8	12.2	13.7	15.5	16.9	23.8	12.9
$(\alpha + \beta R^{-1}) \exp(-\gamma R)$	8.7	10.1	11.1	12.1	13.1	13.7	17.7	11.3
$(\alpha + \beta R^{-2}) \exp(-\gamma R)$	8.2	10.8	11.6	12.5	13.9	15.3	22.2	12.2
$(\alpha + \beta R^{-1} + \delta R) \exp(-\gamma R)$	8.2	10.1	11.3	13.7	17.5	22.6	55.5	16.6
$(\alpha + \beta R^{-1} + \delta R^{-2}) \exp(-\gamma R)$	7.8	10.3	11.7	14.4	18.4	23.7	55.9	16.9
$(\alpha + \beta R^{-1} + \delta R^{-2} + \kappa R^{-3}) \exp(-\gamma R)$	7.7	10.0	11.7	14.5	18.4	22.8	43.8	15.5
$(\alpha + \beta R^{-1} + \delta R + \kappa R^{-2}) \exp(-\gamma R)$	7.9	10.0	11.3	14.3	18.8	25.8	70.9	18.6
$(\alpha + \beta R^{-1} + \delta R + \kappa R^2) \exp(-\gamma R)$	7.9	10.6	13.1	16.5	21.0	27.4	59.7	18.2
$(\alpha + \beta R^{-1} + \delta R + \kappa R^2 + \omega R^{-2}) \exp(-\gamma R)$	7.7	10.5	13.1	16.7	21.9	28.7	62.0	18.6

Bold entries indicate minimal error values for particular distances illustrating the range of validity of the main conclusion i.e. choice of (a+b/R) exp(-gR) expression

Interaction energy components

The only rational way to determine non-empirical atom-atom potentials (NEAAP) is to partition the intermolecular interaction energy ΔE into some well-defined components, which can separate system-specific terms (for example, first-order electrostatic) from other more transferable contributions with different distance dependence requiring different analytical representation. There are many possible interaction energy partitioning methods, mostly related to variational Morokuma scheme [14] or symmetry adapted perturbation theory (SAPT) [15]. On the other hand, non-bonded interactions can be explained in an elegant way

using Hellmann–Feynman theorem [16]. Unfortunately, the Hellmann–Feynman theorem has not been practically demonstrated to provide simple potential functions yet. In this study, we have applied the hybrid variation-perturbation theory (HVPT) [17, 18], in which the total interaction energy for systems containing over 1800 AOs [19] can be partitioned into the following contributions:

$$\Delta E = E_{EL,MTP}^{(10)} + E_{EL,PEN}^{(10)} + E_{EX}^{(10)} + E_{DEL}^{(R0)} + E_{CORR},$$

The first-order multipole electrostatic term $E_{EL,MTP}^{(10)}$ is obtained using SCF monomer cumulative atomic multipole

Fig. 1 Optimal analytic function for exchange energy at equilibrium geometry

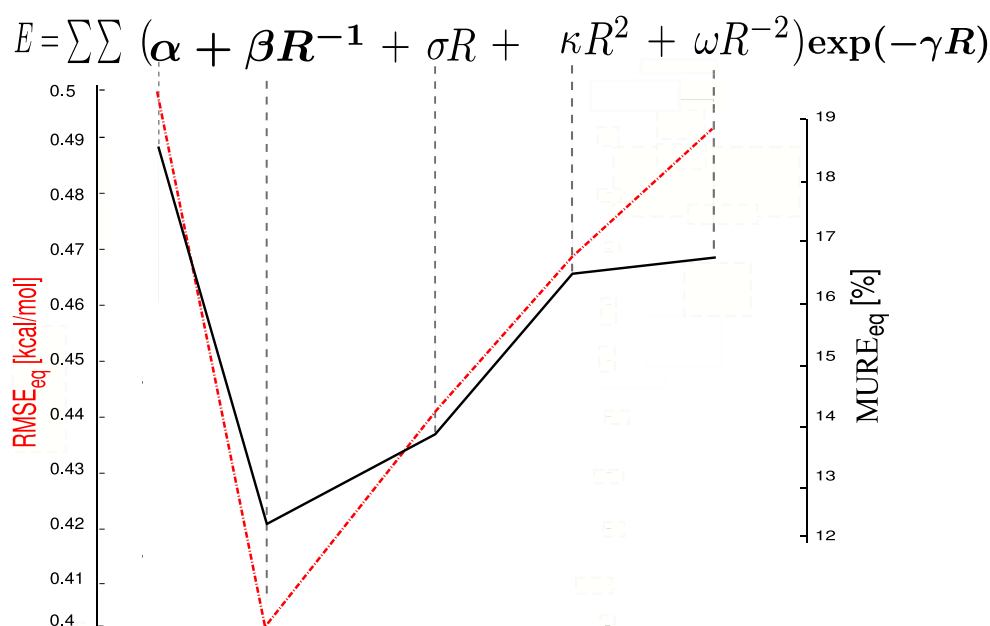
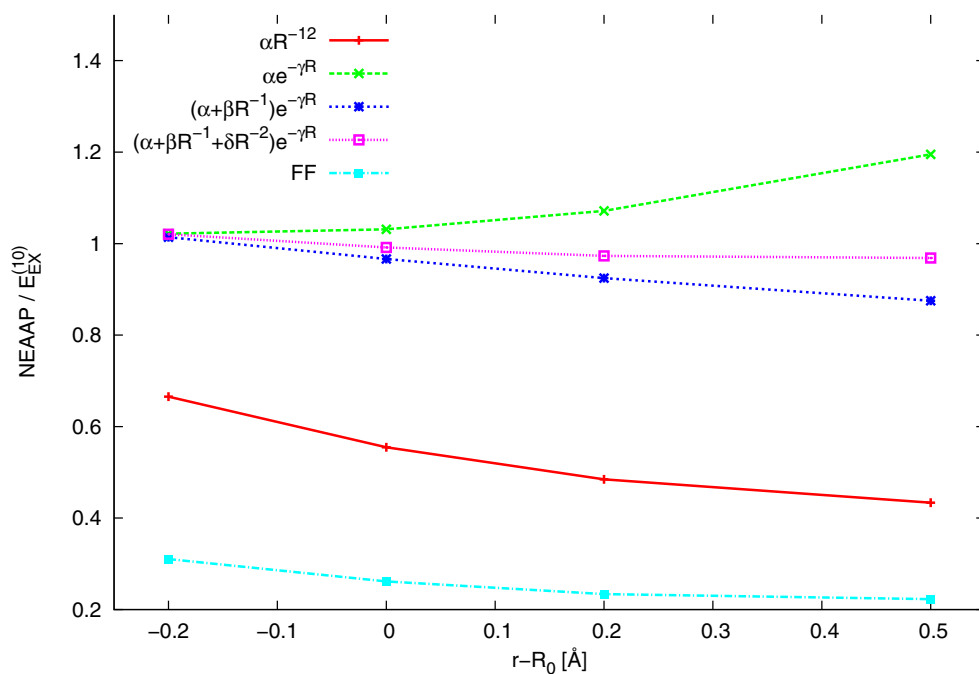


Fig. 2 The distance dependence of the ratio of various functions approximating the repulsion term and exact reference exchange term $E_{EX}^{(10)}$, for the methanol dimer



moments (CAMMs) $M_A^{(k_a)}$ and $M_B^{(k_b)}$ for all atom pairs of A and B molecules [20]:

$$E_{EL, MTP}^{(10)} = \sum_a^A \sum_b^B \sum_{k_a} \sum_{k_b} M_A^{(k_a)} [k_a] T^{(k_a+k_b)} [k_b] M_B^{(k_b)} \\ k_a + k_b \leq L$$

where $M^{(k)}$ is a rank k multipole and $T^{k_a+k_b}$ is the Cartesian interaction tensor containing all the partial derivatives of $|R_{ab}|^{-1}$ of rank $k_a + k_b$. In this contribution, the CAMM expansion was truncated at the R^{-5} term (rank 5) yielding best convergence [21] and calculated using the GAMESS-US ab initio package (activated by adding $\$ELMOM IAMM=n$ $\$END$ to the input file, where n is

Fig. 3 The distance dependence of the ratio of various functions approximating repulsion term and exact reference exchange term $E_{EX}^{(10)}$ values for acetate-methanol complex

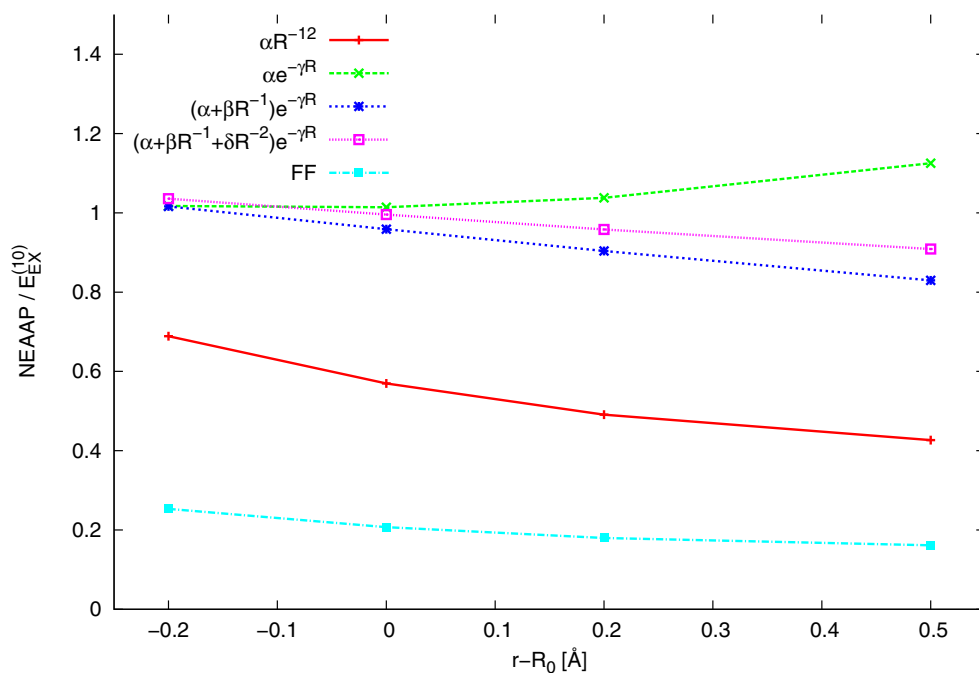
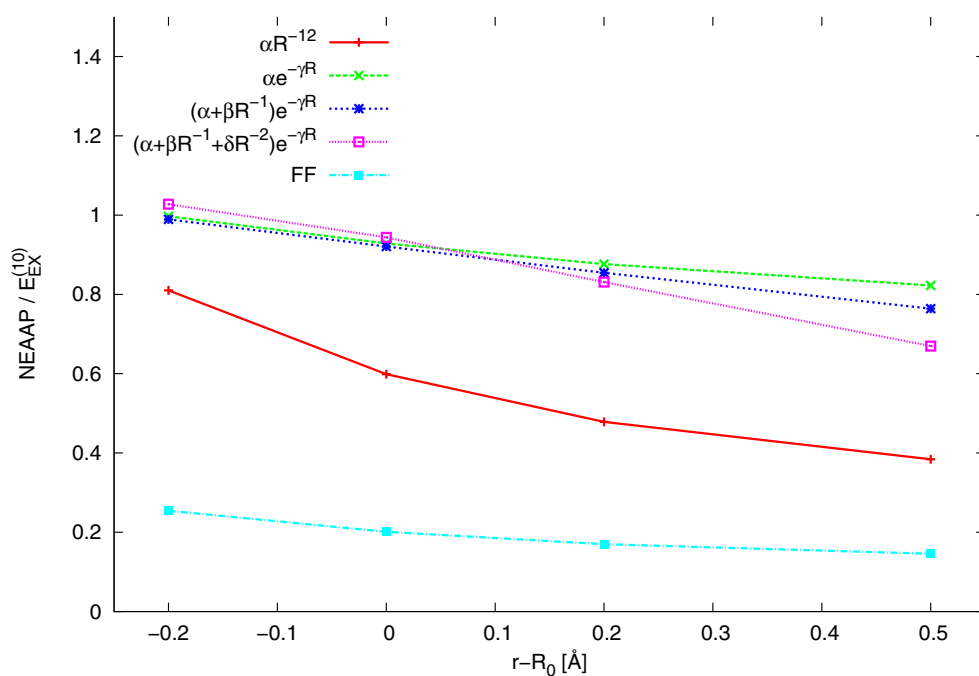


Fig. 4 The distance dependence of the ratio of various functions approximating repulsion term and exact reference exchange term $E_{EX}^{(10)}$ values for methylamine-methanol complex



the desired highest rank) [22]. Higher atomic multipole moments can easily be transformed into alternative point charge models [23, 24].

The complete first-order electrostatic term $E_{EL}^{(10)}$ is calculated in the dimer basis set as the first-order perturbational correction within the polarization approximation [17] and is equivalent to the analogous term defined within Symmetry Adapted Perturbation Theory (SAPT) [15].

$$\begin{aligned}
 E_{EL}^{(10)} &= \langle \Psi_A \Psi_B | \hat{H}_{AB} - \hat{H}_A - \hat{H}_B | \Psi_A \Psi_B \rangle / \langle \Psi_A \Psi_B | \Psi_A \Psi_B \rangle \\
 &= \sum_a^A \sum_b^B Z_a Z_b R_{ab}^{-1} + \sum_r^{AB} \sum_s^{AB} \sum_t^{AB} \sum_u^{AB} D_{rs}^A(D) D_{tu}^B(D) \langle rs | tu \rangle \\
 &+ \sum_r^{AB} \sum_s^{AB} \sum_b^B D_{rs}^A(D) \langle r | Z_b R_{1b}^{-1} | s \rangle \\
 &+ \sum_t^{AB} \sum_u^{AB} \sum_a^A D_{tu}^B(D) \langle t | Z_a R_{1a}^{-1} | u \rangle
 \end{aligned}$$

where monomer electron densities $D_{rs}^A(D)$, $D_{tu}^B(D)$ have been obtained in dimer basis set $D = A + B$, whereas Z_a , Z_b

denote nuclear charges, $\langle rs | tu \rangle$, $\langle r | Z_b R_{1b}^{-1} | s \rangle$, two-electron repulsion and one-electron nuclear attraction integrals, respectively.

The electrostatic penetration term $E_{EL,PEN}^{(10)}$ is defined as the difference between the entire electrostatic energy $E_{EL}^{(10)}$ and its multipole component $E_{EL,MT P}^{(10)}$:

$$E_{EL,PEN}^{(10)} = E_{EL}^{(10)} - E_{EL,MT P}^{(10)}$$

Taking mutually orthogonalized monomer wavefunctions obtained in the dimer basis set as a starting point, the first-order Heitler–London interaction energy $E^{(10)}$ is calculated as the difference between the AB dimer energy $E_{AB}(D)$ at iteration zero and the monomer energies $E_A(D)$ and $E_B(D)$ obtained in the dimer basis set:

$$E^{(10)} = E_{AB}(D) - E_A(D) - E_B(D)$$

Neglecting small Murrell delta term, the exchange repulsion component $E_{EX}^{(10)}$ is then defined as the difference

Table 4 Root mean square errors (RMSE) for short-range interaction energy components (in [kcal/mol]) approximated by $(\alpha + \beta R^{-1}) \exp(-\gamma R)$ function; R/R_{eq} values given in parenthesis

	RMSE(0.85)	RMSE(1.00)	RMSE(1.25)
$E_{EL,PEN}^{(10)}$	7.34	3.09	0.98
$E_{EX}^{(10)}$	2.79	1.26	0.34
$E_{DEL}^{(R0)}$	1.69	1.22	0.72
$E_{EX}^{(10)} + E_{DEL}^{(R0)}$	2.03	0.91	0.41
$E_{EL,PEN}^{(10)} + E_{EX}^{(10)} + E_{DEL}^{(R0)}$	8.68	3.24	0.94

Table 5 α , β and γ NEAAP parameters for exchange term $E_{EX}^{(10)}$ (a.u.)

contact	α	β	γ
H–H	8.07574773e+01	−1.87842011e+02	2.60272196e+00
C–H	1.08945456e+01	1.17274724e+00	1.82765379e+00
H–N	2.10150271e+01	−3.19179556e+01	1.87159307e+00
H–O	1.84654540e+01	−3.06952584e+01	1.95116187e+00
C–C	7.15660935e+02	−2.39375435e+03	1.92479648e+00
C–N	1.97976749e+01	3.85610972e+03	2.25503526e+00
C–O	6.18252120e+04	−2.82918123e+05	2.86485235e+00
N–N	2.01300373e+03	−7.55988423e+03	2.04260802e+00
N–O	3.39661651e+03	−4.91074360e+03	2.39791080e+00
O–O	2.98401193e+03	6.37198459e+02	2.49335288e+00

between the Heitler–London $E^{(10)}$ term defined above and the electrostatic component $E_{EL}^{(10)}$:

$$E_{EX}^{(10)} = E^{(10)} - E_{EL}^{(10)}$$

Another component $E_{DEL}^{(R0)}$, called the delocalization term, covers higher-order (R) induction and exchange-deformation interactions [25] and is obtained as the difference between the converged SCF interaction energy ΔE^{SCF} and the Heitler–London term $E^{(10)}$:

$$E_{DEL}^{(R0)} = \Delta E^{SCF} - E^{(10)}$$

It has to be noted that further partitioning $E_{DEL}^{(R0)}$ yields strongly basis set-dependent induction and charge transfer terms whereas their sum, i.e., delocalization term does not display such dependency [26]. Dispersion and exchange dispersion terms obtained within SAPT approach ($E_{DISP}^{(2)} + E_{EX-DISP}^{(2)}$) as well as first-order correlation correction $\epsilon(1)$ can be closely approximated by atom–atom potentials that include damping functions D_{as} [27] which represents inter- and intra-molecular correlation term E_{CORR} .

High quality of D_{as} functions allows to supplement SCF interaction energy [27] by correlation effects avoiding

well-known deficiencies of DFT or MP2 methods to represent dispersion interactions.

Thus, the total interaction energy used in this study can be expressed as the sum of the following terms:

$$\Delta E = E_{EL, MTP}^{(10)} + E_{EL, PEN}^{(10)} + E_{EX}^{(10)} + E_{DEL}^{(R0)} + D_{as},$$

where both long-range $E_{EL, MTP}^{(10)}$ and D_{as} terms scale with the number of atoms squared $O(A^2)$ and will be supplemented by similarly scaling atom–atom potentials derived in this work to approximate the short-range $E_{EL, PEN}^{(10)}$, $E_{EX}^{(10)}$ and $E_{DEL}^{(R0)}$ terms yielding complete nonempirical estimate of major nonbonded interactions applicable in any force field.

Results and discussion

Selection of the optimal functional form for the exchange repulsion term

Due to the critical importance of the repulsion term in empirical force fields [9], we extensively tested various possible functions starting with the most popular ones such as

Table 6 α , β and γ NEAAP parameters for delocalization component $E_{DEL}^{(R0)}$ (a.u.)

contact	α	β	γ
H–H	1.43313295e+03	1.43347328e+03	1.45137051e+03
C–H	−3.34374591e+01	8.27529346e+00	2.46669101e+00
H–N	−6.03027868e+00	6.19782453e+00	1.93834270e+00
H–O	−4.43751171e+00	2.71105734e+00	2.04151238e+00
C–C	−3.57124523e+01	−1.42264717e+02	2.04559232e+00
C–N	2.14345865e+03	5.17142579e+04	3.54918875e+00
C–O	−3.94230327e+00	1.81948699e+01	1.38730014e+00
N–N	−1.39410649e+02	5.56285349e+02	1.75555911e+00
N–O	−5.83250170e+02	1.55276368e+03	2.10644431e+00
O–O	−2.33891971e+03	4.71728778e+03	2.47343097e+00

Table 7 α , β and γ NEAAP parameters for penetration term $E_{EL,PEN}^{(10)}$ (a.u.)

contact	α	β	γ
H–H	1.62871696e+00	2.26905873e+04	4.54951049e+00
C–H	−3.38110031e-01	−3.33076551e+00	1.38327627e+00
H–N	−2.69691770e+00	3.36751606e+00	1.86353492e+00
H–O	−1.03106690e+01	5.92067460e+00	2.13247195e+00
C–C	1.57165940e-01	−1.02424207e+00	6.74948316e-01
C–N	1.63595055e+03	−9.80531178e+03	2.19443829e+00
C–O	−2.22799682e+04	1.05606596e+05	2.60365699e+00
N–N	−8.39021449e+03	3.78601968e+04	2.56001028e+00
N–O	−1.81525054e+00	9.74306338e+00	9.13237912e-01
O–O	7.06756222e+05	−2.01116758e+06	3.67775467e+00

αR^{-12} , $\alpha \exp(-\gamma R)$ and ending with $(\alpha + \beta R^{-1} + \delta R + \kappa R^2 + \omega R^{-3}) \exp(-\gamma R)$, as well as simpler intermediate versions. As reference data, we assumed values of the first-order exchange term obtained for 660 dimers of biomolecular complexes using the aug-cc-pVDZ basis set, generated from the S66 training set [28]. This set included 23 hydrogen-bonded, 23 dispersion-dominated, and 20 mixed molecular complexes composed of hydrogen, carbon, nitrogen, and oxygen only. Additional inclusion of sulphur, phosphorus, as well as halogen complexes involved in sigma-hole bonding is planned in future. Besides six shortest original distances defined by the ratio R/R_{eq} , we generated an additional S66x4 set to cover shorter distances critical for repulsion interactions. The parameters α , β , γ , δ , κ , and ω that appear in the longest functional form given above approximating the nonempirical exchange energy, were optimized by nonlinear least-squares fitting with all weights equal 1 using Powell's conjugate direction method [29]. The corresponding total root mean square errors (RMSE) as well as for distances close to equilibrium structures (eq) are given in Table 1, together with mean unsigned error (MUE) (Table 2) and mean unsigned relative error (MURE) values (Table 3). The results presented in Tables 1–3 indicate that the αR^{-12} function fails completely, whereas among all functions considered $(\alpha + \beta R^{-1}) \exp(-\gamma R)$ seems to provide an optimal representation of the exchange repulsion interaction around equilibrium distances as illustrated in Fig. 1. The distance dependence of the ratio of various approximations and exact reference

$E_{EX}^{(10)}$ values is shown in Fig. 2 for the methanol dimer. Analogous plots for acetate–methanol, methylamine–methanol, and methylammonium–methanol complexes are shown as Figs. 3, 4, and 5. Again, the αR^{-12} function from the Amber force field or NEAAP seem to yield considerably underestimated repulsion over the entire distance range. Very poor performance of Amber repulsive FF term could be due to direct coupling it to R^{-6} van der Waals component via imprecisely defined well depth and equilibrium distance. Since the optimal function $(\alpha + \beta R^{-1}) \exp(-\gamma R)$ resembles the conventional force field expression, it could be easily incorporated into existing molecular mechanics or dynamics packages. Computationally costly contribution of the three-body interactions dominated by induction term seems to be negligible and has only small influence on final geometries.

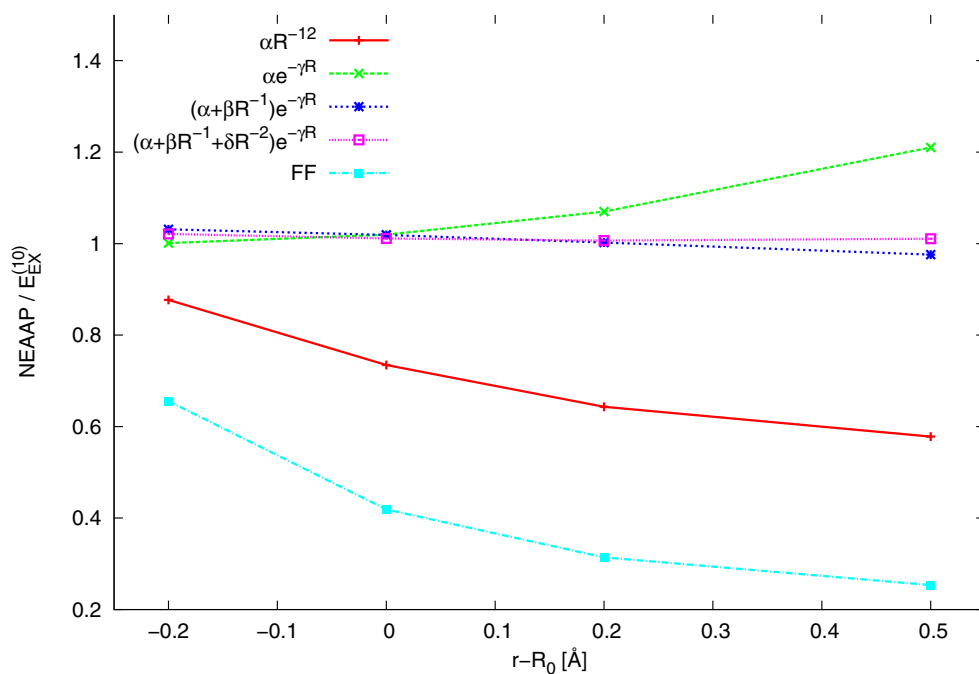
Atom–atom representation of the remaining interaction energy terms

Due to the presumably critical role of the exchange repulsion term in determining equilibrium geometries, the $(\alpha + \beta R^{-1}) \exp(-\gamma R)$ function has also been applied to approximate the remaining short-range components, namely delocalization $E_{DEL}^{(R0)}$ and electrostatic penetration $E_{EL,PEN}^{(10)}$, in order to compare them on an equal footing. The values of root mean square errors around equilibrium presented in Table 4 indicate that the delocalization term $E_{DEL}^{(R0)}$ could be also reasonably represented by $(\alpha + \beta R^{-1}) \exp(-\gamma R)$.

Table 8 Equilibrium distances [Å] between inhibitors and residues at the active site of urokinase determined by force field [10], NEAAP potential functions and ab initio calculations [10]. Values in parentheses indicate deviation from MP2 equilibrium short contacts

inhibitor#	SER 190			GLY219		
	Tripos FF	NEAAP	MP2	Tripos FF	NEAAP	MP2
1	2.88 (−0.20)	2.88 (−0.20)	3.08	1.56 (−0.50)	1.96 (−0.10)	2.06
2	1.84 (−0.56)	2.55 (0.15)	2.40	1.62 (−0.40)	2.02 (−0.0)	2.02
3	3.19 (−0.50)	3.69 (0.0)	3.69	2.02 (−0.30)	2.22 (−0.10)	2.32
4	3.11 (+0.41)	2.50 (−0.20)	2.70	2.08 (+0.10)	1.88 (−0.10)	1.98

Fig. 5 The distance dependence of the ratio of various approximations of the exchange repulsion term and exact reference $E_{EX}^{(10)}$ values for methylammonium-methanol complex



On the other hand, the electrostatic penetration component $E_{EL,PEN}^{(10)}$ seems to require a more complex functional form, for example like the one recently proposed by Tafipolsky and Engels [30]. Due to the relatively small contribution of electrostatic penetration effects (on average, electrostatic penetration at equilibrium in the S66 test set is 14.8 times smaller than exchange repulsion), we kept its functional

form the same as for the exchange and delocalization terms for the sake of simplicity. The corresponding α , β , and γ parameters are given in Tables 5–7. Combining $E_{DEL}^{(R0)}$ and $E_{EX}^{(10)}$, or $E_{DEL}^{(R0)}$, $E_{EX}^{(10)}$ and $E_{EL,PEN}^{(10)}$ terms leads to some RMSE reduction (Table 4) due to error compensation, but we did not resort to this in order to keep a clear meaning for all interaction energy components.

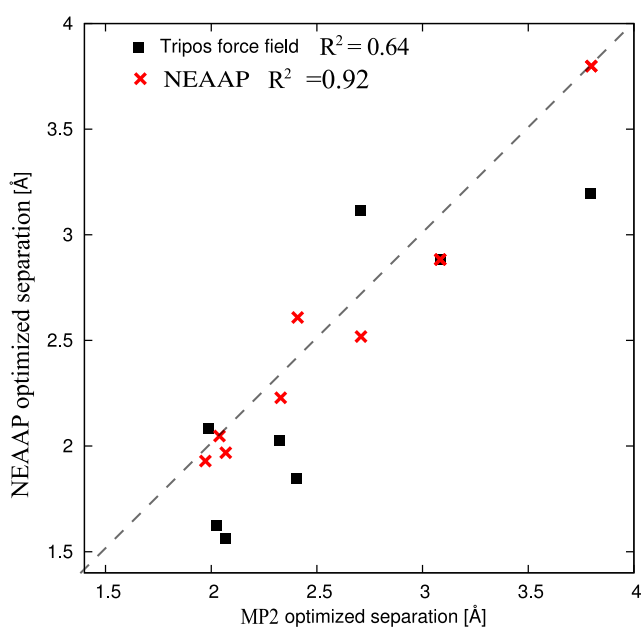


Fig. 6 Equilibrium separations predicted based on nonempirical atom–atom potentials (NEAAP) or the Tripos force field, compared to corresponding ab initio MP2 results for urokinase and its inhibitors

Testing nonempirical atom–atom potentials for inhibitor–active site complexes

The NEAAP potentials derived for S66 model biomolecular complexes have been tested for several urokinase–inhibitor complexes [10], where a force field docking resulted in several short contacts shown in Table 8. Equilibrium distances for active site amino acid–inhibitor contacts obtained using the standard Tripos 5.2 force field and our atom–atom potentials are compared in Fig. 6 alongside MP2 results. Clearly, the application of nonempirical atom–atom potentials results in considerable improvement, yielding a correlation coefficient $R^2 = 0.92$ between NEAAP contact distances and MP2 data, in contrast to the force field value of $R^2 = 0.64$. This feature can be very useful in improving the quality of structural predictions, which can be of practical importance in drug design and scoring. It is possible that the occurrence of artefact short contacts in simulations based on conventional force fields is underreported in the literature as locating it requires significant computational effort and incorrect data are overshadowed by other results while calculating statistical averages.

Conclusions

Derivation of universal nonempirical atom–atom potentials NEAAP from interaction energy components defined within hybrid variation-perturbation theory HVPT opens the possibility for systematic improvements to force field non-bonded terms that are critical for more accurate modeling of molecular materials. This study indicates that the analytical functions αR^{-12} or $\alpha \exp(-\beta R)$ commonly used to represent repulsive interactions are not adequate. By using the more appropriate but still relatively simple form $(\alpha + \beta R^{-1}) \exp(-\gamma R)$ it is possible to obtain a better description of exchange repulsion over a wide range of distances, especially around equilibrium. The application of derived NEAAPs resulted in a considerable improvement of the structural characteristics for an enzyme–inhibitor complex that exhibited artefact short contacts when a conventional force field was applied in the past [10].

Acknowledgments This work was supported by Wrocław Research Centre EIT+ within the project "Biotechnologies and advanced medical technologies" - BioMed (POIG.01.01.02-02-003/08) co-financed by the European Regional Development Fund (Operational Programme Innovative Economy, 1.1.2). Partial financing by a statutory activity subsidy from Polish Ministry of Science and Higher Education for Faculty of Chemistry of Wrocław University of Technology is also acknowledged. Calculations were performed in supercomputer centers in Wrocław (WCSS), Poznań (PCSS), and Warsaw (ICM). The authors are indebted to Dr. Karol M. Langner from the University of Virginia at Charlottesville, VA, USA, for stimulating discussion and valuable comments.

References

- Sokalski WA, Lowrey AH, Roszak S, Lewchenko V, Blaisdell J, Hariharan PC, Kaufman JJ (1986) Nonempirical atom–atom potentials for main components of intermolecular energy. *J Comp Chem* 7:693–700
- Gilson MK, Honig BH (1988) Energetics of charge–charge interactions in proteins. *Proteins: Struct Funct Dyn* 3:32–52
- Roterman I, Gilson KD, Scheraga HA (1989) A comparison of the CHARMM, AMBER and ECEPP potentials for peptides. 1. Conformational predictions for the tandemly repeated peptide (*asn – ala – asn – pro*)₉. *J Biomol Str Dyn* 7:391–419
- Gresh N, Claverie P, Pullman A (1986) Intermolecular interactions: elaboration on an additive procedure including an explicit charge-transfer contribution. *Int J Quantum Chem* 29:101–118
- Singh UC, Kollman PA (1985) A water dimer potential based on ab initio calculations using Morokuma component analyses. *J Chem Phys* 83:4033–4040
- Torheyden M, Jansen G (2006) A new potential energy surface for the water dimer obtained from separate fits of ab initio electrostatic, induction, dispersion and exchange energy contributions. *Mol Phys* 104:2101–2138
- Parish RM, Sherill CD (2014) Spatial assignment of symmetry adapted perturbation theory interaction energy components: the atomic SAPT. *J Chem Phys* 141:044115
- Schmidt JR, Yu K, McDaniel JG (2015) Transferable next-generation force fields from simple liquids to complex materials. *Acc Chem Res* 48:548–556
- Zgarbová M, Otyepka M, Šponer J, Hobza P, Jurečka P (2010) Large-scale compensation of errors in pairwise-additive empirical force fields: comparison of AMBER intermolecular terms with rigorous DFT-SAPT calculations. *Phys Chem Chem Phys* 12:10476–10493
- Grzywa R, Dyguda-Kazimierowicz E, Sieńczyk M, Feliks M, Sokalski WA, Oleksyszyn J (2007) The molecular basis of urokinase inhibition: from the nonempirical analysis of intermolecular interactions to the prediction of binding affinity. *J Mol Model* 13:677–683
- Murrell JN, Teixeira JJ (1970) Dependence of exchange energy on orbital overlap. *Chem Phys Lett* 19:521–525
- Sokalski WA, Chojnacki H (1978) Approximate exchange perturbation study of inter-molecular interactions in molecular complexes. *Int J Quantum Chem* 13:679–692
- Sokalski WA, Roszak S, Lowrey AH, Hariharan P, Koski WS, Kaufman JJ (1983) Crystal structure studies using ab-initio potential functions from partitioned MODPOT/VRDDO SCF energy calculations. I. *N₂* and *CO₂* test cases. II. Nitromethane *CH₃NO₂*. *Int J Quantum Chem: Quantum Chemistry Symp* 17:375–391
- Kitaura K, Morokuma K (1976) New energy decomposition scheme for molecular-interactions within Hartree–Fock approximation. *Int J Quantum Chem* 10:325–340
- Jeziorski B, Moszyński R, Szalewicz K (1994) Perturbation theory approach to intermolecular potential energy surfaces of van der Waals complexes. *Chem Rev* 94:1887–1930
- Politzer P, Murray J, Clark T (2015) Mathematical modeling and physical reality in noncovalent interactions. *J Mol Model* 21:52
- Sokalski WA, Roszak S, Pecul K (1988) An efficient procedure for decomposition of the SCF interaction energy into components with reduced basis set dependence. *Chem Phys Lett* 153:153–159
- Szefczyk B, Mulholland A, Ranaghan K, Sokalski WA (2004) Differential transition state stabilization in enzyme catalysis: quantum chemical analysis of interactions in the chorismate mutase reaction and prediction of the optimal catalytic field. *J Am Chem Soc* 126:16148–16159
- Langner KM, Janowski T, Gora R, Dziekoński P, Sokalski WA, Pulay P (2011) The ethidium-UA/AU intercalation site: effect of model fragmentation and backbone charge state. *J Chem Theor Comp* 7:2600–2609
- Sokalski WA, Poirier RA (1983) Cumulative atomic multipole representation of the molecular charge distribution and its basis set dependence. *Chem Phys Lett* 98:86–92
- Sokalski WA, Sawaryn A (1987) Correlated molecular and cumulative atomic multipole moments. *J Chem Phys* 87:526–534
- Schmidt MW, Baldrige KK, Boatz JA, Elbert ST, Gordon MS, Jensen JH, Koseki S, Matsunaga N, Nguyen KA, Su S, Windus TL, Dupuis M, Montgomery JA (1993) General atomic and molecular electronic structure system. *J Comp Chem* 14:1347–1363
- Sokalski WA, Shibata M, Ornstein RL, Rein R (1993) Point-charge representation of multicenter multipole moments in calculation of electrostatic properties. *Theor Chim Acta* 85:209–216
- Devereux M, Raghunathan S, Federov DG, Meuwly M (2014) A novel, computationally efficient multipolar model employing distributed charges for molecular dynamics simulations. *J Chem Theor Comp* 10:4229–4241

25. Chałasiński G, Szczyński MM (1994) Origins of structure and energetics of van der Waals clusters from ab initio calculations. *Chem Rev* 94:1723–1765
26. Sokalski WA, Roszak S (1991) Efficient techniques for the decomposition of intermolecular interaction energy at SCF level and beyond. *J Mol Struct(THEOCHEM)* 234:387–400
27. Podeszwa R, Pernal K, Patkowski K, Szalewicz K (2010) Extension of the Hartree–Fock plus dispersion method by first-order correlation effects. *J Phys Chem Lett* 1:550–555
28. Řezáč J, Riley KE, Hobza P (2011) S66: A well-balanced database of benchmark interaction energies relevant to biomolecular structures. *J Chem Theor Comput* 7:2427–2438
29. Powell MJD (1964) An efficient method for finding the minimum of a function of several variables without calculating derivatives. *Comp J* 7:155–162
30. Tafipolsky M, Engels B (2011) Accurate intermolecular potentials with physically grounded electrostatics. *J Chem Theor Comp* 7:1791–1803