

Supplementary information for: “First-principles many-body non-additive polarization energies from monomer and dimer calculations only : A case study on water”

Rory A. J. Gilmore, Martin T. Dove, and Alston J. Misquitta

School of Physics and Astronomy and the Thomas Young Centre for Theory and Simulation of Materials at Queen Mary University of London, London E1 4NS, U.K.

E-mail:

1 Model specifications

1.1 DIFF functional form

The DIFF functional form is not fixed, but instead is determined by the best theoretical understanding available. At present we utilize an anisotropic Born–Mayer functional form¹ but we have also used alternative forms^{2,3} sometimes with better results.

Following Misquitta & Stone¹ we represent the potential V_{int} as

$$V_{\text{int}} = \sum_{a \in A} \sum_{b \in B} V_{\text{pair}}[ab](r_{ab}, \Omega_{ab}) + V_{\text{pol}}[\text{cluster}], \quad (1)$$

where, a and b label sites in the interacting molecules A and B , r_{ab} is the inter-site separation, Ω_{ab} is a suitable set of angular coordinates that describes the relative orientation of the local axis systems

on these sites (see ch. 12 in ref. 4), and $V_{\text{pair}}[ab]$ is the pair-wise site–site potential defined as

$$V_{\text{pair}}[ab] = V_{\text{sr}}[ab] + V_{\text{elst}}[ab] + V_{\text{disp}}[ab]. \quad (2)$$

Note that the polarization term cannot be written as a pair-wise sum over sites as it is a manifestly many-body term.

The short-range term $V_{\text{sr}}[ab]$ describes the exchange–repulsion energy, the electrostatic penetration energy, and all other short-range terms, including the charge-delocalization energy:

$$V_{\text{sr}}[ab] = G \exp [-\alpha_{ab}(\Omega_{ab})(r_{ab} - \rho_{ab}(\Omega_{ab}))], \quad (3)$$

where $\rho_{ab}(\Omega_{ab})$ is the shape function for this pair of sites, which depends on their relative orientation described by Ω_{ab} , and α_{ab} is the hardness parameter which will be taken to be independent of orientation. $G = 10^{-3}$ hartree is a constant energy which determines the units of $V_{\text{sr}}[ab]$. The shape-function $\rho_{ab}(\Omega_{ab})$ for site pair ab is dependent on the relative orientation of these sites Ω_{ab} and is given by

$$\rho_{ab}(\Omega_{ab}) = \rho_{ab}^a(\Omega_{ab}) + \rho_{ab}^b(\Omega_{ab}) \quad (4)$$

where $\rho_{ab}^a(\Omega_{ab}) = \sum_{l,k} \rho_{lk}^a C_{lk}(\theta_a, \phi_a)$ is the shape function for atom a and $C_{lk}(\theta, \phi) = \frac{4\pi}{2l+1} Y_{l,m}(\theta, \phi)$ is a renormalised spherical harmonic term.

The shape function $\rho_{ab}(\Omega_{ab})$ is best described in local axis systems that reflect the local symmetries of the sites a and b . These symmetries could be approximate. For example, a convenient choice for the local z -axis at a carbon atom in a benzene molecule might be to have it point from the carbon to the bonded hydrogen atom. With this choice of z -axis, an approximate cylindrical symmetry may be imposed. In which case, the potential parameters would be quite simple. But we now need to transform from these local axis systems to the global axis as the molecular configurations are defined in the global, or laboratory frame. This transformation is done using the

S -functions defined by eqs. 3.3.7 in ref.⁴ and is given by (eq. 12.2.6 in ref.⁴)

$$\rho_{ab}(\Omega_{ab}) = \sum_{l_a l_b j k_a k_b} \rho_{l_a l_b j}^{k_a k_b} \bar{S}_{l_a l_b j}^{k_a k_b}. \quad (5)$$

We do not use the most general S -function in our potentials, but only the special cases: $\bar{S}_{l0l}^{k0}$ and $\bar{S}_{0ll}^{0k}$. Since we do not use mixed terms in the sum, this leads to a very intuitive result that the shape function of a pair of sites is the sum of the shape functions of the individual sites. This is so because these special S -functions can be written quite simply as

$$\bar{S}_{l0l}^{k0} = C_{l,k}(\theta, \phi)^*, \quad (6)$$

where the renormalized spherical harmonics (in the Racah definition) are defined as

$$C_{l,k}(\theta, \phi) = \sqrt{\frac{4\pi}{2l+1}} Y_{lm}(\theta, \phi). \quad (7)$$

We can use the real components of the renormalized spherical harmonics (defined below) to get

$$\bar{S}_{l0l}^{\kappa 0} = C_{l,\kappa}(\theta_a, \phi_a), \quad (8)$$

where the Greek letter κ has been used in place of k to indicate this is the real component and the angles now have subscripts a to indicate they are the polar coordinates describing the site–site vector from a to b in the local axis system of site a . Likewise, we define

$$\bar{S}_{0ll}^{0\kappa} = C_{l,\kappa}(\theta_b, \phi_b). \quad (9)$$

Now we can write the (approximate) shape function as

$$\rho_{ab}(\Omega_{ab}) = \rho^a(\theta_a, \phi_a) + \rho^b(\theta_b, \phi_b), \quad (10)$$

where

$$\rho^a(\theta_a, \phi_a) = \sum_{l\kappa} \rho_{l\kappa}^a C_{l,\kappa}(\theta_a, \phi_a), \quad (11)$$

with a similar expression for $\rho^b(\theta_b, \phi_b)$.

We can interpret ρ^a as the shape function of site a . This is a very useful concept when developing atom–atom potentials with the aim of *transferability*, where it is important to define the parameters in the potential in terms of the properties of the atomic sites. However Misquitta & Stone¹ have argued that this interpretation is only valid at first order. When second-order terms are included then there is a coupling between the parameters from sites a and b as would happen, for example, if there was a strong charge-delocalization between the sites. This happens for the O..H interaction in water. So in the DIFF models we have specific parameters sets for the O..O, H..H, and O..H interactions; i.e., transferability is not imposed. Indeed, it cannot be imposed without compromising the accuracy of the models.

$V_{\text{elst}}[ab]$ is the expanded electrostatic energy:

$$V_{\text{elst}}[ab] = V_{\text{elst}}[ab](r_{ab}, \Omega_{ab}, Q_t^a, Q_u^b) \quad (12)$$

$$= \sum_{tu} Q_t^a T_{tu}^{ab} Q_u^b, \quad (13)$$

where Q_t^a is the multipole moment of rank t for site a , where, using the notation of ref. 4, the angular momenta are indexed as $t = l\kappa = 00, 10, 11c, 11s, 20, 21c, 21s, 22c, 22s, \dots$. Notice that we do not (as yet) damp the electrostatic model. Here T_{tu}^{ab} is a T -tensor defined as (see eq. 3.3.14 in Stone⁴)

$$T_{tu}^{ab} \equiv T_{l_1\kappa_1, l_2\kappa_2}^{ab} \quad (14)$$

$$= \left[\frac{(l_1 + l_2)!}{l_1! l_2!} \right] \bar{S}_{l_1 l_2 l_1 + l_2}^{\kappa_1 \kappa_2} r_{ab}^{-(l_1 + l_2 + 1)}. \quad (15)$$

The polarization term $V_{\text{pol}}[\text{cluster}]$ has been defined in the main body of the paper but is provided here for completeness. In the DIFF models the polarization energy is defined *between*

molecules and not within molecules. This is because the polarizabilities defined using the ISA-Pol algorithm already include intramolecular polarization effects. So the polarization model is defined not in terms of sites, but in terms of molecules. Further, as this is a manifestly many-body term, it cannot be written in terms of pairs of molecules, but is defined as

$$V_{\text{pol}}[\text{cluster}] = \sum_A E_{\text{pol,cl}}(A), \quad (16)$$

where the classical polarization energy of a molecule A in a cluster of molecules is defined as:

$$E_{\text{pol,cl}}(A) = \frac{1}{2} \sum_{a \in A} \sum_{B \neq A} \sum_{b \in B} \sum_{tu} \Delta Q_t^a f_{n(tu)}(\beta_{\text{pol}}^{ab} R_{ab}) T_{tu}^{ab} Q_u^b, \quad (17)$$

where the terms are as defined above for the electrostatic case. $f_{n(tu)}(\beta_{\text{pol}}^{ab} R_{ab})$ is a damping function of order n , that will usually be taken to be the Tang–Toennies incomplete Gamma function of order $n + 1$:

$$f_n^{TT}(\beta R) = P(n + 1, \beta R) = 1 - \exp(-\beta R) \sum_{k=0}^n \frac{(\beta R)^k}{k!}. \quad (18)$$

Here n is a function of the tensor ranks t and u , and if $t = l_1 \kappa_1$ and $u = l_2 \kappa_2$, then $n = l_1 + l_2 + 1$. We assume that the damping depends only on the distance R_{ab} between sites a and b and not on their relative orientation. This is an approximation that needs assessment, but we do not address this issue in this paper. The strength of the damping is governed by the damping parameter β_{pol}^{ab} . In the above expression, ΔQ_t^a is the change in multipole moment t at a due to the self-consistent polarization of site a in the field of all sites on *other* molecules, and is given by

$$\Delta Q_t^a = - \sum_{a' \in A} \sum_{B \neq A} \sum_{b \in B} \sum_{t'v} \alpha_{tt'}^{aa'} f_{n(t'v)}(\beta_{\text{pol}}^{a'b} R_{a'b}) T_{t'v}^{a'b} (Q_v^b + \Delta Q_v^b), \quad (19)$$

where $\alpha_{tt'}^{aa'}$ is the distributed polarizability for sites (a, a') which describes the response of the multipole moment component Q_t^a at site a to the t' -component of the field at site a' . To find ΔQ_t^a

we need to solve eq. (19) iteratively. If ΔQ_v^b is dropped from the right-hand-side of this equation then the resulting ΔQ_t^a , when inserted in eq. (17) leads to the second-order polarization energy, $E_{\text{pol,cl}}^{(2)}$.

In the polarization models used in this paper we assume the Tang–Toennies form for the damping functions.⁵ Further we use the localized form^{6,7} of the distributed polarizability tensor, that is, the non-local polarizability $\alpha_{tu}^{aa'}$ in eq. (19) is replaced by $\alpha_{tu}^a \delta_{aa'}$, where $\delta_{aa'}$ is the Kronecker-delta and α_{tu}^a is the localized polarizability tensor of the same rank.

The dispersion energy $V_{\text{disp}}[ab]$ depends on the anisotropic dispersion coefficients $C_n^{ab}(\Omega_{ab})$ for the pair of sites, and on a damping function f_n that we will take to be the Tang–Toennies⁸ incomplete gamma functions of order $n + 1$:

$$V_{\text{disp}}[ab] = - \sum_{n=6}^{12} f_n(\beta_{\text{disp}}^{ab} r_{ab}) C_n^{ab}(\Omega_{ab}) r_{ab}^{-n} \quad (20)$$

1.2 Model parameters

The DIFF model parameters presented here are defined in local axis frame for each atom in the water molecule. In the notation used in the Orient program⁹ the local axes are defined to be as follows:

Axes

O z between H1 and H2 x from H1 to H2

H1 z from O to H1 x from H1 to H2

H2 z from O to H2 x from H2 to H1

End

This choice places the water molecule in the xz-plane with the z-axes on each H-atom pointing outwards, along the O-H bond, and that for the O-atom bisecting the H-O-H angle and pointing from the O towards the H-atoms.

The DIFF models are all created for the water molecule in a fixed geometry given as (in atomic units):

O	0.000000000	0.000000000	0.000000000
H1	-1.45365196	0.000000000	-1.12168732
H2	1.45365196	0.000000000	-1.12168732

In some of the water clusters the monomer geometries differ from the one above. In this case the DIFF model parameters presented below were moved onto the sites and the local axis system was kept the same. That is, the DIFF model parameters were transferred, without change, to the new monomer geometries, with the local axis system and parameters kept the same.

The polarization model parameters are given in the main paper, but are reproduced below for convenience.

Table 1: L1pol $x = 0.0$ (IP)

Pair	l_a	l_b	ρ	α
O O	00	00	0.575293E+01	0.189637E+01
	00	10	-0.405000E-02	
	10	00	-0.405000E-02	
	00	20	0.166090E-01	
	00	22c	-0.117827E+00	
	20	00	0.166090E-01	
	22c	00	-0.117827E+00	
O H	00	00	0.470612E+01	0.193432E+01
	00	10	-0.265887E+00	
	00	11c	0.179670E-01	
	10	00	-0.199970E-01	
	20	00	0.724500E-02	
	22c	00	-0.169075E+00	
H H	00	00	0.376139E+01	0.199594E+01
	00	10	-0.215258E+00	
	00	11c	0.483480E-01	
	10	00	-0.215258E+00	
	11c	00	0.483480E-01	

Table 2: L1pol $x = 0.5$

Pair	l_a	l_b	ρ	α
O O	00	00	0.574133E+01	0.189882E+01
	00	10	-0.651900E-02	
	10	00	-0.651900E-02	
	00	20	0.164840E-01	
	00	22c	-0.118650E+00	
	20	00	0.164840E-01	
	22c	00	-0.118650E+00	
O H	00	00	0.471851E+01	0.192714E+01
	00	10	-0.279558E+00	
	00	11c	0.190500E-01	
	10	00	-0.184210E-01	
	20	00	0.794900E-02	
	22c	00	-0.167358E+00	
H H	00	00	0.375748E+01	0.200219E+01
	00	10	-0.211388E+00	
	00	11c	0.468130E-01	
	10	00	-0.211388E+00	
	11c	00	0.468130E-01	

Table 3: DIFF-L1pol $x = 1.0$

Pair	l_a	l_b	ρ	α
O O	00	00	0.572103E+01	0.190426E+01
	00	10	-0.787800E-02	
	10	00	-0.787800E-02	
	00	20	0.166740E-01	
	00	22c	-0.118533E+00	
	20	00	0.166740E-01	
	22c	00	-0.118533E+00	
O H	00	00	0.473665E+01	0.191820E+01
	00	10	-0.296016E+00	
	00	11c	0.207420E-01	
	10	00	-0.171990E-01	
	20	00	0.847100E-02	
	22c	00	-0.164328E+00	
H H	00	00	0.375798E+01	0.199913E+01
	00	10	-0.207660E+00	
	00	11c	0.449220E-01	
	10	00	-0.207660E+00	
	11c	00	0.449220E-01	

Table 4: L1pol $x = 1.5$

Pair	l_a	l_b	ρ	α
O O	00	00	0.566957E+01	0.194093E+01
	00	10	-0.693800E-02	
	10	00	-0.693800E-02	
	00	20	0.162530E-01	
	00	22c	-0.116227E+00	
	20	00	0.162530E-01	
	22c	00	-0.116227E+00	
O H	00	00	0.477758E+01	0.189670E+01
	00	10	-0.315240E+00	
	00	11c	0.248030E-01	
	10	00	-0.164550E-01	
	20	00	0.953500E-02	
	22c	00	-0.158684E+00	
H H	00	00	0.377554E+01	0.196057E+01
	00	10	-0.207309E+00	
	00	11c	0.406770E-01	
	10	00	-0.207309E+00	
	11c	00	0.406770E-01	

Table 5: DIFF-L2pol $x = 1.0$

Pair	l_a	l_b	ρ	α
O O	00	00	0.570240E+01	0.188230E+01
	00	10	-0.288290E-01	
	10	00	-0.288290E-01	
	00	20	0.829100E-02	
	00	22c	-0.122959E+00	
	20	00	0.829100E-02	
	22c	00	-0.122959E+00	
O H	00	00	0.481023E+01	0.191122E+01
	00	10	-0.338792E+00	
	00	11c	0.258230E-01	
	10	00	-0.721500E-02	
	20	00	0.115020E-01	
	22c	00	-0.165928E+00	
H H	00	00	0.367186E+01	0.201785E+01
	00	10	-0.135984E+00	
	00	11c	0.385040E-01	
	10	00	-0.135984E+00	
	11c	00	0.385040E-01	

Table 6: L3pol $x = 0.0$ (IP)

Pair	l_a	l_b	ρ	α
O O	00	00	0.585650E+01	0.184777E+01
	00	10	0.168560E−01	
	10	00	0.168560E−01	
	00	20	0.109910E−01	
	00	22c	−0.108392E+00	
	20	00	0.109910E−01	
	22c	00	−0.108392E+00	
O H	00	00	0.467033E+01	0.201391E+01
	00	10	−0.190363E+00	
	00	11c	0.127970E−01	
	10	00	−0.126610E−01	
	20	00	0.400000E−04	
	22c	00	−0.178575E+00	
H H	00	00	0.370112E+01	0.195166E+01
	00	10	−0.166826E+00	
	00	11c	0.468670E−01	
	10	00	−0.166826E+00	
	11c	00	0.468670E−01	

Table 7: L3pol $x = 0.5$

Pair	l_a	l_b	ρ	α
O O	00	00	0.578240E+01	0.186982E+01
	00	10	−0.879600E−02	
	10	00	−0.879600E−02	
	00	20	0.851400E−02	
	00	22c	−0.120035E+00	
	20	00	0.851400E−02	
	22c	00	−0.120035E+00	
O H	00	00	0.475443E+01	0.193819E+01
	00	10	−0.254994E+00	
	00	11c	0.177050E−01	
	10	00	−0.421200E−02	
	20	00	0.604100E−02	
	22c	00	−0.172302E+00	
H H	00	00	0.367475E+01	0.202611E+01
	00	10	−0.157909E+00	
	00	11c	0.398460E−01	
	10	00	−0.157909E+00	
	11c	00	0.398460E−01	

Table 8: DIFF-L3pol $x = 1.0$

Pair	l_a	l_b	ρ	α
O O	00	00	0.573132E+01	0.188375E+01
	00	10	-0.170380E-01	
	10	00	-0.170380E-01	
	00	20	0.959900E-02	
	00	22c	-0.122018E+00	
	20	00	0.959900E-02	
	22c	00	-0.122018E+00	
O H	00	00	0.478851E+01	0.190116E+01
	00	10	-0.308985E+00	
	00	11c	0.193160E-01	
	10	00	-0.472900E-02	
	20	00	0.899500E-02	
	22c	00	-0.169783E+00	
H H	00	00	0.369505E+01	0.203243E+01
	00	10	-0.151968E+00	
	00	11c	0.375710E-01	
	10	00	-0.151968E+00	
	11c	00	0.375710E-01	

Table 9: L3pol $x = 1.5$

Pair	l_a	l_b	ρ	α
O O	00	00	0.566135E+01	0.192579E+01
	00	10	-0.163670E-01	
	10	00	-0.163670E-01	
	00	20	0.108010E-01	
	00	22c	-0.116228E+00	
	20	00	0.108010E-01	
	22c	00	-0.116228E+00	
O H	00	00	0.481617E+01	0.187771E+01
	00	10	-0.362695E+00	
	00	11c	0.232330E-01	
	10	00	-0.641400E-02	
	20	00	0.965000E-02	
	22c	00	-0.169048E+00	
H H	00	00	0.374351E+01	0.197627E+01
	00	10	-0.135674E+00	
	00	11c	0.385730E-01	
	10	00	-0.135674E+00	
	11c	00	0.385730E-01	

Table 10: Polarisation damping parameters used for each model used in this work. The column titled “IP” indicates the damping based on the ionization potential of water (see text for details), and “DIFF” indicates the optimized damping for the DIFF models. This table is also presented in the main paper.

	IP		DIFF	
	$x = 0$	$x = 0.5$	$x = 1$	$x = 1.5$
L3:				
β_{OO}	1.926	1.588	1.25	0.912
β_{OH}	1.926	1.698	1.47	1.242
β_{HH}	1.926	1.963	2.00	2.037
L2:				
β_{OO}	–	–	1.25	–
β_{OH}	–	–	1.57	–
β_{HH}	–	–	2.00	–
L1:				
β_{OO}	1.926	1.588	1.25	0.912
β_{OH}	1.926	1.803	1.68	1.557
β_{HH}	1.926	1.963	2.00	2.037

Table 11: Dispersion coefficients from the localized ISA-Pol model and site-site damping parameters. All terms in atomic units.

Pair (ab)	C_6^{ab}	C_8^{ab}	C_{10}^{ab}	C_{12}^{ab}	β_{disp}^{ab}
O O	24.34089	489.9063	12519.45	238364.1	1.7794
O H	4.335086	55.94859	1174.193	13116.46	1.9011
H H	0.7833591	4.356823	90.61106	771.3764	2.0227

Table 12: Non-zero components of the DF-ISA rank 4 multipole model in the local axes frame. Note that symmetry is not imposed so that there are small differences in the multipoles on the two hydrogen sites.

O			
t	Q_t^O		
00	-0.825458		
10	-0.170731		
20	0.013320		
22c	0.446098		
30	-0.111202		
32c	-0.116581		
40	-0.395115		
42c	0.449626		
44c	0.017959		
H1		H2	
t	Q_t^{H1}	t	Q_t^{H2}
00	0.413222	00	0.413227
10	0.016268	10	0.016266
11c	-0.022715	11c	-0.022713
20	0.026170	20	0.026166
21c	-0.012333	21c	-0.012333
22c	0.023062	22c	0.023061
30	0.022590	30	0.022592
31c	0.009128	31c	0.009123
32c	-0.000725	32c	-0.000723
33c	-0.001680	33c	-0.001680
40	-0.047819	40	-0.047805
41c	0.039014	41c	0.039005
42c	-0.027919	42c	-0.027914
43c	-0.000806	43c	-0.000803
44c	0.005395	44c	0.005393

Table 13: α_{tu}^{OO} for the ISA-Pol L3pol model. Terms are expressed in the local-axis frame. Units are atomic units.

t, u	α_{tu}^{OO}	t, u	α_{tu}^{OO}
10,10	6.938561067132	21s,31s	-5.695850943476
10,20	0.283014503203	21s,33s	-5.544559210436
10,22c	1.297065784153	22c,22c	31.003555795636
10,30	-1.924692628858	22c,30	-2.134821966453
10,32c	-5.065926516432	22c,32c	6.387970997851
11c,11c	6.758614571817	22s,22s	28.975306251152
11c,21c	2.609015559476	22s,32s	-5.423168252926
11c,31c	5.869123330783	30,30	197.513252679545
11c,33c	-6.605301843408	30,31c	-0.013502436204
11s,11s	7.522295544935	30,21c	39.341637391107
11s,21s	0.534498269999	30,33c	-0.016042300561
11s,31s	-3.263639064343	31c,31c	196.876749295609
11s,33s	-9.492140869903	31c,32c	0.025580817673
20,20	27.634920511971	31c,33c	22.885787318830
20,22s	6.178197154040	31s,31s	194.533538449987
20,30	-1.850509303928	31s,33s	32.298750884673
20,32c	-1.859913227916	32c,32c	237.101696439329
21c,21c	27.989582837200	32c,33c	0.025727884497
21c,31c	3.459918679126	32s,32s	235.740220888427
21c,33c	5.231198193522	33c,33c	258.178588469019
21s,21s	33.110619559210	33s,33s	202.634991496274

Table 14: α_{tu}^{HH} for the ISA-Pol L3pol model. Terms are expressed in the local-axis frame. Units are atomic units.

t, u	α_{tu}^{HH}	t, u	α_{tu}^{HH}
10,10	2.133530241229	21c,30	0.491372594083
10,11c	0.014806076479	21c,31c	3.244889231824
10,21c	-1.962195867167	21c,32c	-0.164192398715
10,21s	0.558394807976	21c,33c	0.502961820740
10,22c	-0.139007994819	21s,21s	0.862161214828
10,30	2.624600448864	21s,22s	0.067763055234
10,31c	-0.387671591787	21s,31s	4.658283502883
10,32c	-0.285952156735	21s,32s	-0.299828269759
10,33c	-0.039088323730	21s,33s	0.123635495859
11c,10	0.014806076479	22c,22c	2.258497604970
11c,11c	0.768893987100	22c,30	-0.014559136681
11c,20	-0.060959104003	22c,31c	0.073289640942
11c,21c	0.204839449553	22c,32c	1.932260882992
11c,22c	0.146065219982	22c,33c	0.846069547066
11c,30	-0.084814571913	22s,22s	2.358486780599
11c,31c	0.025623412166	22s,31s	-0.083428643181
11c,32c	0.374664191337	22s,32s	2.019910136457
11c,33c	-0.399931921642	22s,33s	0.526987026651
11s,11s	0.787159280680	30,30	-2.732173577046
11s,21s	0.553458688663	30,31c	-0.976616012299
11s,22s	0.183476194197	30,32c	5.916676794459
11s,31s	-0.581574706631	30,33c	2.459537631952
11s,32s	0.335756030493	31c,31c	11.389988339701
11s,33s	-0.332939366208	31c,32c	4.464304862624
20,20	4.693088404029	31c,33c	4.569874943117
20,21c	-0.677855615155	31s,31s	13.964067932662
20,22c	0.725066406875	31s,32s	3.094566951749
20,30	1.210062157884	31s,33s	-0.301357361307
20,31c	1.041758490711	32c,32c	0.281721397105
20,32c	0.105983816308	32c,33c	-2.903496260436
20,33c	0.018679739312	32s,32s	-4.032641998208
21c,21c	1.423976650100	32s,33s	0.783324615248
21c,22c	0.121006332199	33s,33s	11.308337868268

Table 15: α_{tu}^{OO} for the ISA-Pol L2pol model. Terms are expressed in the local-axis frame. Units are atomic units.

t, u	α_{tu}^{OO}
10,10	6.978439619408
10,20	0.298266643849
10,22c	1.284142535503
11c,11c	6.923882571524
11c,21c	2.590522565151
11s,11s	7.712578185325
11s,21s	0.590880476972
20,20	27.998687464761
20,22c	6.115649837756
21c,21c	27.431664736301
21s,21s	33.400963369514
22c,22c	31.418111682887
22s,22s	30.005436201383

Table 16: α_{tu}^{HH} for the ISA-Pol L2pol model. Terms are expressed in the local-axis frame. Units are atomic units.

t, u	α_{tu}^{HH}	t, u	α_{tu}^{HH}
10,10	2.080105897425	11s,22s	0.302067978718
10,11c	0.053116146435	20,20	4.814477514669
10,20	-1.854642739739	20,21c	-0.684744785441
10,21c	0.655677997532	20,22c	0.730912286810
10,22c	-0.203666996438	21c,21c	1.438600109316
11c,11c	0.712203468085	21c,22c	0.121807515267
11c,20	-0.123521050531	21,21s	0.870977300150
11c,21c	0.166372245640	21s,22s	0.064467545108
11c,22c	0.159427069532	22c,22c	2.257351971348
11s,11s	0.699536201785	22s,22s	2.378902651023
11s,21s	0.443538935284		

Table 17: α_{tu}^{OO} for the ISA-Pol L1pol model. Terms are expressed in the local-axis frame. Units are atomic units.

t, u	α_{tu}^{OO}
10,10	6.636912518476
11c,11c	6.397939492453
11s,11s	6.746166610856

Table 18: α_{tu}^{HH} for the ISA-Pol L1pol model. Terms are expressed in the local-axis frame. Units are atomic units.

t, u	α_{tu}^{HH}
10,10	2.221233947114
10,11c	-0.013585535815
11c,11c	1.021508106671
11s,11s	1.188309162710

2 Plots for the two-body interaction

Figure 1: The total interaction energy along the profile of the dimer global minimum for all DIFF models.

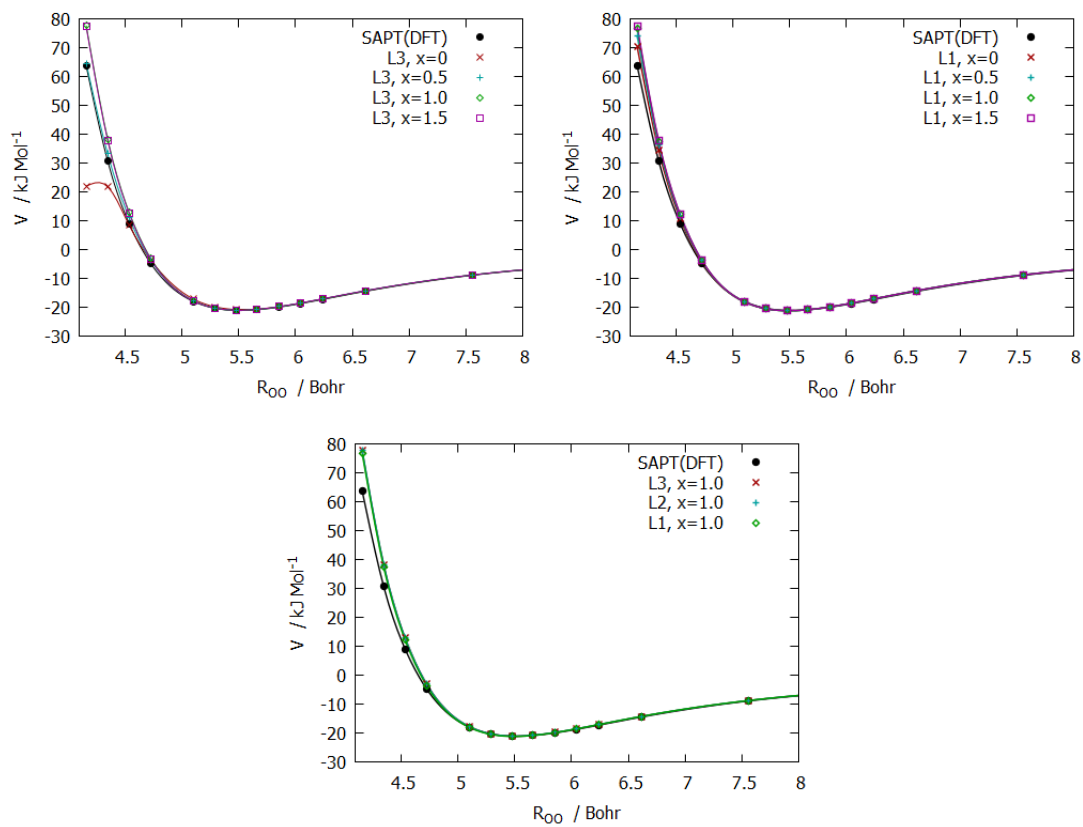
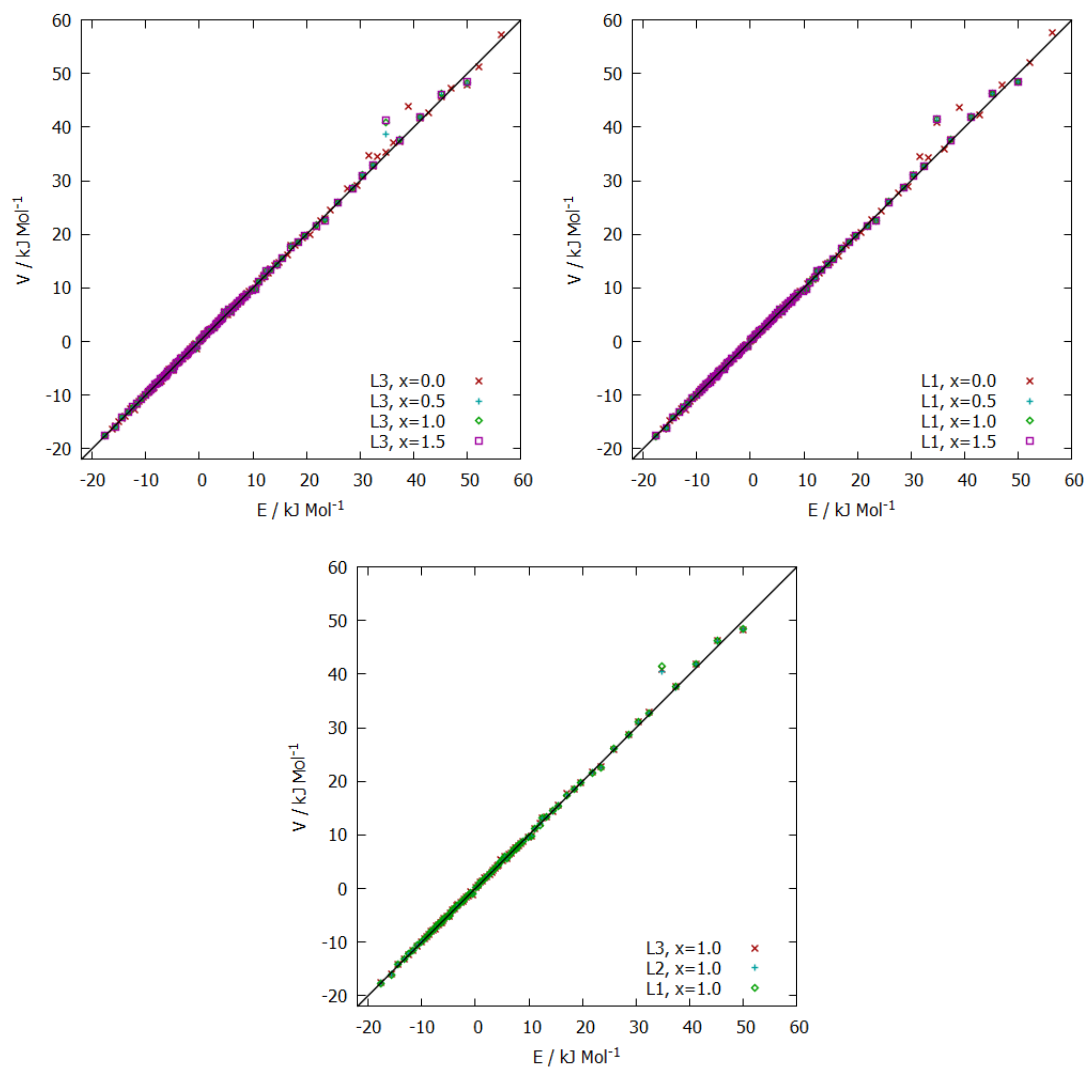


Figure 2: Scatter plot of total interaction energy for all DIFF models V versus SAPT(DFT) total interaction energy E .



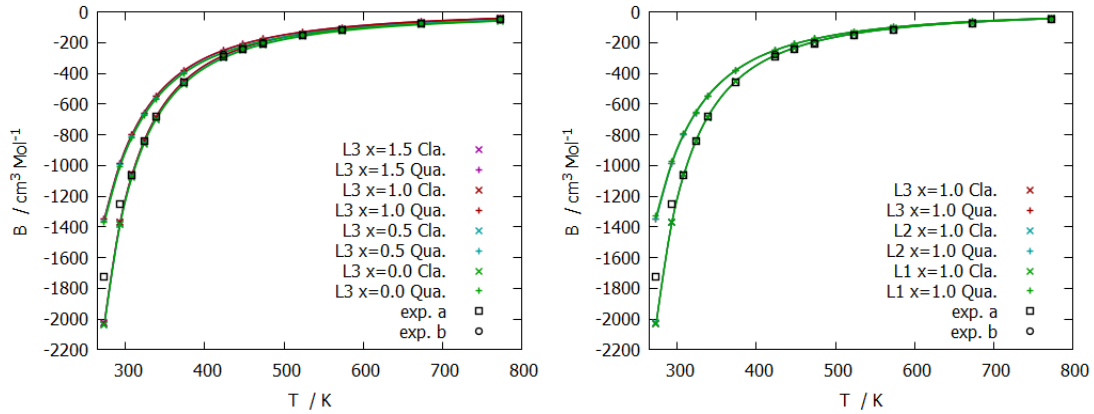
2.1 Second virial coefficient

The second virial coefficient $B(T)$ is calculated using

$$B(T) = -\frac{1}{2} \int \int (e^{E_{\text{int}}/kT} - 1) d\Omega dr^3 + \frac{\hbar^2}{24(kT)^3} \left(\frac{\langle \mathbf{F}^2 \rangle_0}{M} + \sum_{\alpha} \frac{\langle \mathbf{T}_{\alpha}^2 \rangle_0}{I_{\alpha\alpha}} \right) \quad (21)$$

where the first term above is the classical result $B(T)_{\text{Cl}}$ from integrating the Mayer function (the integration here is over separations and orientations) and the second term gives the quantum correction. Here $\langle \mathbf{F}^2 \rangle_0$ and $\langle \mathbf{T}_{\alpha}^2 \rangle_0$ are the mean square force and components of mean square torque on the molecule respectively and $I_{\alpha\alpha}$ are the molecule's moments of inertia.

Figure 3: Second virial coefficient for water for all L3 models and for the $x = 1.0$ L2 and L1 models plotted against temperature. Experimental data taken from Mas et al. (2000).¹⁰



3 Water cluster structures

Sources for the water cluster structures:

- $(\text{H}_2\text{O})_3$: Two sets were used. One from Liu *et al.*¹¹ and the 600 trimer set from Akin-Ojo and Szalewicz.¹²
- $(\text{H}_2\text{O})_6$: These structures were taken from Bates & Tschumper.¹³
- $(\text{H}_2\text{O})_{16}$: These were from Yoo and Xantheas¹⁴ and are also provided by Góra *et al.*¹⁵ in their SI. Note that for the boat-a structure given in these references the O and H sites are not ordered. The structure with the correct ordering is provided below.
- $(\text{H}_2\text{O})_{24}$: As for $(\text{H}_2\text{O})_{16}$.

3.1 $(\text{H}_2\text{O})_{16}$ boat-a isomer

This is the structure with the atoms ordered according to the water molecule they belong to:

48

boat-a: MP2/aug-cc-pVTZ (opt) E = -1221.536142 a.u.

O	0.12742587	1.14684972	-1.58027855
H	-0.71725804	1.53849679	-1.87225414
H	-0.00427839	0.17400371	-1.58650759
O	4.23247769	-0.26271901	-1.24153281
H	4.97071820	-0.25118561	-1.85737926
H	3.72092881	0.57286259	-1.40972697
O	-2.42866591	2.23165734	-1.85770461
H	-2.68572875	3.03888825	-2.31286191
H	-2.49329627	2.42735616	-0.88330238
O	-4.23429048	0.26350591	1.23919725
H	-3.72176515	-0.57149271	1.40711309

H	-4.97199711	0.25127986	1.85568760
O	2.76159464	2.38202662	1.41229013
H	3.27367235	1.54634517	1.57145967
H	3.16989060	3.05563656	1.96351943
O	2.52685413	-2.46453621	-0.77925849
H	1.63876110	-2.24899979	-1.11640748
H	3.10657890	-1.74710268	-1.09353796
O	-2.52476503	2.46384789	0.77992184
H	-1.63736906	2.24746385	1.11841417
H	-3.10585832	1.74726359	1.09349706
O	-2.76186721	-2.38048819	-1.41506704
H	-3.27465488	-1.54522054	-1.57436672
H	-3.16870741	-3.05438894	-1.96703452
O	-2.79264028	-1.95698424	1.44172471
H	-1.85687894	-1.75682833	1.63115432
H	-2.78903010	-2.31202936	0.53726329
O	4.08043374	0.07727714	1.57432064
H	3.56805875	-0.70756435	1.83742253
H	4.32624834	-0.10262958	0.64945285
O	-0.11657254	-1.55258996	-1.14833997
H	-0.97416083	-1.94951222	-1.39379643
H	-0.16336722	-1.46208551	-0.17107250
O	2.79243517	1.95862738	-1.44410922
H	1.85680583	1.75649358	-1.63217101
H	2.78916008	2.31417338	-0.53981244
O	-4.08049979	-0.07673860	-1.57722005
H	-3.56777805	0.70839215	-1.83878183

```

H -4.32866081  0.10305124 -0.65302683
O -0.12623544 -1.14911860  1.58450896
H  0.71887096 -1.54099591  1.87520218
H  0.00519896 -0.17621894  1.59153759
O  2.43054159 -2.23102984  1.85838070
H  2.49447002 -2.42706168  0.88398964
H  2.68572680 -3.03885326  2.31351095
O  0.11727910  1.55092462  1.15243575
H  0.97484648  1.94864826  1.39674109
H  0.16355228  1.45911335  0.17525371

```

4 Hexamer data

Table 19: Water hexamer isomer intermolecular energies and many-body decomposition. Reference CCSD(T)-F12 and MB-pol model energies are from Medders *et al.*¹⁶ The columns n B show the n -body non-additive interaction energies, and the total interaction energy is given in column “2B-6B”. The sum of the terms of 3B to 6B is given in column “>2B”. All energies are in kJ mol^{-1} .

Model	2B-6B	2B	3B	4B	5B	6B	>2B
Prism							
CCSD(T)-F12	-200.957	-161.711	-36.735	-2.761	0.251	0.000	-39.245
MB-pol	-201.543	-163.008	-36.568	-2.175	0.209	0.000	-38.534
DIFF-L3pol	-202.522	-158.044	-40.359	-4.429	0.299	0.012	-44.478
DIFF-L2pol	-201.371	-158.025	-39.561	-4.154	0.357	0.012	-43.345
DIFF-L1pol	-201.491	-158.930	-38.532	-4.227	0.188	0.010	-42.560

Table 20: Water hexamer isomer intermolecular energies and many-body decomposition. See the caption to Table 19 for an explanation of the columns.

Model	2B-6B	2B	3B	4B	5B	6B	>2B
Cage							
CCSD(T)-F12	-199.869	-159.828	-37.823	-2.217	0.041	0.000	-39.999
MB-pol	-200.204	-161.000	-37.363	-1.966	0.125	0.000	-39.204
DIFF-L3pol	-199.435	-154.967	-40.915	-3.686	0.160	-0.025	-44.467
DIFF-L2pol	-197.694	-154.743	-39.758	-3.356	0.188	-0.024	-42.951
DIFF-L1pol	-198.644	-156.088	-39.077	-3.528	0.072	-0.021	-42.555

Table 21: Water hexamer isomer intermolecular energies and many-body decomposition. See the caption to Table 19 for an explanation of the columns.

Model	2B-6B	2B	3B	4B	5B	6B	>2B
Book-1							
CCSD(T)-F12	-198.070	-149.619	-43.764	-4.518	-0.167	0.000	-48.450
MB-pol	-196.648	-149.787	-42.927	-3.849	-0.041	0.000	-46.818
DIFF-L3pol	-197.651	-146.005	-44.903	-6.337	-0.355	-0.049	-51.645
DIFF-L2pol	-196.382	-145.936	-44.022	-6.068	-0.308	-0.044	-50.445
DIFF-L1pol	-198.735	-147.359	-44.480	-6.434	-0.409	-0.051	-51.376

Table 22: Water hexamer isomer intermolecular energies and many-body decomposition. See the caption to Table 19 for an explanation of the columns.

Model	2B-6B	2B	3B	4B	5B	6B	>2B
Book-2							
CCSD(T)-F12	-196.899	-150.038	-42.593	-4.184	-0.083	0.000	-46.860
MB-pol	-195.853	-150.331	-41.965	-3.556	0.000	0.000	-45.521
DIFF-L3pol	-196.080	-145.769	-43.950	-6.043	-0.277	-0.039	-50.310
DIFF-L2pol	-194.686	-145.725	-42.989	-5.715	-0.221	-0.034	-48.961
DIFF-L1pol	-196.131	-147.212	-42.741	-5.860	-0.280	-0.036	-48.919

Table 23: Water hexamer isomer intermolecular energies and many-body decomposition. See the caption to Table 19 for an explanation of the columns.

Model	2B-6B	2B	3B	4B	5B	6B	>2B
Bag							
CCSD(T)-F12	-195.016	-146.481	-43.639	-4.853	-0.083	0.041	-48.576
MB-pol	-193.719	-147.444	-42.467	-3.765	-0.041	0.000	-46.275
DIFF-L3pol	-195.552	-143.327	-44.970	-6.769	-0.531	0.047	-52.224
DIFF-L2pol	-193.862	-143.275	-43.878	-6.313	-0.441	0.047	-50.586
DIFF-L1pol	-195.347	-144.690	-43.828	-6.393	-0.481	0.044	-50.657

Table 24: Water hexamer isomer intermolecular energies and many-body decomposition. See the caption to Table 19 for an explanation of the columns.

Model	2B-6B	2B	3B	4B	5B	6B	>2B
Ring							
CCSD(T)-F12	-193.593	-135.645	-49.664	-7.447	-0.794	-0.041	-57.906
MB-pol	-190.915	-135.728	-48.701	-6.066	-0.418	0.000	-55.186
DIFF-L3pol	-192.801	-134.565	-47.909	-8.854	-1.340	-0.131	-58.236
DIFF-L2pol	-192.230	-134.785	-47.508	-8.563	-1.255	-0.118	-57.445
DIFF-L1pol	-197.511	-136.184	-49.872	-9.725	-1.557	-0.171	-61.327

Table 25: Water hexamer isomer intermolecular energies and many-body decomposition. See the caption to Table 19 for an explanation of the columns.

Model	2B-6B	2B	3B	4B	5B	6B	>2B
Cyclic-boat-1							
CCSD(T)-F12	-189.367	-133.971	-47.864	-6.819	-0.669	-0.041	-55.354
MB-pol	-187.275	-133.971	-47.279	-5.648	-0.376	0.000	-53.304
DIFF-L3pol	-187.391	-131.902	-46.071	-8.154	-1.149	-0.112	-55.488
DIFF-L2pol	-187.222	-132.176	-45.916	-7.938	-1.087	-0.102	-55.046
DIFF-L1pol	-191.391	-133.581	-47.585	-8.783	-1.301	-0.139	-57.809

Table 26: Water hexamer isomer intermolecular energies and many-body decomposition. See the caption to Table 19 for an explanation of the columns.

Model	2B-6B	2B	3B	4B	5B	6B	>2B
Cyclic-boat-2							
CCSD(T)-F12	-188.949	-133.720	-47.823	-6.736	-0.669	-0.041	-55.228
MB-pol	-187.275	-133.971	-47.237	-5.648	-0.376	0.000	-53.262
DIFF-L3pol	-187.961	-132.491	-46.165	-8.050	-1.141	-0.112	-55.469
DIFF-L2pol	-187.691	-132.736	-45.895	-7.865	-1.089	-0.103	-54.954
DIFF-L1pol	-192.553	-134.151	-48.133	-8.805	-1.320	-0.141	-58.401

5 Energies for 16-mers and 24-mers

Table 27: Decomposition of many-body energies up to four-body contributions, for each model using DIFF (i.e. $x = 1$) damping compared with the SAMBA energies from Góra *et al.*¹⁵ The final column gives the 3B non-additive energies where the molecular properties have been replaced by those from the water monomer in a conformation optimized using CCSD(T)/cc-pVTZ using the Psi4 program.

Model	2B	3B	4B	5B	>5B	>4B	2B-4B	3B(opt.geom.)
4444-a								
SAMBA	-572.898	-135.662	-3.837	—	—	—	-712.397	—
DIFF-L3pol	-547.526	-155.976	-9.685	3.364	-0.245	3.119	-713.187	-151.447
DIFF-L2pol	-546.715	-148.381	-8.565	3.175	-0.301	2.875	-703.661	-144.902
DIFF-L1pol	-550.645	-157.518	-14.594	1.521	-0.320	1.201	-722.757	-153.392
4444-b								
SAMBA	-566.288	-141.013	-3.933	—	—	—	-711.234	—
DIFF-L3pol	-541.546	-160.606	-11.721	2.586	-0.161	2.424	-713.873	-155.949
DIFF-L2pol	-540.797	-152.345	-9.809	2.756	-0.159	2.596	-702.951	-148.789
DIFF-L1pol	-545.097	-162.315	-15.100	1.199	-0.181	1.018	-722.512	-158.058
boat-a								
SAMBA	—	—	—	—	—	—	—	—
DIFF-L3pol	-531.955	-167.257	-17.023	1.523	0.097	1.620	-716.235	-162.406
DIFF-L2pol	-531.351	-159.495	-15.345	1.670	0.077	1.746	-706.191	-155.735
DIFF-L1pol	-536.249	-166.501	-18.932	0.478	0.03	0.508	-721.682	-162.093
boat-b								
SAMBA	-556.493	-152.206	-8.481	—	—	—	-717.18	—
DIFF-L3pol	-530.607	-168.049	-17.790	1.220	0.164	1.384	-716.446	-163.180
DIFF-L2pol	-530.093	-160.155	-15.766	1.528	0.16	1.688	-706.014	-156.384
DIFF-L1pol	-535.048	-167.883	-18.851	0.561	0.179	0.740	-721.782	-163.425
anti-boat								
SAMBA	-553.288	-150.498	-11.979	—	—	—	-715.765	—
DIFF-L3pol	-530.969	-166.678	-18.956	1.207	0.288	1.495	-716.603	-161.912
DIFF-L2pol	-530.696	-159.556	-17.364	1.292	0.204	1.496	-707.616	-155.802
DIFF-L1pol	-534.791	-170.783	-23.615	-0.677	0.039	-0.638	-729.189	-166.228

Table 28: Decomposition of interaction energies for two variants of the (H₂O)₂₄ tetradecahedron isomers and the differences between them, using the DIFF models. We also present the CCpol23+, CC-pol-8s+NB and reference SAMBA results from Góra *et al.*¹⁵ In the DIFF-L3pol(opt) rows we present energies obtained using the DIFF-L3pol models with molecular properties (multipoles and polarizabilities) replaced with those from the water monomer in a conformation optimized using CCSD(T)/cc-pVTZ using the Psi4 program.

Model	2B	3B	4B	5B	>5B	2B-4B	2B-∞B
ISOMER 316							
SAMBA	−801.370	−246.864	−29.292			−1077.526	
DIFF-L3pol	−769.749	−257.040	−43.971	−2.921	0.628	−1070.760	−1073.053
DIFF-L3pol(opt)		−249.883	−42.548	−2.809			
DIFF-L2pol	−769.741	−244.892	−39.843	−1.996	0.676	−1054.476	−1055.796
DIFF-L1pol	−778.151	−259.324	−45.453	−3.629	0.528	−1082.927	−1086.028
CCpol-8s-NB	−769.254	−191.104	−27.882	−1.460	0.360	−988.240	
CCpol23+	−769.387	−214.434	−27.188	−1.339	0.314	−1011.009	
ISOMER 308							
SAMBA	−808.048	−241.798	−28.305			−1078.151	
DIFF-L3pol	−772.943	−255.382	−43.003	−2.473	0.728	−1071.328	−1073.074
DIFF-L3pol(opt)		−248.242	−41.597	−2.361			
DIFF-L2pol	−772.855	−243.089	−38.998	−1.667	0.707	−1054.941	−1055.902
DIFF-L1pol	−781.169	−255.401	−45.545	−3.772	0.439	−1082.115	−1085.448
CCpol-8s-NB	−773.203	−185.623	−30.882	−2.556	0.079	−989.708	
CCpol23+	−773.576	−210.497	−28.125	−1.929	0.126	−1012.197	
Difference: 316 - 308							
SAMBA	6.678	−5.067	−0.987			0.625	
DIFF-L3pol	3.194	−1.658	−0.968	−0.448	−0.100	0.568	0.021
DIFF-L3pol(opt)		−1.641	−0.951	−0.448		0.463	
DIFF-L2pol	3.114	−1.803	−0.845	−0.329	−0.031	0.465	0.106
DIFF-L1pol	3.018	−3.923	0.092	0.143	0.089	−0.812	−0.580
CCpol-8s-NB	3.954	−5.481	2.996	1.092	0.280	1.469	
CCpol23+	4.188	−3.941	0.937	0.590	0.188	1.188	

References

- (1) Misquitta, A. J.; Stone, A. J. Ab Initio Atom-Atom Potentials Using CamCASP: Theory and Application to Many-Body Models for the Pyridine Dimer. *J. Chem. Theory Comput.* **2016**, *12*, 4184–4208, PMID: 27467814.
- (2) Van Vleet, M. J.; Misquitta, A. J.; Stone, A. J.; Schmidt, J. R. Beyond Born–Mayer: Improved Models for Short-Range Repulsion in Ab Initio Force Fields. *J. Chem. Theory Comput.* **2016**, *12*, 3851–3870.
- (3) Van Vleet, M. J.; Misquitta, A. J.; Schmidt, J. R. New Angles on Standard Force Fields: Toward a General Approach for Treating Atomic-Level Anisotropy. *J. Chem. Theory Comput.* **2018**, *14*, 739–758.
- (4) Stone, A. J. *The Theory of Intermolecular Forces*, 2nd ed.; Oxford University Press, Oxford, 2013.
- (5) Tang, K. T.; Toennies, J. P. An improved simple model for the van der Waals potential based on universal damping functions for the dispersion coefficients. *J. Chem. Phys.* **1984**, *80*, 3726–3741.
- (6) Misquitta, A. J.; Stone, A. J. Accurate induction energies for small organic molecules: I. Theory. *J. Chem. Theory Comput.* **2008**, *4*, 7–18.
- (7) Misquitta, A. J.; Stone, A. J.; Price, S. L. Accurate induction energies for small organic molecules: II. Models and Numerical details. *J. Chem. Theory Comput.* **2008**, *4*, 19–32.
- (8) Tang, K. T.; Toennies, J. P. The damping function of the van der Waals attraction in the potential between rare gas atoms and metal surfaces. *Surf. Sci. Lett.* **1992**, *279*, 203–206.
- (9) Stone, A. J.; Dullweber, A.; Engkvist, O.; Fraschini, E.; Hodges, M. P.; Meredith, A. W.; Nutt, D. R.; Popelier, P. L. A.; Wales, D. J. ORIENT: a program for studying interactions

between molecules, version 5.0. University of Cambridge, 2019; <http://www-stone.ch.cam.ac.uk/programs.html\#Orient>, Accessed: Aug 2019.

- (10) Mas, E. M.; Bukowski, R.; Szalewicz, K.; Groenenboom, G. C.; Wormer, P. E. S.; van der Avoird, A. Water pair potential of near spectroscopic accuracy. I. Analysis of potential surface and virial coefficients. *J. Chem. Phys.* **2000**, *113*, 6687–6701.
- (11) Liu, C.; Qi, R.; Wang, Q.; Piquemal, J.-P.; Ren, P. Capturing many-body interactions with classical dipole induction models. *J. Chem. Theory Comput.* **2017**, *13*, 2751–2761.
- (12) Akin-Ojo, O.; Szalewicz, K. How Well Can Polarization Models of Pairwise Nonadditive Forces Describe Liquid Water? *J. Chem. Phys.* **2013**, *138*, 024316.
- (13) Bates, D. M.; Tschumper, G. S. CCSD (T) complete basis set limit relative energies for low-lying water hexamer structures. *J. Phys. Chem. A* **2009**, *113*, 3555–3559.
- (14) Yoo, S.; Xantheas, S. S. Structures, energetics, and spectroscopic fingerprints of water clusters n= 2–24. *Handbook of Computational Chemistry* **2017**, 1139–1173.
- (15) Góra, U.; Cencek, W.; Podeszwa, R.; van der Avoird, A.; Szalewicz, K. Predictions for Water Clusters from a First-Principles Two- and Three-Body Force Field. *J. Chem. Phys.* **2014**, *140*, 194101.
- (16) Medders, G. R.; Götz, A. W.; Morales, M. A.; Bajaj, P.; Paesani, F. On the Representation of Many-Body Interactions in Water. *J. Chem. Phys.* **2015**, *143*, 104102.