OPLS3e: Extending Force Field Coverage for Drug-Like Small Molecules: Supplementary Information

Katarina Roos[2,3], Chuanjie Wu[1], Wolfgang Damm[1], Mark Reboul[1], James M. Stevenson[1], Chao Lu[1], Markus K. Dahlgren[1], Sayan Mondal[1], Wei Chen[1], Lingle Wang[1], Robert Abel[1], Richard A. Friesner[2], Edward D. Harder[1]

[1]Schrodinger, Inc., 120 West 45th Street, New York, New York 10036, United States, [2]Department of Chemistry, Columbia University, 3000 Broadway, New York, New York 10027, United States, [3]Department of Cell and Molecular Biology, Uppsala University, Biomedical Centre, BOX 596, SE-751 24 Uppsala, Sweden

Corresponding Author: ed.harder@schrodinger.com

1. Additional details regarding the charge model development

OPLS3e partial charges are assigned by minimizing the following objective function,

$$\chi = \sum_k b_k (\varphi_k^{QM} - \varphi_k^{MM})^2 + \sum_{ij} a_{ij} (q_{ij} - q_{ij}^0)^2 \quad \text{(S1)}$$

where the first term is the electrostatic potential difference between QM and MM summed over a set of grid points surrounding the molecule. The second term is the difference in bond-charge increment between atoms i and j relative to the starting reference parameter q^0_{ij}. The weight a_{ij} is 100 e^{-1} for most chemical moieties and 10000 e^{-1} for a select group of functional groups shown below.

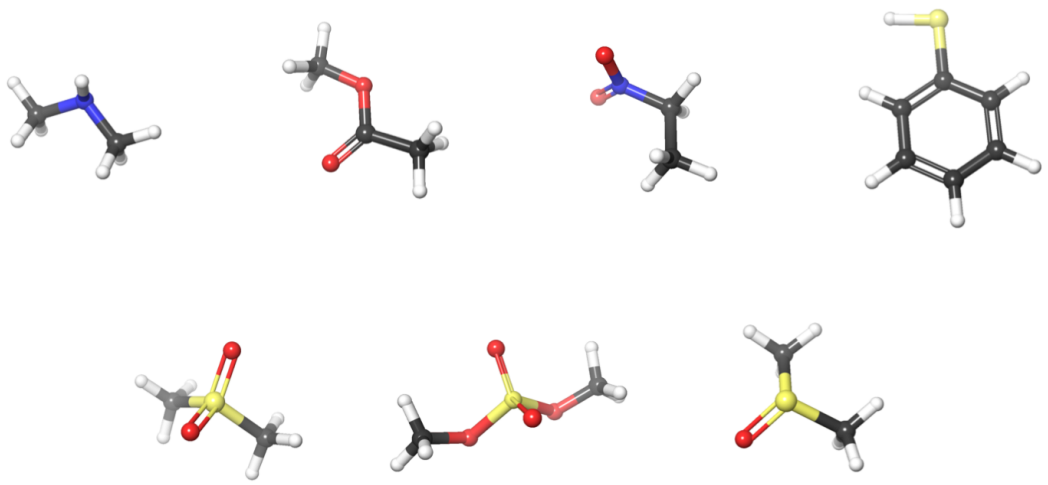


Figure S1. Illustration of chemical groups assigned higher restraints to reference BCIs in OPLS3e; amine, formate, nitro, aryl thiol, sulfone, sulfate, and sulfoxide groups.

2. Additional details regarding the torsion model development

Table S1. RMS Error comparison of candidate QM reference levels with respect to 13759 CCSD(T)/CBS single point energies calculated with Equation S2 for 1156 torsion profiles taken from B3LYP/6-31G* geometries.

Model	RMSE (kcal/mol)
LMP2/cc-pVTZ(-f)	0.58
M06-2X/cc-pVTZ(-f)	0.32

2a. The CCSD(T)/CBS limit calculation for torsional profiles

$$E_{CCSD(T)/CBS} = E_{HF,CBS} + E_{MP2_correlation,CBS} + \Delta_{CCSD(T)-MP2} \quad \text{(S2)}$$

$$E_{HF,CBS} = \frac{e^{a \cdot E_{HF,aVTZ} - E_{HF,aVDZ}}}{e^{a}-1},\ (a = 1.39873) \quad \text{(S3)}$$

$$E_{MP2_correction,CBS} = \frac{27 \cdot E_{MP2_correlation,aVTZ} - 8 \cdot E_{MP2_{correlation,aVDZ}}}{27-8} \quad \text{(S4)}$$

$$\Delta_{CCSD(T)-MP2} = E_{CCSD(T)//6-31gm} - E_{MP2//6-31gm} \quad \text{(S5)}$$

The coefficients in Equations S2~S5 were derived by using the two-point forms proposed by Helgaker *et al*. [1] with aVDZ and aVTZ basis sets. The 6-31gm basis set is adapted from 6-31g* for broader coverage of elements, and is shown below.

2b. The 6-31gm basis set

```
basis={
default=6-31g*
! HYDROGEN
s, H , 18.7311370, 2.8253944, 0.64012169, 0.16127776
c, 1.3, 0.033494604, 0.23472695, 0.81375733
c, 4.4, 1
! CARBON
s, C , 3047.5249000, 457.3695200, 103.9486800, 29.2101550,
9.2866630, 3.1639270, 7.8682723, 1.8812885, 0.54424926,
0.16871448
c, 1.6, 0.0018347371, 0.014037323, 0.068842622, 0.23218444,
0.46794135, 0.36231199
c, 7.9, -0.11933242, -0.16085415, 1.1434564
c, 10.10, 1.0000000
```

```
p, C , 7.8682723, 1.8812885, 0.54424926, 0.16871448
c, 1.3, 0.068999067, 0.31642396, 0.74430829
c, 4.4, 1.0000000
d, C , 0.2500000
c, 1.1, 1
! NITROGEN
s, N , 4173.5115000, 627.4579100, 142.9020900, 40.2343290,
12.8202130, 4.3904370, 11.6263620, 2.7162798, 0.77221840,
0.2120315
c, 1.6, 0.0018347722, 0.013994627, 0.068586552, 0.23224087,
0.46906995, 0.36045520
c, 7.9, -0.11496118, -0.16911748, 1.1458519
c, 10.10, 1.0000000
p, N , 11.6263620, 2.7162798, 0.77221840, 0.2120315
c, 1.3, 0.067579744, 0.32390730, 0.74089514
c, 4.4, 1.0000000
d, N , 0.2500000
c, 1.1, 1
! OXYGEN
s, O , 5484.6717000, 825.2349500, 188.0469600, 52.9645000,
16.8975700, 5.7996353, 15.5396160, 3.5999336, 1.0137618,
0.27000582
c, 1.6, 0.0018310744, 0.013950172, 0.068445078, 0.23271434,
0.47019290, 0.35852085
c, 7.9, -0.11077755, -0.14802626, 1.1307670
c, 10.10, 1.0000000
p, O , 15.5396160, 3.5999336, 1.0137618, 0.27000582
c, 1.3, 0.070874268, 0.33975284, 0.72715858
c, 4.4, 1.0000000
d, O , 0.2500000
c, 1.1, 1
! CALCIUM
s,Ca,35264.86,5295.503,1206.02,339.6839,108.6264,36.92103,706.309
6,167.81870,53.825580,20.01638,7.970279,3.212059,14.19518,4.88082
8,2.16039,.98789
9,.449517,.187387,1.032271,.381171,.0651310,.02601
c,1.6,.0018135010,.01388493,.06836162,.2356188,.4820639,.3429819
c,7.12,.002448225,.03241504,.1226219,-.04316965,-.61269950,-
.448754
c,13.18,.010845,.2088333,.031503380,-.5526518,-.5437997,-
.06669342
c,19.21,-.0443972,-.3284563,1.16301
c,22.22,1
p,Ca,706.3096,167.81870,53.825580,20.01638,7.970279,3.212059,14.1
9518,4.880828,2.16039,.987899,.449517,.187387,1.032271,.381171,.0
651310,.02601
c,1.6,.004020371,.03100601,.1337279,.3367983,.4631281,.2257532
c,7.12,-.01289621,-.01025198,.1959781,.4357933,.3996452,.09713636
c,13.15,-.4298621,.006935829,.9705933
c,16.16,1
```

```
! BROMINE ECP10MDF_VDZ
ecp,Br,10,4,3
1
2,1.000000,.000000
3
2,70.024257,49.962834
2,31.178412,370.014205
2,7.156593,10.241439
4
2,46.773471,99.112244
2,46.184120,198.253046
2,21.713858,28.261740
2,20.941792,56.623366
6
2,50.698839,-18.605853
2,50.644764,-27.923280
2,15.447509,-.379693
2,15.500259,-.780583
2,2.800391,.035968
2,1.077480,.094397
2
2,14.465606,-1.091269
2,21.234065,-2.887691
4
2,46.773471,-198.224488
2,46.184120,198.253046
2,21.713858,-56.523480
2,20.941792,56.623366
6
2,50.698839,18.605853
2,50.644764,-18.615520
2,15.447509,.379693
2,15.500259,-.520389
2,2.800391,-.035968
2,1.077480,.062931
2
2,14.465606,.727513
2,21.234065,-1.443846
s,Br,2808.60,421.180,50.3457,17.9133,3.80531,1.74968,.448555,.164
498
c,1.8,0.001606,.008393,.069578,-.389908,.694497,.491354,.022637,-
0.003723
c,1.8,-.000635,-.003492,-.025195,.150113,-.366226,-
.383422,0.714468,.535253
c,7.7,1
c,8.8,1
!id=ECP10MDF_VDZ elementType=Br minL=1 maxL=1
p,Br,105.752,27.6368,6.59656,2.78522,1.07812,.393537,.127469
c,1.7,.005341,-0.083084,.447766,.550617,.123500,-.003771,.002278
c,1.7,-.001308,0.022921,-.145029,-.209037,.093730,.605021,.457123
```

```
c,7.7,1
!id=ECP10MDF_VDZ elementType=Br minL=2 maxL=2
d,Br,143.865,46.1163,17.3694,6.95107,2.75607,1.01178,.4291
c,1.7,.010237,.076083,0.229807,.403347,.409728,.162790,.006991
c,7.7,1
! IODINE ECP28MDF_VDZ
ecp,I,28,4,3
1
2,1.000000,.000000
3
2,40.015835,49.994293
2,17.429747,281.025317
2,9.005484,61.573326
4
2,15.355466,67.442841
2,14.971833,134.881137
2,8.960164,14.675051
2,8.259096,29.375666
6
2,15.068908,35.439529
2,14.555322,53.176057
2,6.718647,9.067195
2,6.456393,13.206937
2,1.191779,.089335
2,1.291157,.052380
4
2,19.458609,-21.842040
2,19.349260,-28.468191
2,4.823767,-.243713
2,4.884315,-.320804
4
2,15.355466,-134.885683
2,14.971833,134.881137
2,8.960164,-29.350101
2,8.259096,29.375666
6
2,15.068908,-35.439529
2,14.555322,35.450704
2,6.718647,-9.067195
2,6.456393,8.804624
2,1.191779,-.089335
2,1.291157,.034920
4
2,19.458609,14.561360
2,19.349260,-14.234095
2,4.823767,.162475
2,4.884315,-.160402
!id=ECP28MDF_VDZ elementType=I minL=0 maxL=0
s,I,2437.23,358.331,14.5856,9.18532,2.10199,1.03944,.317845,.122187
```

```
c,1.8,0.000430,.002292,.393905,-.925011,.935200,.391057,.012428,-
0.001314
c,1.8,.000179,.001077,.166524,-.415575,.630059,.324446,-
0.811965,-.492687
c,7.7,1
c,8.8,1
!id=ECP28MDF_VDZ elementType=I minL=1 maxL=1
p,I,18.3299,11.4794,2.75125,1.21708,.342164,.111445
c,1.6,.083100,-.258792,0.665756,.451823,.029025,-.002953
c,1.6,-.026616,.088347,-.301465,-0.151315,.605685,.549572
c,6.6,1
!id=ECP28MDF_VDZ elementType=I minL=2 maxL=2
d,I,51.2059,15.7468,4.51661,2.04408,.869864,.3013
c,1.6,.004042,-.004992,0.293292,.512688,.310977,.041872
c,6.6,1
}
```

3. Additional details regarding the ASP228 predicted pKa shift calculation

To calculate the pKa shift of Asp288, we performed FEP simulations between protonated and deprotonated forms of Asp using the charge FEP protocol with the post-simulation charge correction method [2]. One FEP simulation is run for model Asp in solvent with both termini capped, giving the deprotonation free energy of model Asp, ΔG(model). The other simulations are run for Asp228 in the protein-ligand complex, giving the deprotonation free energy of Asp228 in complex, ΔG(complex). In these simulations, both the ligand and Asp32 were fixed at the charged states. The reason is that Asp32 interacts strongly with the positively charged moiety of the ligand, stabilizing their charged states. The difference in the deprotonation free energy between model Asp and complex is related to the pKa shift of Asp228 as ΔG(complex) - ΔG(model) = RT*ln10*(pKa_complex – pKa_model). Using this formula with the known pKa of model Asp, 4.0, we obtained the pKa of Asp228 in the complexes. The FEP simulations for model Asp and the complexes last 30 and 100 ns, respectively. The first 50 ns of the

complex simulations were excluded from free energy calculation to remove the initial equilibration stage.

Using the pKa values of Asp228 in the complexes, we applied the pKa correction method [3] to the binding free energy according to

$$\Delta\Delta G = \Delta\Delta G_{D228-} - RT \ln \frac{1+10^{pK_a^{L23}-\mathrm{pH}}}{1+10^{pK_a^{L11}-\mathrm{pH}}}, \quad \text{(S6)}$$

where ΔΔGD228- is the binding free energy when Asp228 is fixed at its deprotonated state, pKaL11 and pKaL23 are the pKa values of Asp228 in the complexes of ligands 11 and 23, respectively, R is the ideal gas constant, and T is temperature.

4. OPLS3e parameters

Parameters for the molecules in Figure 1 Panel B of the manuscript and molecules covering the solvation free energy and protein-ligand binding free energy calculations are provided in opls3e_si.zip. Virtual site definitions (LC2, LC3) follow from the definitions described in the *Desmond Users Guide (Release 3.6.1.1/0.8) D. E. Shaw Research*.

5. Hydration and transfer free energies

Starting structures, OPLS3e parameters, experimental and calculated small molecule hydration free energies, transfer free energies from water to chloroform and transfer free energies from water to toluene are provided in opls3e_si.zip

6. Protein-ligand binding FEP

Starting structures, OPLS3e parameters, experimental and calculated protein-ligand binding FEP free energies for all presented public data sets are provided in opls3e_si.zip.

Table S2. Protein structures used in the protein-ligand FEP calculations.

System	Series	PDB id
BACE		4djw
CDK2		1h1q
JNK1		2gmx
MCL1		4hw3
P38		3fln
PTP1B		2qbs
Thrombin		2zff
Tyk2		4gih
FXA	1	3iit
	2	3iit
	3	2vvv
	4	2j95
	5	3iit
	6	2j95
CHK1	1	2ayp
	2	2hy0
	3	2ydk
	4	2ydj
	5	3pa3
	6	3u9n
	7	2e9v
BACE core	1	4djw

	2	4djw
BACE heterocycles	1	3nsh
	2	3nsh
	3	3nsh

References

(1) Helgaker, T.; Ruden, T. A.; Jørgensen, P.; Olsen, J.; Klopper, W. A Priori Calculation of Molecular Properties to Chemical Accuracy. *J. Phys. Org. Chem.* **2004**, *17* (11), 913–933.

(2) Chen, W.; Deng, Y.; Russell, E.; Wu, Y.; Abel, R.; Wang, L. Accurate calculation of relative binding free energies between ligands with different net charges. *(Submitted)*

(3) de Oliveira, C.; Yu, H. S.; Chen, W.; Abel, R.; Wang, L. A rigorous free energy perturbation approach to estimating relative binding affinities between ligands with multiple protonation and tautomeric states. *(Submitted)*