Supplementary Information: Benchmarking Quantum Mechanical Levels of Theory for Valence Parametrization in Force Fields

Pavan Kumar Behara,[*,†] Hyesu Jang,[‡,¶] Joshua T. Horton,[§] Trevor Gokey,[‖] David L. Dotson,[⊥,#] Simon Boothroyd,[@] Christopher I. Bayly,[¶] Daniel J. Cole,[§] Lee-Ping Wang,[‡] and David L. Mobley[†,‖]

†*Department of Pharmaceutical Sciences, University of California, Irvine, California 92697, USA*

‡*Chemistry Department, The University of California at Davis, Davis, California, 95616, USA*

¶*OpenEye Scientific Software, Santa Fe, New Mexico 87508, USA*

§*School of Natural and Environmental Sciences, Newcastle University, Newcastle upon Tyne NE1 7RU, UK*

‖*Department of Chemistry, University of California, Irvine, California 92697, USA*

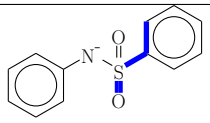
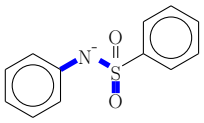
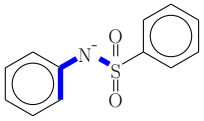
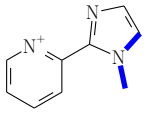
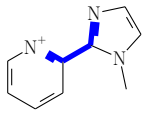
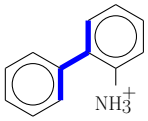
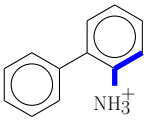
⊥*The Open Force Field Initiative, Open Molecular Software Foundation, Davis, California 95616, USA*

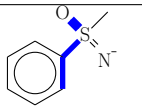
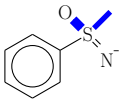
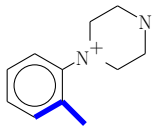
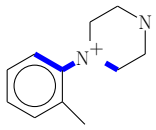
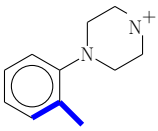
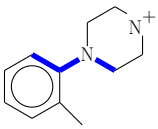
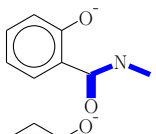
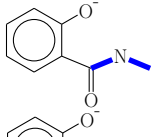
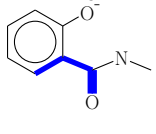
#*Datryllic LLC, Phoenix, Arizona 85003, USA*

@*Boothroyd Scientific Consulting Ltd., London WC2H 9JQ, UK.*

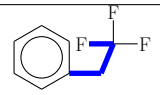
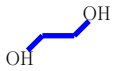
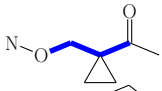
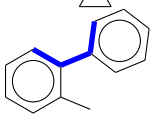
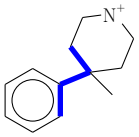
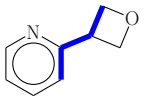
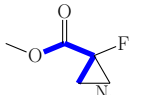
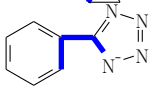
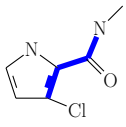
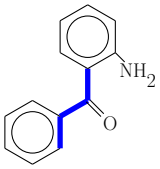
E-mail: pbehara@uci.edu

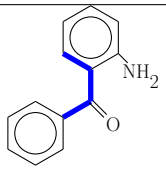
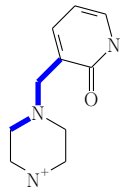
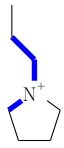
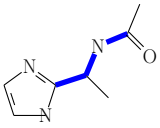
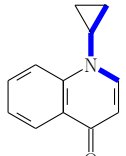
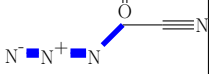
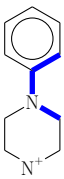
SI 1.1 Individual errors per molecule

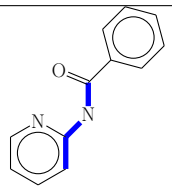
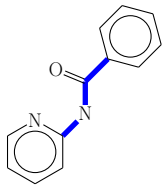
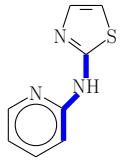
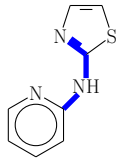
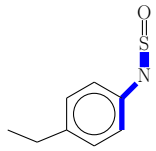
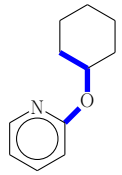
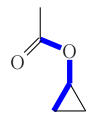
Index	Molecule	MP2/heavy-aug-cc-pVTZ	B97-D3BJ/def2-TZVP	M05-2X-D3/DZVP	M06-2X-D3/DZVP	M08-HX-D3/DZVP	ωB97M-D3BJ/DZVP	ωB97M-V/DZVP	PW6B95-D3BJ/DZVP	PW6B95-D3/DZVP	MP2/aug-cc-pVTZ	DSD-BLYP-D3BJ/Heavy-aug-cc-pVTZ	**B3LYP-D3BJ/DZVP**	B3LYP-NL/DZVP	B3LYP-D3MBJ/DZVP	B3LYP-D3BJ/def2-TZVP	B3LYP-D3BJ/def2-TZVPD	B3LYP-D3BJ/def2-TZVPP	B3LYP-D3BJ/def2-TZVPPD	B3LYP-D3BJ/QZVP	B3LYP-D3BJ/6-31+G**	B3LYP-D3BJ/6-311+G**
1		0.64	0.14	0.38	0.22	0.28	0.20	0.27	0.16	0.16	0.54	0.29	**0.20**	0.20	0.17	0.20	0.16	0.19	0.16	0.19	0.20	0.31
2		1.51	0.48	0.20	0.20	0.25	0.21	0.23	0.31	0.38	1.65	0.73	**0.35**	0.55	0.29	0.40	0.34	0.40	0.34	0.36	0.50	0.74
3		0.37	0.37	0.25	0.20	0.25	0.11	0.22	0.22	0.23	0.26	0.27	**0.22**	0.27	0.28	0.40	0.35	0.41	0.35	0.34	0.62	0.35
4		0.91	0.50	0.37	0.57	0.88	0.12	0.22	0.69	0.67	1.12	0.77	**0.86**	0.87	0.92	0.61	0.37	0.59	0.36	0.42	0.53	0.36
5		0.66	0.37	0.31	0.46	0.58	0.09	0.25	0.51	0.52	0.84	0.61	**0.70**	0.66	0.77	0.47	0.41	0.46	0.41	0.43	0.52	0.33
6		0.17	0.03	0.28	0.28	0.23	0.17	0.14	0.17	0.20	0.16	0.13	**0.10**	0.08	0.12	0.04	0.06	0.03	0.05	0.05	0.05	0.01
7		0.96	0.90	0.21	0.26	0.40	0.58	0.61	0.70	0.68	1.03	0.87	**0.65**	0.78	0.73	0.80	0.78	0.85	0.80	0.81	0.55	0.60
8		0.19	0.16	0.83	0.44	0.40	0.70	0.81	0.17	0.18	0.13	0.07	**0.10**	0.06	0.07	0.08	0.10	0.12	0.12	0.12	0.15	0.20
9		0.06	0.26	0.22	0.23	0.12	0.20	0.21	0.04	0.07	0.15	0.09	**0.02**	0.01	0.02	0.15	0.16	0.16	0.15	0.14	0.13	0.11

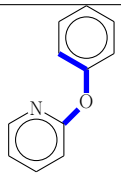
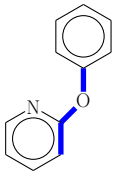
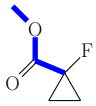
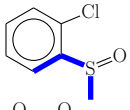
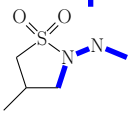
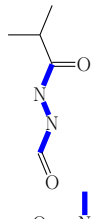
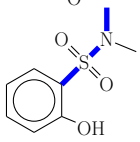
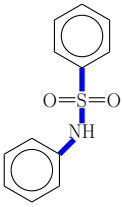
Index	Molecule	MP2/heavy-aug-cc-pVTZ	B97-D3BJ/def2-TZVP	M05-2X-D3/DZVP	M06-2X-D3/DZVP	M08-HX-D3/DZVP	ωB97M-D3BJ/DZVP	ωB97M-V/DZVP	PW6B95-D3BJ/DZVP	PW6B95-D3/DZVP	MP2/aug-cc-pVTZ	DSD-BLYP-D3BJ/Heavy-aug-cc-pVTZ	**B3LYP-D3BJ/DZVP**	B3LYP-NL/DZVP	B3LYP-D3MBJ/DZVP	B3LYP-D3BJ/def2-TZVP	B3LYP-D3BJ/def2-TZVPD	B3LYP-D3BJ/def2-TZVPP	B3LYP-D3BJ/def2-TZVPPD	B3LYP-D3BJ/QZVP	B3LYP-D3BJ/6-31+G**	B3LYP-D3BJ/6-311+G**
10		0.10	0.89	1.30	1.06	1.18	1.07	1.14	0.56	0.60	0.18	0.05	**0.40**	0.44	0.39	0.32	0.46	0.34	0.45	0.40	0.15	0.24
11		0.06	0.19	0.07	0.11	0.08	0.06	0.05	0.07	0.04	0.08	0.03	**0.14**	0.12	0.12	0.08	0.09	0.09	0.08	0.08	0.09	0.14
12		0.11	0.15	0.17	0.13	0.08	0.08	0.06	0.02	0.02	0.16	0.03	**0.09**	0.09	0.08	0.15	0.12	0.15	0.13	0.14	0.16	0.17
13		0.68	0.40	0.52	0.22	0.27	0.30	0.35	0.17	0.20	0.52	0.49	**0.38**	0.30	0.31	0.40	0.42	0.40	0.40	0.39	0.43	0.38
14		0.06	0.09	0.08	0.06	0.09	0.05	0.05	0.08	0.08	0.07	0.08	**0.02**	0.02	0.02	0.10	0.10	0.11	0.10	0.10	0.06	0.08
15		0.18	0.91	0.60	0.45	0.31	0.61	0.66	0.43	0.47	0.24	0.23	**0.64**	0.44	0.62	0.66	0.67	0.66	0.66	0.66	0.60	0.63
16		1.04	0.33	0.21	0.74	0.58	0.88	0.73	0.46	0.56	0.99	0.83	**0.53**	0.45	0.53	0.36	0.44	0.38	0.44	0.39	0.53	0.36
17		0.03	0.10	0.06	0.02	0.06	0.04	0.06	0.12	0.09	0.08	0.02	**0.06**	0.06	0.05	0.08	0.04	0.07	0.03	0.02	0.03	0.06
18		0.95	0.41	0.46	0.33	0.33	0.45	0.26	0.42	0.41	1.02	1.03	**0.53**	0.94	0.61	0.64	0.48	0.63	0.48	0.47	1.01	0.43

Index	Molecule	MP2/heavy-aug-cc-pVTZ	B97-D3BJ/def2-TZVP	M05-2X-D3/DZVP	M06-2X-D3/DZVP	M08-HX-D3/DZVP	ωB97M-D3BJ/DZVP	ωB97M-V/DZVP	PW6B95-D3BJ/DZVP	PW6B95-D3/DZVP	MP2/aug-cc-pVTZ	DSD-BLYP-D3BJ/Heavy-aug-cc-pVTZ	**B3LYP-D3BJ/DZVP**	B3LYP-NL/DZVP	B3LYP-D3MBJ/DZVP	B3LYP-D3BJ/def2-TZVP	B3LYP-D3BJ/def2-TZVPD	B3LYP-D3BJ/def2-TZVPP	B3LYP-D3BJ/def2-TZVPPD	B3LYP-D3BJ/QZVP	B3LYP-D3BJ/6-31+G**	B3LYP-D3BJ/6-311+G**
19		0.62	0.45	0.14	0.19	0.56	0.17	0.21	0.50	0.49	0.72	0.65	**0.59**	0.65	0.61	0.59	0.56	0.61	0.56	0.56	0.58	0.47
20		0.10	0.23	0.11	0.08	0.04	0.02	0.04	0.05	0.02	0.12	0.04	**0.13**	0.09	0.12	0.14	0.13	0.13	0.12	0.13	0.13	0.18
21		0.76	1.20	0.19	0.18	0.51	0.14	0.19	0.73	0.54	1.04	0.82	**1.36**	1.36	1.37	0.96	0.71	0.93	0.72	0.78	0.91	0.81
22		1.25	0.80	0.22	0.20	0.44	0.24	0.32	0.60	0.43	1.64	1.02	**1.24**	1.24	1.28	0.72	0.55	0.69	0.56	0.59	0.87	0.68
23		0.58	1.08	0.40	0.20	0.69	0.35	0.34	0.78	0.63	0.71	0.82	**1.17**	1.29	1.14	1.14	0.88	1.14	0.88	0.95	0.93	0.80
24		0.40	0.29	0.62	0.46	0.38	0.14	0.20	0.14	0.24	0.25	0.26	**0.20**	0.13	0.19	0.18	0.21	0.20	0.21	0.21	0.24	0.26

Index	Molecule	MP2/heavy-aug-cc-pVTZ	B97-D3BJ/def2-TZVP	M05-2X-D3/DZVP	M06-2X-D3/DZVP	M08-HX-D3/DZVP	ωB97M-D3BJ/DZVP	ωB97M-V/DZVP	PW6B95-D3BJ/DZVP	PW6B95-D3/DZVP	MP2/aug-cc-pVTZ	DSD-BLYP-D3BJ/Heavy-aug-cc-pVTZ	**B3LYP-D3BJ/DZVP**	B3LYP-NL/DZVP	B3LYP-D3MBJ/DZVP	B3LYP-D3BJ/def2-TZVP	B3LYP-D3BJ/def2-TZVPD	B3LYP-D3BJ/def2-TZVPP	B3LYP-D3BJ/def2-TZVPPD	B3LYP-D3BJ/QZVP	B3LYP-D3BJ/6-31+G**	B3LYP-D3BJ/6-311+G**
25		0.19	0.18	0.15	0.20	0.12	0.07	0.04	0.05	0.04	0.08	0.11	**0.16**	0.09	0.14	0.13	0.17	0.15	0.18	0.16	0.15	0.17
26		0.15	0.16	0.68	0.45	0.37	0.36	0.30	0.30	0.42	0.11	0.13	**0.38**	0.31	0.38	0.12	0.17	0.10	0.16	0.15	0.40	0.34
27		0.75	0.34	0.49	0.63	0.70	0.35	0.31	0.28	0.40	0.95	0.44	**0.25**	0.22	0.31	0.13	0.15	0.12	0.15	0.16	0.17	0.16
28		0.33	0.28	0.42	0.28	0.23	0.23	0.23	0.29	0.36	0.48	0.27	**0.40**	0.33	0.37	0.36	0.34	0.36	0.34	0.35	0.42	0.36
29		0.64	0.20	0.58	0.35	0.43	0.27	0.29	0.30	0.33	0.65	0.36	**0.20**	0.13	0.22	0.14	0.14	0.15	0.14	0.14	0.16	0.17
30		0.17	0.14	0.26	0.21	0.30	0.16	0.17	0.26	0.26	0.19	0.07	**0.23**	0.25	0.20	0.17	0.18	0.16	0.17	0.17	0.16	0.18
31		0.30	0.24	0.58	0.51	0.50	0.02	0.08	0.33	0.40	0.26	0.31	**0.05**	0.17	0.04	0.05	0.08	0.05	0.08	0.08	0.12	0.16
32		0.63	0.65	0.76	0.74	1.05	0.29	0.32	0.98	0.98	0.67	0.53	**0.82**	0.95	0.88	0.70	0.38	0.69	0.38	0.44	0.46	0.38
33		0.78	0.21	0.24	0.17	0.43	0.30	0.28	0.33	0.42	1.02	0.83	**0.28**	0.42	0.38	0.48	0.48	0.51	0.50	0.47	0.43	0.18
34		0.14	0.49	0.28	0.41	0.49	0.15	0.14	0.42	0.47	0.17	0.20	**0.17**	0.21	0.18	0.38	0.39	0.38	0.39	0.37	0.32	0.34

Index	Molecule	MP2/heavy-aug-cc-pVTZ	B97-D3BJ/def2-TZVP	M05-2X-D3/DZVP	M06-2X-D3/DZVP	M08-HX-D3/DZVP	ωB97M-D3BJ/DZVP	ωB97M-V/DZVP	PW6B95-D3BJ/DZVP	PW6B95-D3/DZVP	MP2/aug-cc-pVTZ	DSD-BLYP-D3BJ/Heavy-aug-cc-pVTZ	**B3LYP-D3BJ/DZVP**	B3LYP-NL/DZVP	B3LYP-D3MBJ/DZVP	B3LYP-D3BJ/def2-TZVP	B3LYP-D3BJ/def2-TZVPD	B3LYP-D3BJ/def2-TZVPP	B3LYP-D3BJ/def2-TZVPPD	B3LYP-D3BJ/QZVP	B3LYP-D3BJ/6-31+G**	B3LYP-D3BJ/6-311+G**
35		0.43	0.81	0.29	0.24	0.48	0.26	0.18	0.54	0.53	0.60	0.34	**0.54**	0.83	0.61	0.84	0.88	0.84	0.88	0.86	1.01	0.78
36		0.28	0.65	0.47	0.61	0.59	0.56	0.55	0.58	0.58	0.42	0.41	**0.78**	0.65	0.78	0.70	0.69	0.70	0.68	0.67	0.74	0.79
37		0.31	0.13	0.16	0.17	0.18	0.05	0.06	0.12	0.11	0.17	0.23	**0.08**	0.09	0.07	0.13	0.13	0.13	0.12	0.12	0.09	0.12
38		0.49	0.69	0.48	0.44	0.46	0.20	0.22	0.24	0.26	0.51	0.31	**0.36**	0.30	0.33	0.40	0.39	0.41	0.40	0.39	0.36	0.38
39		0.49	0.39	1.03	0.83	0.78	0.67	0.74	0.38	0.48	0.40	0.26	**0.18**	0.20	0.18	0.28	0.32	0.33	0.34	0.33	0.12	0.16
40		0.20	0.08	0.11	0.09	0.10	0.11	0.10	0.11	0.12	0.15	0.08	**0.09**	0.07	0.09	0.13	0.12	0.13	0.12	0.12	0.18	0.16
41		0.18	0.22	0.49	0.34	0.25	0.42	0.46	0.15	0.20	0.12	0.08	**0.18**	0.15	0.15	0.21	0.22	0.22	0.23	0.23	0.18	0.18

Index	Molecule	MP2/heavy-aug-cc-pVTZ	B97-D3BJ/def2-TZVP	M05-2X-D3/DZVP	M06-2X-D3/DZVP	M08-HX-D3/DZVP	ωB97M-D3BJ/DZVP	ωB97M-V/DZVP	PW6B95-D3BJ/DZVP	PW6B95-D3/DZVP	MP2/aug-cc-pVTZ	DSD-BLYP-D3BJ/Heavy-aug-cc-pVTZ	**B3LYP-D3BJ/DZVP**	B3LYP-NL/DZVP	B3LYP-D3MBJ/DZVP	B3LYP-D3BJ/def2-TZVP	B3LYP-D3BJ/def2-TZVPD	B3LYP-D3BJ/def2-TZVPP	B3LYP-D3BJ/def2-TZVPPD	B3LYP-D3BJ/QZVP	B3LYP-D3BJ/6-31+G**	B3LYP-D3BJ/6-311+G**
42		0.44	0.59	0.64	0.43	0.79	0.22	0.24	0.69	0.69	0.49	0.69	**0.77**	0.83	0.82	0.75	0.67	0.75	0.69	0.68	0.81	0.65
43		1.03	0.75	0.14	0.23	0.30	0.65	0.53	0.25	0.27	1.16	0.55	**0.28**	0.32	0.32	0.36	0.37	0.40	0.37	0.35	0.35	0.30
44		1.03	1.02	0.16	0.25	0.50	0.55	0.42	0.51	0.46	0.86	1.05	**0.48**	0.60	0.53	1.07	1.05	1.11	1.05	1.05	0.69	0.67
45		0.82	0.56	0.12	0.25	0.35	0.54	0.35	0.29	0.18	0.61	0.83	**0.27**	0.40	0.35	0.72	0.70	0.75	0.70	0.71	0.41	0.37
46		0.36	0.79	0.34	0.32	0.47	0.13	0.21	0.66	0.60	0.32	0.14	**0.75**	0.71	0.78	0.68	0.69	0.68	0.69	0.67	1.06	1.06
47		0.41	0.33	0.22	0.35	0.42	0.32	0.26	0.24	0.25	0.51	0.26	**0.33**	0.23	0.38	0.21	0.21	0.21	0.21	0.20	0.24	0.28
48		0.17	0.38	0.06	0.20	0.28	0.10	0.07	0.32	0.40	0.15	0.06	**0.38**	0.32	0.42	0.31	0.29	0.32	0.29	0.29	0.28	0.31

Index	Molecule	MP2/heavy-aug-cc-pVTZ	B97-D3BJ/def2-TZVP	M05-2X-D3/DZVP	M06-2X-D3/DZVP	M08-HX-D3/DZVP	ωB97M-D3BJ/DZVP	ωB97M-V/DZVP	PW6B95-D3BJ/DZVP	PW6B95-D3/DZVP	MP2/aug-cc-pVTZ	DSD-BLYP-D3BJ/Heavy-aug-cc-pVTZ	**B3LYP-D3BJ/DZVP**	B3LYP-NL/DZVP	B3LYP-D3MBJ/DZVP	B3LYP-D3BJ/def2-TZVP	B3LYP-D3BJ/def2-TZVPD	B3LYP-D3BJ/def2-TZVPP	B3LYP-D3BJ/def2-TZVPPD	B3LYP-D3BJ/QZVP	B3LYP-D3BJ/6-31+G**	B3LYP-D3BJ/6-311+G**
49		0.30	0.10	0.13	0.08	0.13	0.12	0.21	0.12	0.12	0.19	0.06	**0.05**	0.07	0.07	0.10	0.09	0.10	0.09	0.10	0.13	0.07
50		0.17	0.31	0.22	0.20	0.22	0.14	0.33	0.21	0.20	0.20	0.06	**0.06**	0.24	0.05	0.09	0.10	0.07	0.10	0.11	0.11	0.05
51		0.31	0.42	0.24	0.13	0.08	0.18	0.12	0.08	0.10	0.21	0.18	**0.25**	0.10	0.25	0.31	0.33	0.31	0.31	0.30	0.11	0.19
52		0.56	0.39	1.42	1.22	1.21	1.06	1.09	0.97	1.03	0.69	0.41	**0.82**	0.82	0.86	0.17	0.26	0.17	0.26	0.23	0.34	0.59
53		0.45	0.63	1.26	1.19	1.15	0.93	0.98	0.92	0.96	0.61	0.61	**1.11**	1.01	1.19	0.67	0.62	0.68	0.61	0.61	1.21	1.29
54		0.50	1.31	0.59	0.55	0.32	0.62	0.66	0.49	0.41	0.49	0.77	**0.57**	0.59	0.49	1.20	1.43	1.27	1.42	1.34	1.16	0.92
55		0.19	0.30	0.45	0.27	0.31	0.38	0.45	0.15	0.15	0.34	0.11	**0.33**	0.39	0.31	0.43	0.43	0.43	0.42	0.41	0.43	0.52
56		2.33	0.50	0.61	0.51	0.65	0.63	0.62	0.34	0.29	2.48	1.14	**0.61**	0.56	0.64	0.21	0.24	0.22	0.23	0.23	0.60	0.61

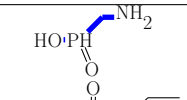
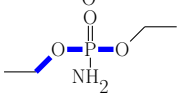
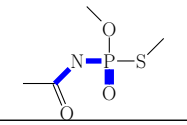
Index	Molecule	MP2/heavy-aug-cc-pVTZ	B97-D3BJ/def2-TZVP	M05-2X-D3/DZVP	M06-2X-D3/DZVP	M08-HX-D3/DZVP	ωB97M-D3BJ/DZVP	ωB97M-V/DZVP	PW6B95-D3BJ/DZVP	PW6B95-D3/DZVP	MP2/aug-cc-pVTZ	DSD-BLYP-D3BJ/Heavy-aug-cc-pVTZ	**B3LYP-D3BJ/DZVP**	B3LYP-NL/DZVP	B3LYP-D3MBJ/DZVP	B3LYP-D3BJ/def2-TZVP	B3LYP-D3BJ/def2-TZVPD	B3LYP-D3BJ/def2-TZVPP	B3LYP-D3BJ/def2-TZVPPD	B3LYP-D3BJ/QZVP	B3LYP-D3BJ/6-31+G**	B3LYP-D3BJ/6-311+G**
57		0.06	0.12	0.29	0.32	0.33	0.24	0.25	0.34	0.38	0.06	0.03	**0.29**	0.30	0.27	0.06	0.15	0.09	0.15	0.15	0.34	0.35
58		0.13	0.22	0.20	0.19	0.25	0.24	0.18	0.29	0.26	0.14	0.12	**0.27**	0.23	0.25	0.20	0.20	0.20	0.20	0.21	0.27	0.35
59		0.38	0.52	1.00	0.84	1.16	0.55	0.65	0.61	0.66	0.44	0.32	**0.42**	0.50	0.47	0.25	0.21	0.24	0.21	0.20	0.25	0.18

Table S1: RMSE in torsion profile energies (kcal/mol) with different functional/basis set combinations for all the molecules in this benchmark set with respect to the reference level CCSD(T)/CBS//MP2/heavy-aug-cc-pVTZ. The dihedral atoms involved in the torsion are highlighted in blue (in some cases where a fourth atom is not tagged it is the hydrogen not depicted in these diagrams). OpenFF's current default level of theory being used in training force fields, B3LYP-D3BJ/DZVP, is boldfaced and the columns to the right of it have the same functional but different basis set and dispersion corrections. The columns to the left of it have different functionals, mostly within the DZVP basis set unless specified otherwise.

SI 1.2 Compute Infrastructure

The Molecular Sciences Software Institute's (MolSSI) Quantum Chemistry Archive (QCA)[1] is a key part of OpenFF's science infrastructure and we have built an automated dataset submission and validation workflow using the `openff-qcsubmit`[2] package. QC Archive is a dynamic database of quantum chemistry datasets for various molecules and properties of interest. It encompasses diverse subsets of calculations, including optimized geometries, multi-dimensional torsional scans, vibrational frequency calculations, single-point energy calculations, and reaction-based calculations. These calculations span different levels of theory and serve various scientific pursuits. They contribute to endeavors such as constructing machine

learning models, training and benchmarking physics-based potentials for molecular simulations, and much more. The GitHub repository from OpenFF, `qca-dataset-submission`, serves as a node for interacting with the QCA server. Figure S1 depicts a workflow for OpenFF dataset generation, which starts with a set of molecule representations either in 2D or 3D, and calculation inputs were built in the format required by the backend packages `QCFractal+QCEngine`.[3] Each entry in the calculation input being sent to the server for computation corresponds to a single molecule and an associated QM calculation schema.

A QM calculation schema in each dataset entry specifies the program to use, the type of calculation to perform such as torsion scan or single point energy calculation, and QM calculation details such as functional, basis set, convergence criteria, maximum number of steps, SCF properties to derive from the final wave function, etc., and each of those inputs were validated to have the correct input syntax that can be parsed by the program. This input validation ensures we can skip any calculations which would have failed due to invalid input syntax before even submitting to the compute cluster.

The `QCFractal` package is utilized to instantiate a compute manager and compute workers on a cluster, enabling the execution of a substantial number of independent calculations or jobs in parallel. A compute manager monitors the job completion status and compute workers grab any incomplete jobs from the compute manager and send back completed jobs. The compute managers interface with the QCA server to update the QCA database with finished calculations. Automated error-cycling has been set up using `GitHub actions`, a backend tool provided by GitHub for workflow management, allowing resubmission of failed jobs and also generating high level summaries on the progress of the calculations. These automated error reports help in diagnosing whether the failures were due to early preemption of compute workers, or due to lack of resources, such as low memory for calculations with larger molecules, or redundant failures due to a systematic issue with the compute environment on the worker compute node, and other failure modes that may occur during computation.

Several software packages were used in combination to perform the calculations and generate quantum chemistry data. We used Psi4[4–19] for electronic structure calculations. Psi4 is a versatile, open source, quantum chemistry package. Psi4 is highly parallel, and has the most efficient implementation of density-fitting method. Psi4 has fast analytic gradients for many high level methods such as Møller–Plesset perturbation theory (MP2) as well as for most functional and dispersion corrections. For geometry optimizations an external optimizer, GeomeTRIC,[20] was used. GeomeTRIC uses a translation-rotation-internal coordinate (TRIC) system which significantly speeds up geometry optimizations. And for torsion scans, the TorsionDrive[21] program, which uses a wavefront propagation algorithm and multiple initial guesses in parallel, independent of the search direction, to avoid hysteresis in the atomic configurations involved in driving the torsion. The completed calculations can be conveniently accessed from the QCA server at any time, and from anywhere in the world. Accessing the dataset only requires providing the dataset type, name, and the compute specification.

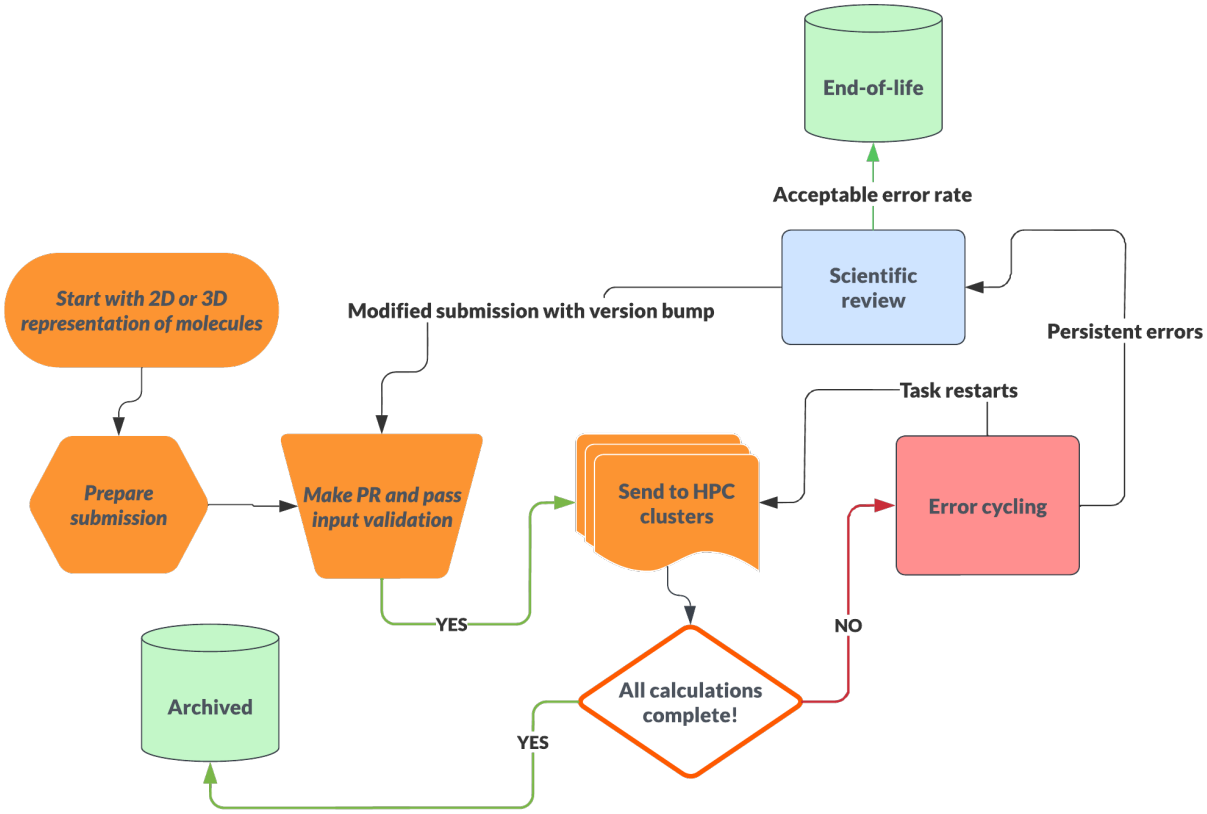


Figure S1: OpenFF dataset generation workflow, qca-dataset-submission. Starting with a set of 2D or 3D representations of molecules and using openff-qcsubmit [2] a dataset submission can be prepared with the required level of theory and other hyperparameters that are necessary for a calculation. After the inputs are ready one can make a pull request on this repository, and a reviewer will then process the submission and after approval the calculations are submitted to QCArchive server, a public repository of data hosted by MolSSI. QCFractal is the backend server software that manages the calculations to be done and sends them to compute managers on HPC clusters hosted at different compute centers. Finished calculations are sent back to QCArchive.

SI 1.3 Dipole moment comparison

Dipole moment, an electrostatic property, is an overall measure of geometry and electronic structure within a molecule and getting it right is crucial for intermolecular interactions. For all the methods tested, the deviations in the magnitude and the direction of dipole moment vectors with respect to the reference level of theory, CCSD(T)/haDZ//MP2/haTZ, are tabulated in Table S2. All the methods tested give overall comparable performance statistics relative to one another, with an RMSE in dipole magnitudes of around 0.6 Debye, and deviations in dipole moment vectors of around 0.1-0.2 degrees. Dipole moment is a key metric for modeling the quality of electrostatic interactions, and we refer the reader to more

detailed studies that benchmarked electronic structure methods for electrostatic properties, such as polarizabilities, hyperpolarizabilities, and magnetizabilities.[22–25]

Table S2: The deviations in dipole moment magnitudes (in Debye) with respect to the reference CCSD(T)/heavy-aug-cc-pVDZ//MP2/heavy-aug-cc-pVTZ level of theory, for the whole benchmark set and the neutral and charged subsets. The 95% confidence intervals, calculated with cinnabar, were presented on the side. The deviations are minimal for all the theory levels benchmarked and the values are comparable to each other. Large deviations were seen on the charged subset of molecules. Origin (0, 0, 0) is chosen as the reference point for the dipoles.

	RMSE in Dipole moments (in Debye)		
Specification	Whole set	Neutral subset	Charged subset
B97-D3(BJ)/def2-TZVP	$0.6754^{0.7645}_{0.5899}$	$0.1572^{0.1654}_{0.1491}$	$0.9499^{1.0737}_{0.8267}$
M05-2X-D3/DZVP	$0.6497^{0.7375}_{0.5666}$	$0.3765^{0.3947}_{0.3592}$	$0.8439^{0.9702}_{0.7166}$
M06-2X-D3/DZVP	$0.6172^{0.7049}_{0.5270}$	$0.3071^{0.3231}_{0.2910}$	$0.8231^{0.9644}_{0.6916}$
M08-HX-D3/DZVP	$0.6245^{0.7079}_{0.5421}$	$0.3163^{0.3318}_{0.3001}$	$0.8307^{0.9570}_{0.7022}$
ωB97M-D3(BJ)/DZVP	$0.5877^{0.6777}_{0.4977}$	$0.2156^{0.2283}_{0.2035}$	$0.8091^{0.9500}_{0.6721}$
ωB97M-V/DZVP	$0.5877^{0.6824}_{0.4939}$	$0.2166^{0.2287}_{0.2042}$	$0.8089^{0.9533}_{0.6782}$
PW6B95-D3(BJ)/DZVP	$0.5998^{0.6865}_{0.4959}$	$0.2162^{0.2292}_{0.2025}$	$0.8267^{0.9660}_{0.6879}$
PW6B95-D3/DZVP	$0.5998^{0.6936}_{0.5041}$	$0.2162^{0.2302}_{0.2030}$	$0.8267^{0.9698}_{0.6910}$
MP2/aug-cc-pVTZ	$0.6692^{0.7501}_{0.5908}$	$0.3774^{0.3905}_{0.3636}$	$0.8740^{1.0067}_{0.7593}$
MP2/heavy-aug-cc-pVTZ	$0.6695^{0.7606}_{0.5906}$	$0.3782^{0.3917}_{0.3650}$	$0.8741^{1.0039}_{0.7516}$
DSD-BLYP-D3(BJ)/heavy-aug-cc-pVTZ	$0.6368^{0.7238}_{0.5470}$	$0.3026^{0.3142}_{0.2921}$	$0.8546^{0.9869}_{0.7301}$
B3LYP-D3(BJ)/DZVP	$0.6079^{0.6997}_{0.5145}$	$0.2118^{0.2263}_{0.1983}$	$0.8399^{0.9723}_{0.6993}$
B3LYP-D3MBJ/DZVP	$0.6079^{0.7053}_{0.5200}$	$0.2118^{0.2259}_{0.1971}$	$0.8399^{0.9821}_{0.7114}$
B3LYP-NL/DZVP	$0.6076^{0.6981}_{0.5137}$	$0.2120^{0.2259}_{0.1999}$	$0.8394^{0.9708}_{0.7100}$
B3LYP-D3(BJ)/def2-SV(P)	$0.6400^{0.7308}_{0.5594}$	$0.1511^{0.1579}_{0.1442}$	$0.8999^{1.0268}_{0.7752}$
B3LYP-D3(BJ)/def2-SVP	$0.6483^{0.7395}_{0.5560}$	$0.1481^{0.1546}_{0.1419}$	$0.9124^{1.0480}_{0.7766}$
B3LYP-D3(BJ)/def2-TZVP	$0.6152^{0.7046}_{0.5238}$	$0.0942^{0.1004}_{0.0885}$	$0.8723^{0.9977}_{0.7490}$
B3LYP-D3(BJ)/def2-TZVPD	$0.6148^{0.7087}_{0.5214}$	$0.1058^{0.1147}_{0.0966}$	$0.8703^{1.0025}_{0.7430}$
B3LYP-D3(BJ)/def2-TZVPP	$0.6178^{0.7160}_{0.5274}$	$0.0929^{0.0984}_{0.0872}$	$0.8761^{0.9998}_{0.7432}$
B3LYP-D3(BJ)/def2-TZVPPD	$0.6151^{0.7069}_{0.5259}$	$0.1057^{0.1142}_{0.0975}$	$0.8708^{1.0022}_{0.7373}$
B3LYP-D3(BJ)/def2-QZVP	$0.6148^{0.7113}_{0.5193}$	$0.1001^{0.1076}_{0.0925}$	$0.8710^{0.9954}_{0.7406}$
B3LYP-D3(BJ)/6-31G*	$0.6136^{0.7119}_{0.5247}$	$0.1194^{0.1250}_{0.1132}$	$0.8667^{0.9950}_{0.7414}$
B3LYP-D3(BJ)/6-31+G**	$0.6118^{0.7123}_{0.5219}$	$0.2034^{0.2189}_{0.1895}$	$0.8477^{0.9827}_{0.7150}$
B3LYP-D3(BJ)/6-311+G**	$0.6144^{0.7003}_{0.5181}$	$0.2092^{0.2251}_{0.1940}$	$0.8501^{0.9791}_{0.7107}$

SI 1.4 Energy from a method versus the energy difference wrt reference

We plotted the energy from a method versus the energy difference with respect to the reference theory level and we see a poor correlation between increase in error and the rise in energy.

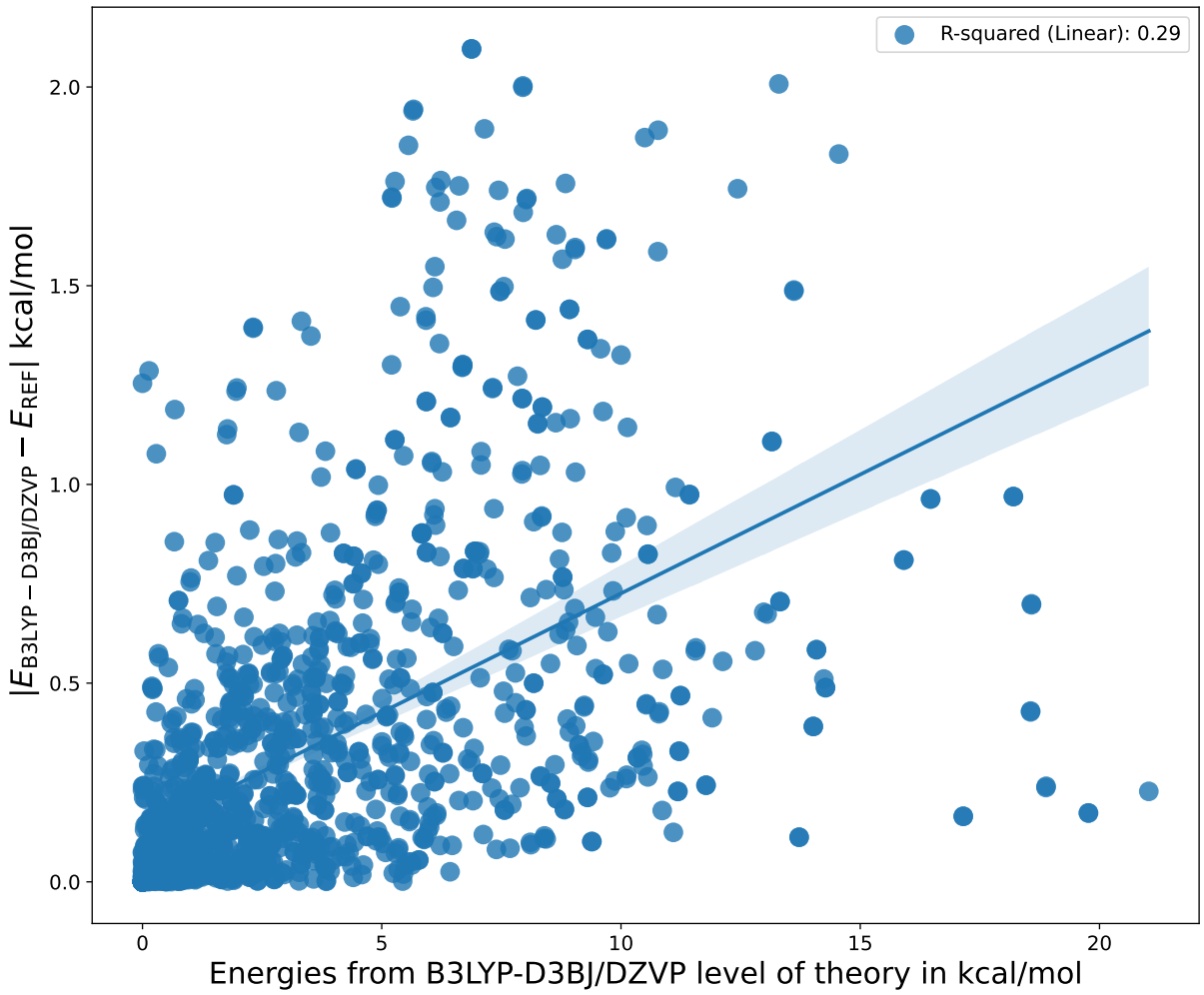


Figure S2: Energy from B3LYP-D3BJ/DZVP method versus the energy difference wrt reference theory level.

SI 1.5 Comparison of torsional barriers

Table S3: Comparison of RMSEs in torsion barriers on the whole set and in the neutral and charged molecules subsets with respect to the reference level CCSD(T)/[haTQZ; δ:haDZ]//MP2/heavy-aug-cc-pVTZ level of theory. The 95% confidence intervals, calculated with cinnabar. The RMSE in barriers is slightly higher than the RMSE in energies for each functional as seen in Table 2. This can be attributed to the coarser grid and estimation of barriers from single point energy calculations.

	RMSE (in kcal/mol)		
Specification	Whole set	Neutral subset	Charged subset
B97-D3(BJ)/def2-TZVP	$0.7632^{0.9065}_{0.6138}$	$0.7502^{0.9668}_{0.5510}$	$0.7764^{0.9769}_{0.5522}$
M05-2X-D3/DZVP	$0.8614^{1.0498}_{0.6603}$	$0.9022^{1.1793}_{0.6361}$	$0.8170^{1.0587}_{0.5449}$
M06-2X-D3/DZVP	$0.6823^{0.8572}_{0.5214}$	$0.7280^{0.9849}_{0.4814}$	$0.6316^{0.8340}_{0.4242}$
M08-HX-D3/DZVP	$0.9174^{1.1190}_{0.6985}$	$0.8721^{1.1457}_{0.6148}$	$0.9619^{1.2855}_{0.6141}$
ωB97X-D3(BJ)/DZVP	$0.6614^{0.8286}_{0.4902}$	$0.6640^{0.8993}_{0.4262}$	$0.6587^{0.9008}_{0.3955}$
ωB97M-D3(BJ)/DZVP	$\mathbf{0.6207^{0.7540}_{0.4738}}$	$\mathbf{0.6235^{0.7986}_{0.4400}}$	$\mathbf{0.6178^{0.8356}_{0.3658}}$
ωB97M-V/DZVP	$0.6462^{0.7973}_{0.4784}$	$0.6239^{0.7943}_{0.4242}$	$0.6686^{0.9159}_{0.4164}$
PW6B95-D3(BJ)/DZVP	$0.7630^{0.9501}_{0.5892}$	$0.6822^{0.8950}_{0.4885}$	$0.8384^{1.1219}_{0.5329}$
PW6B95-D3/DZVP	$0.7809^{0.9783}_{0.5761}$	$0.7118^{0.9450}_{0.5079}$	$0.8465^{1.1436}_{0.5455}$
MP2/aug-cc-pVTZ	$0.7347^{0.8727}_{0.5869}$	$0.5788^{0.7124}_{0.4367}$	$0.8669^{1.0746}_{0.6354}$
MP2/heavy-aug-cc-pVTZ	$0.6904^{0.8141}_{0.5711}$	$0.5553^{0.6643}_{0.4353}$	$0.8066^{0.9910}_{0.6298}$
DSD-BLYP-D3(BJ)/heavy-aug-cc-pVTZ	$0.6909^{0.8102}_{0.5679}$	$0.6399^{0.8217}_{0.4663}$	$0.7399^{0.9144}_{0.5591}$
B3LYP-D3(BJ)/DZVP	$0.7876^{0.9388}_{0.6289}$	$0.6909^{0.8504}_{0.5212}$	$0.8765^{1.1238}_{0.6105}$
B3LYP-D3MBJ/DZVP	$0.8283^{0.9816}_{0.6782}$	$0.7491^{0.9131}_{0.5598}$	$0.9030^{1.1354}_{0.6502}$
B3LYP-NL/DZVP	$0.7932^{0.9353}_{0.6416}$	$0.6860^{0.8281}_{0.5293}$	$0.8906^{1.1126}_{0.6282}$
B3LYP-D3(BJ)/def2-SV(P)	$1.5712^{1.9310}_{1.2255}$	$1.4162^{1.7318}_{1.0583}$	$1.7169^{2.2764}_{1.1068}$
B3LYP-D3(BJ)/def2-SVP	$1.6169^{1.9743}_{1.2637}$	$1.4040^{1.7555}_{1.0368}$	$1.8110^{2.3851}_{1.1604}$
B3LYP-D3(BJ)/def2-TZVP	$0.7121^{0.8722}_{0.5586}$	$0.6882^{0.9033}_{0.4887}$	$0.7361^{0.9519}_{0.5097}$
B3LYP-D3(BJ)/def2-TZVPD	$0.6584^{0.7933}_{0.5089}$	$0.6674^{0.8621}_{0.4633}$	$0.6490^{0.8370}_{0.4395}$
B3LYP-D3(BJ)/def2-TZVPP	$0.7238^{0.8786}_{0.5592}$	$0.7004^{0.9149}_{0.4858}$	$0.7473^{0.9885}_{0.4986}$
B3LYP-D3(BJ)/def2-TZVPPD	$0.6628^{0.7988}_{0.5171}$	$0.6715^{0.8584}_{0.4665}$	$0.6537^{0.8436}_{0.4394}$
B3LYP-D3(BJ)/def2-QZVP	$0.6658^{0.8098}_{0.5141}$	$0.6669^{0.8684}_{0.4694}$	$0.6646^{0.8810}_{0.4508}$
B3LYP-D3(BJ)/6-31G*	$1.1595^{1.3783}_{0.9457}$	$1.0747^{1.2925}_{0.8514}$	$1.2411^{1.5402}_{0.8144}$
B3LYP-D3(BJ)/6-31+G**	$0.7553^{0.8911}_{0.6184}$	$0.8037^{1.0029}_{0.6009}$	$0.7016^{0.8599}_{0.5182}$
B3LYP-D3(BJ)/6-311+G**	$0.7092^{0.8485}_{0.5659}$	$0.7911^{1.0082}_{0.5666}$	$0.6130^{0.7625}_{0.4582}$

References

(1) Smith, D. G. A.; Altarawy, D.; Burns, L. A.; Welborn, M.; Naden, L. N.; Ward, L.; Ellis, S.; Pritchard, B. P.; Crawford, T. D. The MolSSI QCArchive project: An open-source platform to compute, organize, and share quantum chemistry data. *WIREs Computational Molecular Science* **2021**, *11*, e1491, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/wcms.1491.

(2) Horton, J. T.; Boothroyd, S.; Wagner, J.; Mitchell, J. A.; Gokey, T.; Dotson, D. L.; Behara, P. K.; Ramaswamy, V. K.; Mackey, M.; Chodera, J. D.; Anwar, J.; Mobley, D. L.; Cole, D. J. Open Force Field BespokeFit: Automating Bespoke Torsion Parametrization at Scale. *Journal of Chemical Information and Modeling* **2022**, *62*, 5622–5633, Publisher: American Chemical Society.

(3) Smith, D. G. A. et al. Quantum Chemistry Common Driver and Databases (QCDB) and Quantum Chemistry Engine (QCEngine): Automation and interoperability among computational chemistry programs. *The Journal of Chemical Physics* **2021**, *155*, 204801, Publisher: American Institute of Physics.

(4) Smith, D. G. A. et al. PSI4 1.4: Open-source software for high-throughput quantum chemistry. *The Journal of Chemical Physics* **2020**, *152*, 184108.

(5) Turney, J. M. et al. Psi4: an open-source ab initio electronic structure program. *WIREs Computational Molecular Science* **2012**, *2*, 556–565, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/wcms.93.

(6) Parrish, R. M. et al. Psi4 1.1: An Open-Source Electronic Structure Program Emphasizing Automation, Advanced Libraries, and Interoperability. *Journal of Chemical Theory and Computation* **2017**, *13*, 3185–3197, Publisher: American Chemical Society.

(7) Crawford, T. D.; Schaefer III, H. F. *Reviews in Computational*

Chemistry; John Wiley & Sons, Ltd, 2000; pp 33–136, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/9780470125915.ch2.

(8) DePrince, A. E.; Sherrill, C. D. Accuracy and Efficiency of Coupled-Cluster Theory Using Density Fitting/Cholesky Decomposition, Frozen Natural Orbitals, and a t1-Transformed Hamiltonian. *Journal of Chemical Theory and Computation* **2013**, *9*, 2687–2696, Publisher: American Chemical Society.

(9) DePrince, A. E. I.; Sherrill, C. D. Accurate Noncovalent Interaction Energies Using Truncated Basis Sets Based on Frozen Natural Orbitals. *Journal of Chemical Theory and Computation* **2013**, *9*, 293–299, Publisher: American Chemical Society.

(10) Bozkaya, U.; Sherrill, C. D. Analytic energy gradients for the orbital-optimized second-order Møller–Plesset perturbation theory. *The Journal of Chemical Physics* **2013**, *138*, 184103.

(11) Bozkaya, U. Analytic Energy Gradients and Spin Multiplicities for Orbital-Optimized Second-Order Perturbation Theory with Density-Fitting Approximation: An Efficient Implementation. *Journal of Chemical Theory and Computation* **2014**, *10*, 4389–4399, Publisher: American Chemical Society.

(12) Bozkaya, U.; Sherrill, C. D. Analytic energy gradients for the coupled-cluster singles and doubles method with the density-fitting approximation. *The Journal of Chemical Physics* **2016**, *144*, 174103.

(13) Bozkaya, U. A noniterative asymmetric triple excitation correction for the density-fitted coupled-cluster singles and doubles method: Preliminary applications. *The Journal of Chemical Physics* **2016**, *144*, 144108.

(14) Bozkaya, U. Orbital-Optimized Second-Order Perturbation Theory with Density-Fitting and Cholesky Decomposition Approximations: An Efficient Implementation.

Journal of Chemical Theory and Computation **2014**, *10*, 2371–2378, Publisher: American Chemical Society.

(15) Bozkaya, U.; Turney, J. M.; Yamaguchi, Y.; Schaefer, H. F., III; Sherrill, C. D. Quadratically convergent algorithm for orbital optimization in the orbital-optimized coupled-cluster doubles method and in orbital-optimized second-order Møller-Plesset perturbation theory. *The Journal of Chemical Physics* **2011**, *135*, 104103.

(16) Lehtola, S. Assessment of Initial Guesses for Self-Consistent Field Calculations. Superposition of Atomic Potentials: Simple yet Efficient. *Journal of Chemical Theory and Computation* **2019**, *15*, 1593–1604, Publisher: American Chemical Society.

(17) Van Lenthe, J. H.; Zwaans, R.; Van Dam, H. J. J.; Guest, M. F. Starting SCF calculations by superposition of atomic densities. *Journal of Computational Chemistry* **2006**, *27*, 926–932, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/jcc.20393.

(18) Almlöf, J.; Faegri Jr., K.; Korsell, K. Principles for a direct SCF approach to LICAO–MOab-initio calculations. *Journal of Computational Chemistry* **1982**, *3*, 385–399, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/jcc.540030314.

(19) Lehtola, S.; Steigemann, C.; Oliveira, M. J. T.; Marques, M. A. L. Recent developments in libxc — A comprehensive library of functionals for density functional theory. *SoftwareX* **2018**, *7*, 1–5.

(20) Wang, L.-P.; Song, C. Geometry optimization made simple with translation and rotation coordinates. *The Journal of Chemical Physics* **2016**, *144*, 214108.

(21) Qiu, Y.; Smith, D. G. A.; Stern, C. D.; Feng, M.; Jang, H.; Wang, L.-P. Driving torsion scans with wavefront propagation. *The Journal of Chemical Physics* **2020**, *152*, 244116.

(22) Hait, D.; Head-Gordon, M. How Accurate Is Density Functional Theory at Predicting Dipole Moments? An Assessment Using a New Database of 200 Benchmark Values.

Journal of Chemical Theory and Computation **2018**, *14*, 1969–1981, Publisher: American Chemical Society.

(23) Lehtola, S.; Dimitrova, M.; Fliegl, H.; Sundholm, D. Benchmarking Magnetizabilities with Recent Density Functionals. *Journal of Chemical Theory and Computation* **2021**, *17*, 1457–1468, Publisher: American Chemical Society.

(24) Zhou, A.; Schauperl, M.; Nerenberg, P. S. Benchmarking Electronic Structure Methods for Accurate Fixed-Charge Electrostatic Models. *Journal of Chemical Information and Modeling* **2020**, *60*, 249–258, Publisher: American Chemical Society.

(25) Hickey, A. L.; Rowley, C. N. Benchmarking Quantum Chemical Methods for the Calculation of Molecular Dipole Moments and Polarizabilities. *The Journal of Physical Chemistry A* **2014**, *118*, 3678–3687, Publisher: American Chemical Society.