Supplementary Information for

MATCH: An Atom- Typing Toolset for Molecular Mechanics Force Fields

Joseph D. Yesselman[1], Daniel J. Price[2], Jennifer L. Knight[1], and Charles L. Brooks III[1,3,*]

[1]Department of Chemistry and Biophysics Program, University of Michigan, Ann Arbor, MI 48109, USA

[2]GlaxoSmithKline, MDR/CSC, 3.3134B, 5 Moore Drive, Research Triangle Park, NC 27709, USA

[3]Center for Theoretical Biological Physics, University of California San Diego, San Diego, CA 92037, USA

List of the molecules excluded from the CGENFF force field:

NAD;ATP;NADP;PY01;PY02;4HQI;HEX3;13DB;DMB1;DMP1;DMP2;CROT;MECH;FRET;RTOL;RTAL;RTAC;MIND;PTID;PYZN;FLRN;PUR7;PUR9;ISOI;ZTHP;ZIMI;ZFUR;NIC;INDO;INDA;3PRP;ZTHZ;INDI;INDE;3PRL;3HIN;2PRP;ISOX;CA;DCA;UDCA;CDCA;INDZ;2HPR;2IMP;2HPP;AMDN;THNP;HDZ2;INDZ;CSO;THNI

Figure S1: Summary of data for atomic charges in the cross validation studies when considering all atoms that are fully charged regardless of whether the entire molecule is able to be typed and charged

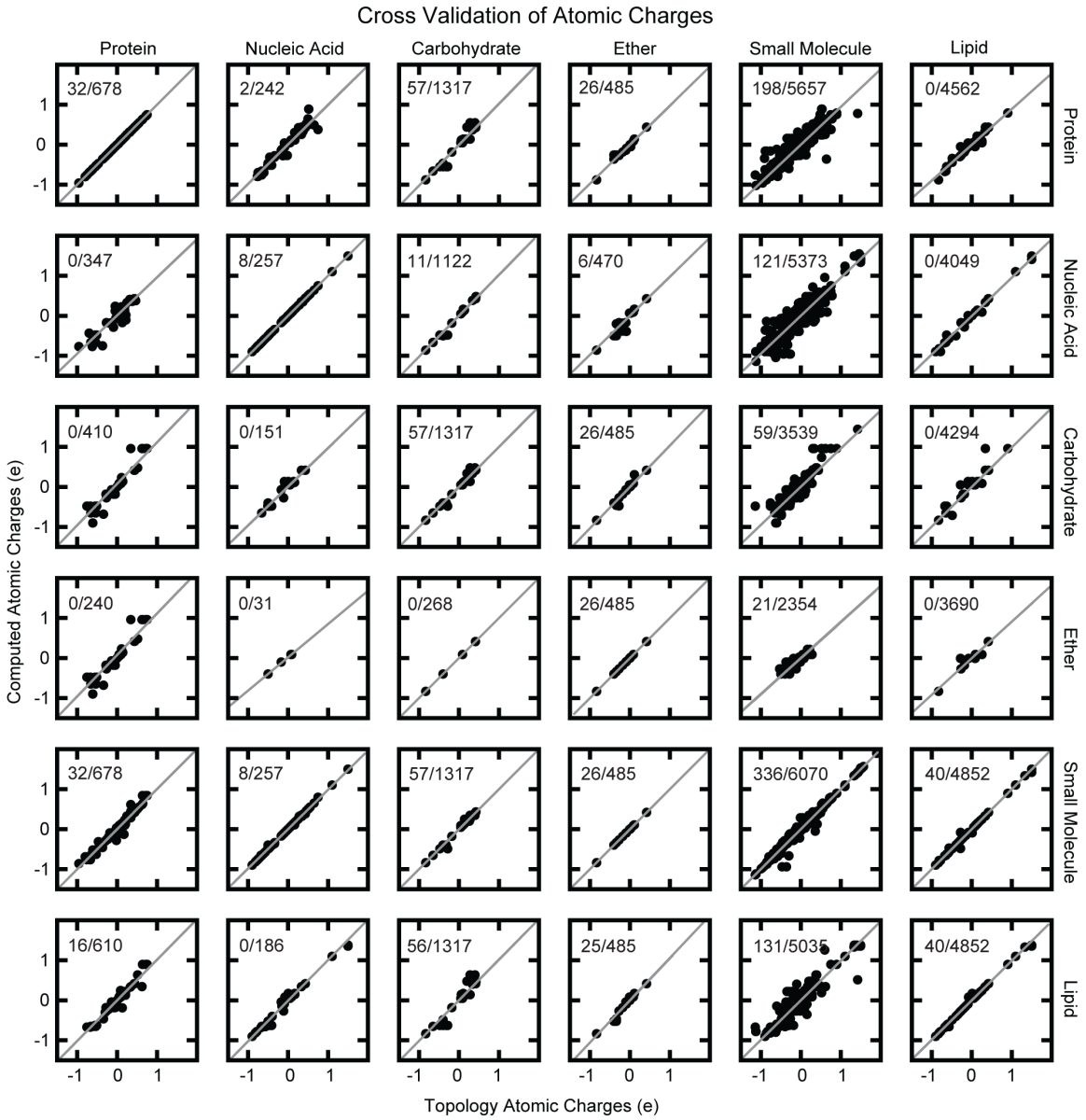

Table S1: Summary of data for charge cross validation studies comparing using atoms from only completely charged molecules and all atoms regardless of complete typing and charging of the molecules that contained them.

		Complete Molecules Only				All Completed Atoms			
MATCH Library	**Force Field**	**#atoms**	**Average Error**	**Average % Error**	**R^2**	**#atoms**	**Average Error**	**Average% Error**	**R^2**
prot	prot	678	0	0.041	1	678	0	0.041	1
prot	na	62	0.035	18.31	0.97	242	0.037	19.664	0.961
prot	carb	1317	0.008	2.581	0.998	1317	0.008	2.581	0.998
prot	ethers	476	0.005	14.213	0.985	485	0.005	13.95	0.985
prot	cgenff	3544	0.024	15.848	0.941	5657	0.03	20.76	0.925
prot	lipid	3	0.031	5.516	1	4562	0.014	6.679	0.97
na	prot	0	NA	NA	NA	347	0.031	17.753	0.917
na	na	257	0.001	0.182	1	257	0.001	0.182	1
na	carb	201	0.014	5.024	0.996	1122	0.008	2.035	0.999
na	ethers	73	0.006	11.71	0.991	470	0.017	13.509	0.907
na	cgenff	2304	0.04	27.578	0.934	5373	0.042	28.609	0.911
na	lipid	3	0.017	3.118	1	4049	0.015	4.986	0.972
carb	prot	0	N/A	N/A	N/A	410	0.087	22.105	0.8
carb	na	0	N/A	N/A	N/A	151	0.018	17.09	0.987
carb	carb	1317	0.001	0.308	0.999	1317	0.001	0.308	0.999
carb	ethers	476	0.004	15.382	0.982	485	0.004	15.096	0.982
carb	cgenff	882	0.019	15.231	0.916	3539	0.023	16.267	0.912
carb	lipid	3	0	0	1	4294	0.018	8.266	0.934
ethers	prot	0	N/A	N/A	N/A	240	0.002	2.161	0.997
ethers	na	0	N/A	N/A	N/A	31	0.03	9.255	0.999
ethers	carb	3	0	0	1	268	0	0	1
ethers	ethers	453	0	0	1	480	0	0	1
ethers	cgenff	302	0	0.522	0.999	2354	0.009	5.426	0.933
ethers	lipid	3	0	0	1	3690	0.007	3.027	0.942
cgenff	prot	678	0.007	6.001	0.992	678	0.007	6.001	0.992
cgenff	na	257	0.009	3.174	0.997	257	0.009	3.174	0.997
cgenff	carb	1317	0.001	0.409	1	1317	0.001	0.409	1
cgenff	ethers	476	0	0.005	1	486	0	0.004	1
cgenff	cgenff	6070	0.003	1.711	0.997	6070	0.003	1.711	0.997
cgenff	lipid	4852	0	0.473	1	4852	0	0.473	1
lipid	prot	302	0.013	6.199	0.985	601	0.011	5.564	0.987

lipid	na	0	N/A	N/A	N/A	186	0.023	17.094	0.984
lipid	carb	1269	0.007	2.988	0.994	1317	0.007	3.006	0.994
lipid	ethers	459	0.008	26.97	0.968	485	0.007	25.524	0.969
lipid	cgenff	2143	0.027	27.49	0.964	5035	0.029	24.421	0.946
lipid	lipid	4852	0.005	3.841	0.992	4852	0.005	3.841	0.992

Table S2: A comparison bond and angle parameter recovery between using atoms from completely parameterized molecules and from considering all atoms that could be parameterized regardless of complete typing/charging/parameterizing of the molecule containing the atoms.

		Complete Molecules Only				**All Completed Atoms**			
Parameter Type	**Parameter**	**R^2**	**Average Error**	**Average % Error**	**count**	**R^2**	**Average Error**	**Average % Error**	**count**
Non-bond	ε_{min}	0.788	0.010	21.250	3623	0.772	0.011	22.680	5611
Non-bond	R_{min}	0.959	0.048	3.240	3623	0.932	0.062	4.803	5611
Bond	B_0	0.977	0.010	0.728	1381	0.904	0.021	1.459	2428
Bond	K_B	0.869	17.511	5.427	1381	0.719	26.656	8.504	2428
Angle	θ_0	0.568	2.160	1.879	2322	0.403	2.695	2.307	4171
Angle	K_θ	0.410	11.616	20.953	2322	0.373	14.138	24.024	4171