

Paired t-test (% NOE fulfilled )

ensemble GaMD	1	1	0	0	0	1	0.021	1
ensemble cMD	0.015	1	0	0.005	0.011	1	1	0.13
RDKit best	0	0	1	1	1	0	0.001	0
RDKit LICUV 10	0	0	0.14	1	1	0.001	0.001	0
RDKit NAMFIS 10	0	0	0.38	0.6	1	0	0	0
RDKit low_energy 10	0.027	0.98	0	0	0	1	1	0.14
RDKit random 10	0	0.13	0	0	0	0.016	1	0.001
RDKit random 1	0.16	0.001	0	0	0	0.001	0	1

Wilcoxon signed rank test  
(% NOE fulfilled )

1	1	0	0	0	1	0.046	1
0.022	1	0.004	0.024	0.021	1	1	0.2
0	0	1	1	1	0.001	0.001	0
0	0	0.61	1	1	0.003	0.002	0
0	0	0.43	0.42	1	0.002	0.004	0
0.031	0.88	0	0	0	1	1	0.11
0	0.24	0	0	0	0.03	1	0.001
0.13	0.002	0	0	0	0.001	0	1

Paired t-test (RMSE)

ensemble GaMD	1	1	0	0	0.004	1	0.18	0.001
ensemble cMD	0.9	1	0	0	0.019	1	1	0.008
RDKit best	0	0	1	1	1	0.042	0.008	0
RDKit LICUV 10	0	0	0.56	1	1	0.04	0.002	0
RDKit NAMFIS 10	0	0	0.91	0.82	1	0.41	0.81	0
RDKit low_energy 10	0.036	0.09	0	0	0.005	1	1	0
RDKit random 10	0.002	0.037	0	0	0.012	0.18	1	0
RDKit random 1	0	0	0	0	0	0	0	1

Wilcoxon signed rank test (RMSE )

1	1	0	0	0	1	0.23	0
0.84	1	0	0	0	1	1	0.003
0	0	1	1	1	0.012	0.03	0
0	0	0.95	1	0.44	0.001	0	0
0	0	0.23	0.006	1	0.007	0.005	0
0.038	0.15	0	0	0	1	1	0
0.003	0.044	0	0	0	0.38	1	0
0	0	0	0	0	0	0	1