# How Molecular Size Impacts RMSD Applications in Molecular Dynamics Simulations

Karen Sargsyan[1*], Cédric Grauffel[1], and Carmay Lim[1,2,*]

[1]*Institute of Biomedical Sciences, Academia Sinica, Taipei 115, Taiwan*
[2]*Department of Chemistry, National Tsinghua University, Hsinchu 300, Taiwan*

**Supplementary Table S1**. Number of aa residues and PDB/CATH codes of proteins in the AMBER10 dataset

| # of aa residues | PDB code | CATH code(s0 |
|---|---|---|
| 41 | 1hy9 | 4.10.40.30 |
| 42 | 1esk | 4.10.60.10 |
| 45 | 1pdc | 2.10.10.10 |
| 46 | 1atx | 2.20.20.10 |
| 49 | 1ify | 1.10.8.10 |
| 50 | 1tfi | 2.20.25.10 |
| 55 | 2vgh | 2.10.160.10 |
| 59 | 1mkn | 2.20.60.10 |
| 60 | 1a7i | 2.10.110.10 |
| 65 | 2alc | 4.10.240.10 |
| 68 | 1n4y | 4.10.70.10 |
| 69 | 1psf | 2.30.30.50 |
| 71 | 1bbi | 2.10.69.10 |
| 72 | 1jw2 | 1.20.1280.40 |
| 75 | 1o6w | 2.20.70.10 |
| 77 | 1eik | 3.90.940.20 |
| 93 | 1fbr | 2.10.70.10 |
| 95 | 1lpl | 2.30.30.190 |
| 97 | 2hgf | 3.50.4.10 |
| 101 | 3vub | 2.30.30.110 |
| 104 | 1skz | 2.10.22.10 |
| 104 | 1bmp | 2.10.90.10 |
| 107 | 1bet | 2.10.90.10 |
| 108 | 1jpc | 2.90.10.10 |
| 110 | 1bfe | 2.30.42.10 |
| 110 | 1aqe | 3.90.10.10 |
| 113 | 1b11 | 2.40.70.10 |
| 119 | 1bak | 2.30.29.30 |
| 127 | 1rfs | 2.102.10.10 |
| 128 | 1eif | 2.30.30.30, 2.40.50.140 |
| 136 | 1lis | 1.20.150.10 |
| 140 | 1jw3 | 3.55.10.10 |
| 140 | 1b1b | 1.10.10.10 |
| 141 | 1ckv | 3.90.56.10 |
| 144 | 1qqh | 2.170.16.10 |
| 144 | 1j7g | 3.50.80.10 |
| 145 | 1at0 | 2.170.16.10 |
| 152 | 1ass | 3.50.7.10 |
| 159 | 1a17 | 1.25.40.10 |
| 164 | 1ib8 | 3.30.300.70, 2.30.30.180 |
| 164 | 1ezg | 2.160.20.50 |
| 168 | 1h6q | 2.170.150.10 |
| 191 | 1b71 | 1.20.1260.10, 2.20.28.10 |

| | | |
|---|---|---|
| 199 | 1dnl | 2.30.110.10 |
| 200 | 1gen | 2.110.10.10 |
| 202 | 1kuu | 3.60.20.20 |
| 211 | 1vzv | 3.20.16.10 |
| 213 | 1lbu | 1.10.101.10, 3.30.1380.10 |
| 220 | 1ddk | 3.60.15.10 |
| 233 | 1lrv | 1.25.10.10 |
| 244 | 1s2o | 3.40.50.1000 |
| 250 | 1amk | 3.20.20.70 |
| 254 | 1bg5 | 3.40.30.10, 1.20.1050.10, 4.10.6.10 |
| 258 | 1plr | 3.70.10.10 |
| 259 | 1eza | 3.50.30.10, 1.10.274.10 |
| 262 | 1lxa | 2.160.10.10, 1.20.1180.10 |
| 264 | 1mat | 3.90.230.10 |
| 268 | 1ako | 3.60.10.10 |
| 274 | 2plc | 3.20.20.190 |
| 274 | 1bwz | 3.10.310.10 |
| 277 | 1a80 | 3.20.20.100 |
| 281 | 1t71 | 3.60.21.10 |
| 282 | 1aj2 | 3.20.20.20 |
| 285 | 1qtw | 3.20.20.150 |
| 286 | 1tml | 3.20.20.40 |
| 290 | 1x2j | 2.120.10.80 |
| 294 | 1c3d | 1.50.10.20 |
| 295 | 1hq0 | 3.60.100.10 |
| 296 | 1aua | 3.40.525.10, 1.10.8.20 |
| 300 | 1a3h | 3.20.20.80 |
| 301 | 1bqb | 3.10.170.10 |
| 315 | 1gyd | 2.115.10.20 |
| 327 | 1hmy | 3.40.50.150, 3.90.120.10 |
| 347 | 1jvb | 3.40.50.720 |
| 349 | 1add | 3.20.20.140 |
| 351 | 1qaz | 1.50.10.100 |
| 353 | 1h6l | 2.120.10.30 |
| 360 | 1jdw | 3.75.10.10 |
| 372 | 1pud | 3.20.20.105 |
| 374 | 1atu | 2.30.39.10, 3.30.497.10 |
| 374 | 1qlj | 3.90.180.10, 3.40.50.720 |
| 379 | 3sil | 2.120.10.10 |
| 418 | 1a2n | 3.65.10.10 |
| 440 | 1otp | 1.20.970.10, 3.40.1030.10, 3.90.1170.30 |
| 451 | 1ulz | 3.40.50.20 |
| 457 | 2bnh | 3.80.10.10 |
| 459 | 1a6r | 3.90.70.10 |
| 470 | 1akl | 2.150.10.10, 3.40.390.10 |
| 492 | 1ayx | 1.50.10.10 |
| 500 | 1a8h | 3.40.50.620, 2.170.220.10, 1.10.730.10 |
| 507 | 1dpe | 3.90.76.10, 3.40.190.10, 3.10.105.10 |

| | | |
|---|---|---|
| 558 | 1uok | 2.60.40.1180 |
| 629 | 1g9g | 1.50.10.10, 2.170.160.10, 4.10.870.10 |
| 639 | 1gof | 2.60.120.260, 2.130.10.80, 2.60.40.10 |
| 752 | 1j0m | 1.50.10.10 |

**Supplementary Table S2**. Number of aa residues and PDB/CATH codes of proteins in the CHARMM20 dataset

| # of aa residues | PDB code | CATH code |
|---|---|---|
| 40 | 3e21 | 1-10-8-10 |
| 51 | 3f2u | 2.40.50.40 |
| 90 | 4mmj | Unassigned |
| 106 | 4ruv | 3.50.7.10 |
| 124 | 4wyn | 3-10-130-10 |
| 125 | 5d82 | 3-10-450-50 |
| 155 | 5c9y | 3.30.530.20 |
| 179 | 4rz9 | Unassigned |
| 197 | 4db6 | 1-25-10-10 |
| 221 | 4uuc | 1-25-40-20 |
| 254 | 4rsw | Unassigned |
| 256 | 4n92 | 3-40-710-10 |
| 226 | 3wp9 | Unassigned |
| 281 | 5cen | 1-10-510-10 |
| 287 | 4nzz | 3-40-50-1820 |
| 311 | 3zo5 | 3-10-20-90 |
| 320 | 4y9s | 3-40-50-1820 |
| 389 | 4yhe | 3-20-20-80 |
| 659 | 4fnv | Unassigned |
| 281 | 4rhg | Unassigned |

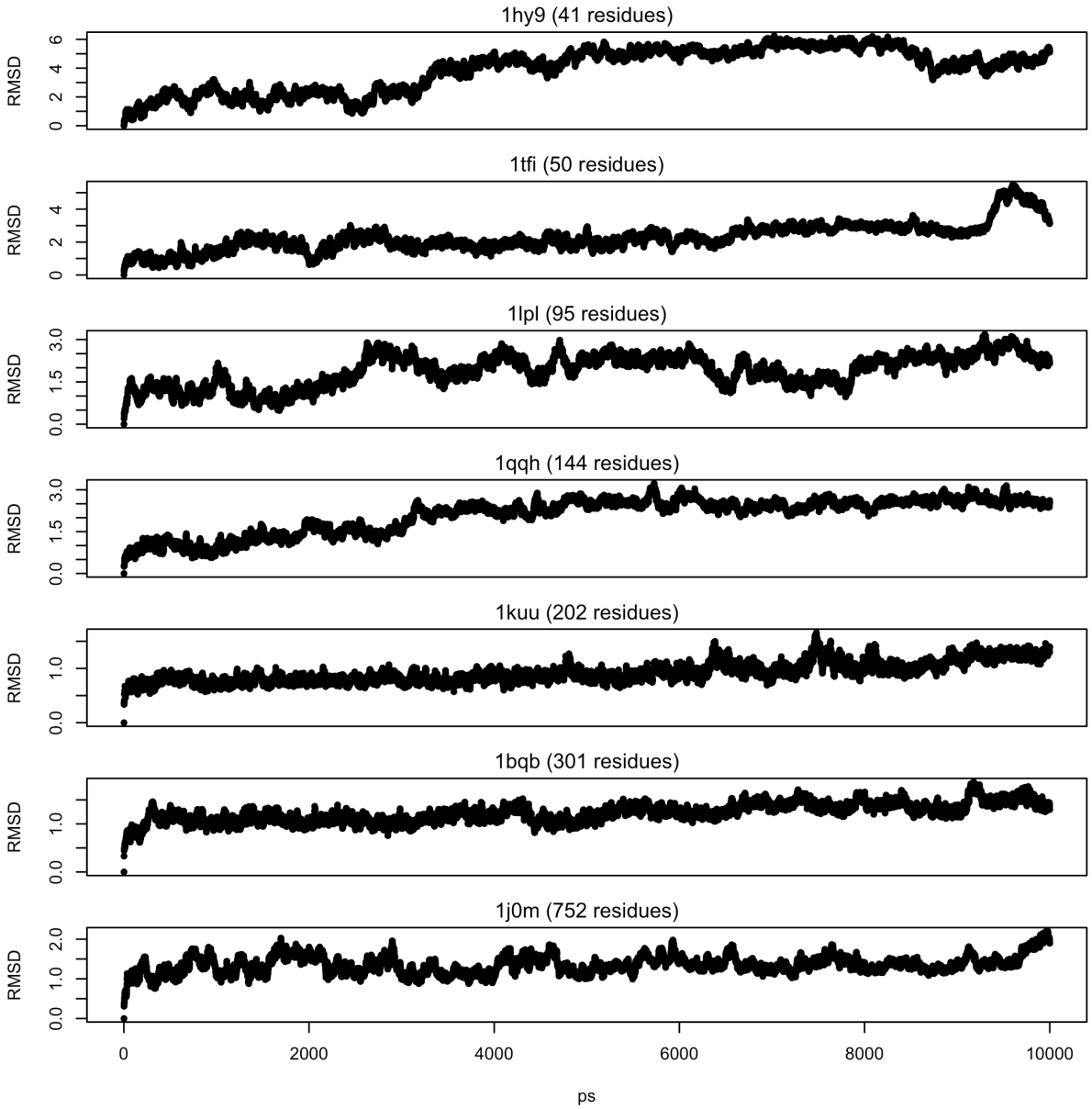


Figure S1. RMSD as a function of simulation time for proteins of various sizes.

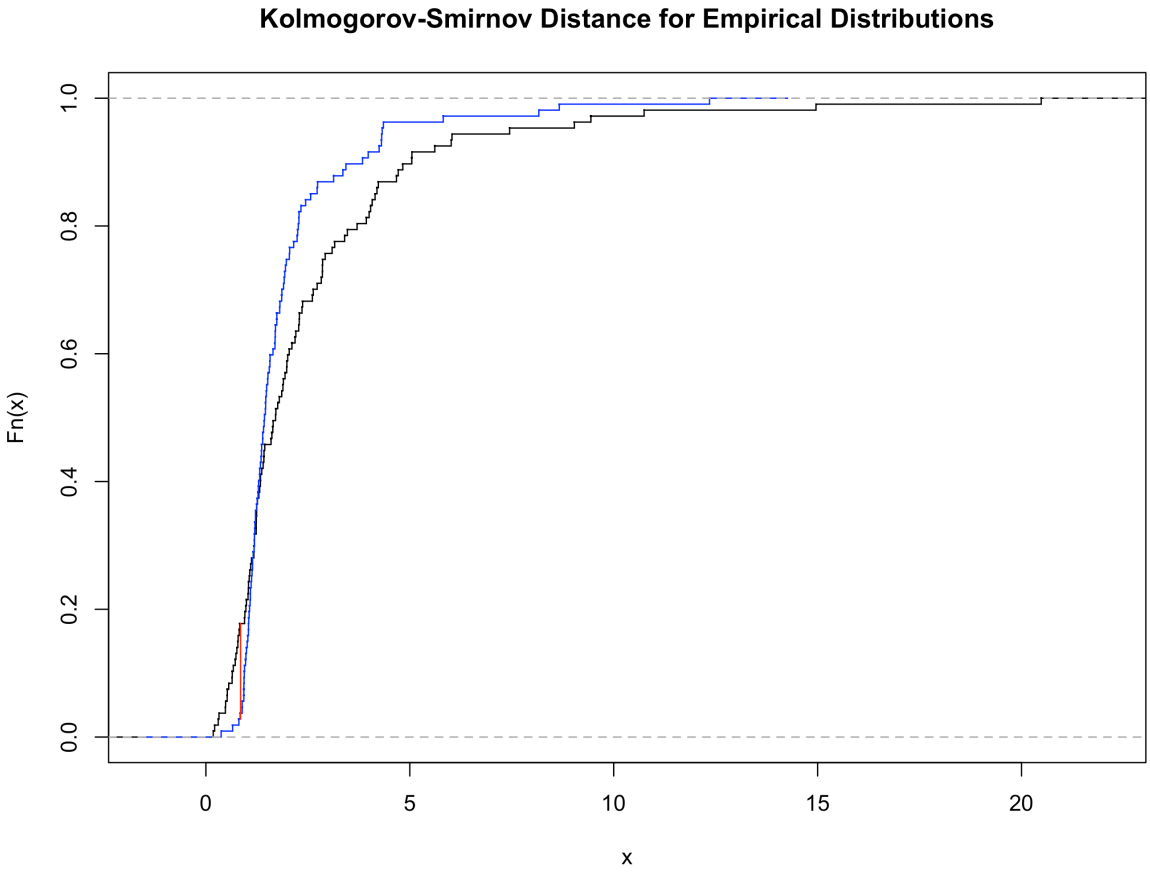


Figure S2. The length of the red segment shows the Kolmogorov-Smirnov distance between experimental cumulative distributions of $d_i^2$ (squares of distances between corresponding atoms) for two different frames of the trajectory (1bet pdb, all frames aligned to the initial frame). Kolmogorov-Smirnov distance is the greatest difference of $F_n(x)$ for two distributions.

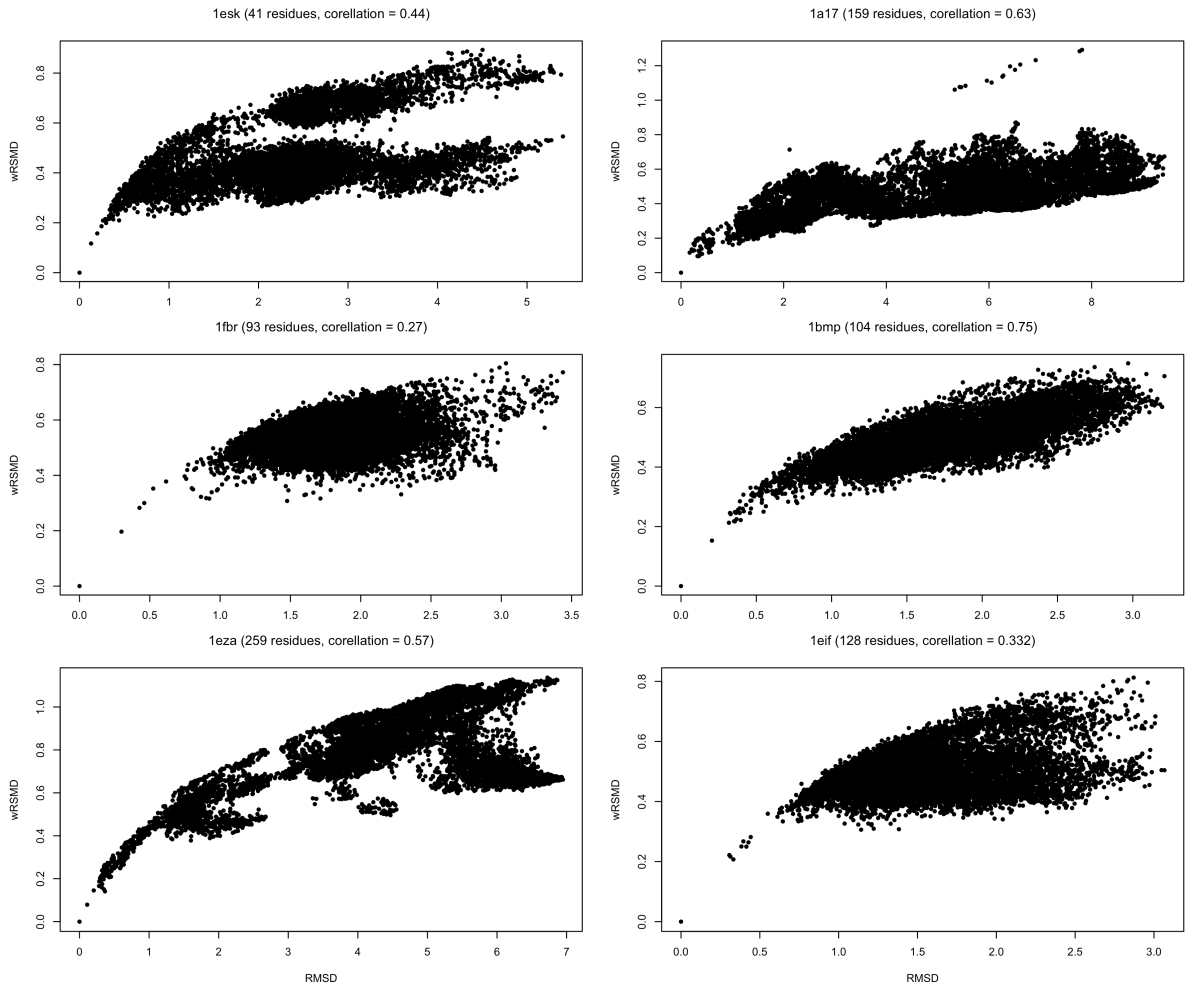


Figure S3. RMSD vs wRMSD plot for various proteins.