Table 1: Selected MC and REMC protocols

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Name | Walker temperature(s) kT(kcal/mol) | Trajectory lentgh (steps) | Move probabilitiesa rot;mut;mut+rot;  mut+mut | Walker swap periodicityb |
| MC | 0.2 | 6 109 | 0; 1; 0.1; 0 | - |
| REMCa | 0.125; 0.25; 0.5; 1 | 1.5 109 | 1; 0; 0.1; 0 | 7.5 106 |
| REMCb | 0.25; 0.5; 1; 2 | 1.5 109 | 1; 0; 0.1; 0 | 7.5 106 |
| REMCc | 0.175; 0.263; 0.395; 0.592; 0.888; 1.333; 2; 3 | 0.75 109 | 0; 1; 0.1; 0 | 7.5 106 |
| REMCd | 0.175; 0.263; 0.395; 0.592; 0.888; 1.333; 2; 3 | 0.75 109 | 1; 0; 0.1; 0 | 7.5 106 |
| REMCe | 0.175; 0.263;0.395;0.592; 0.888; 1.333; 2; 3 | 0.75 109 | 1; 0; 0.1; 0 | 106 |

aProbabilities at each MC step to change, respectively, a rotamer; a sidechain type; a type at one position and a rotamer at another; a type at two positions.bThe interval between attempts to exchange states between two walkers (using a Metropolis test).

Table 2: Reference energies for unfolded state (kcal/mol)

|  |  |  |  |
| --- | --- | --- | --- |
| Ala | -8.519 | Ile | -10.530 |
| Arg | -22.107 | Leu | -12.227 |
| Asn | -16.715 | Lys | -20.999 |
| Asp | -19.699 | Met | -11.985 |
| Cys | -9.426 | Phe | -18.017 |
| Gln | -17.718 | Ser | -11.881 |
| Glu | -20.048 | Thr | -11.413 |
| His(d) | -17.824 | Trp | -19.267 |
| His(e) | -17.824 | Tyr | -20.893 |
| His+ | -21.704 | Val | -9.5654 |

Three His variants are given, protonated on ND1 (d), NE2 (e), or both (+).

Table 3: Test proteins

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| type | PDB | length | acronym | type | PDB | length | acronym |
| PDZ | 1G9O | 91 | NHERF | SH2 | 1A81 | 108 | Syk kinase |
| PDZ | 1R6J | 82 | syntenin | SH2 | 1BM2 | 98 | Grb2 |
| PDZ | 2BYG | 97 | DLH2 | SH2 | 1M61 | 109 | Zap70 |
| SH3 | 1ABO | 58 | Abl | SH2 | 1O4C | 104 | Src kinase |
| SH3 | 1CKA | 57 | c-Crk |  |  |  |  |

Table 3: Designed sequence quality measures

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Number of | Identity | Superfamily tests | | | | |
| Protein | sequences  tested | % to  wildtype | Match  length | Superfamily  E-value | Superfamily  success rate | Family  E-value | Family  success rate |
| 1A81 | 236 | 27 | none |  |  |  |  |
| 1ABO | 203 | 32 | 51/58 | 4.4e-4 | 100% | 2.8e-3 | 100% |
| 1BM2 | 209 | 27 | 78/98 | 4.2e-5 | 100% | 2.6e-3 | 100% |
| 1CKA | 416 | 33 | 40/57 | 1.1e-5 | 100% | 3.4e-3 | 100% |
| 1G9O | 338 | 36 | 79/91 | 7.0e-7 | 100% | 2.5e-3 | 100% |
| 1M61 | 405 | 42 | 97/109 | 7.2e-7 | 100% | 2.6e-4 | 100% |
| 1O4C | 274 | 21 | 95/104 | 2.1e-4 | 100% | 4.5e-3 | 100% |
| 1R6J | 270 | 34 | 74/82 | 9.8e-6 | 100% | 4.6e-3 | 100% |
| 2BYG | 426 | 28 | 59/97 | 1.4e-5 | 100% | 7.1e-3 | 100% |

Table 4: Tests with 10 designed positions

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| rotamersa | lengthb | Protein | CFNc | Heur.d | MC | REMC |
| 2991 | 108(17) | 1A81 3 | gmec | 0.001 | 0.1595 |  |
|  |  | 1A81 4 | gmec | 0. | 0.0317 |  |
|  |  | 1A81 5 | gmec | 0. | 0.0563 |  |
| 2520 | 58(8) | 1ABO 1 | gmec | 0.0675 | 0.9054 | 0.8041 |
|  |  | 1ABO 4 | gmec | 0. | 0.0128 |  |
| 2957 | 98(10) | 1BM2 1 | gmec | 0. | 0.0950 |  |
|  |  | 1BM2 5 | gmec | 0. | 0.1082 |  |
| 2508 | 57(8) | 1CKA 5 | gmec | 0.2859 | 3.2525 | 0. |
| 2819 | 91(15) | 1G9O 3 | gmec | 0.1366 | 0.1366 |  |
|  |  | 1G9O 5 | gmec | 0. | 3.9599 | 0. |
| 2957 | 109(21) | 1M61 1 | gmec | 0. | 0.0776 |  |
|  |  | 1M61 2 | gmec | 3.5105 | 4.5062 | 0.3215 |
|  |  | 1M61 5 | gmec | 0. | 0.0432 |  |
| 3037 | 104(8) | 1O4C 1 | gmec | 0. | 0.1121 |  |
|  |  | 1O4C 2 | gmec | 0. | 0.1046 |  |
|  |  | 1O4C 3 | gmec | 0. | 0.1519 |  |
|  |  | 1O4C 4 | gmec | 0. | 0.1545 |  |
|  |  | 1O4C 5 | gmec | 0. | 0.1753 |  |
| 2773 | 82(10) | 1R6J 1 | gmec | 0. | 2.4022 | 0.3986 |
|  |  | 1R6J 2 | gmec | 0. | 1.0398 | 0.3049 |
|  |  | 1R6J 3 | gmec | 0. | 0.0106 |  |
|  |  | 1R6J 5 | gmec | 0. | 0.0162 |  |
| 2888 | 97(15) | 2BYG 1 | 5.7485 | 0. | 0.0337 |  |
|  |  | 2BYG 3 | gmec | 0. | 0.0833 |  |
|  |  | 2BYG 4 | gmec | 0. | 0.2149 |  |

aTotal number of rotamers available to the system. Each designed position can explore 206 rotamers; the others explore about 10 rotamers each. bTotal protein length (number of Gly+Pro in parentheses). cgmec indicates the GMEC was successfully identified. dFor all four exploration methods and each test, we report the difference between the best energy obtained and the overall best energy (the best over all methods, which may or not be the GMEC). 10-position tests where all four methods found the GMEC are not listed.

Table 5: Tests with 20 and 30 designed positions

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 20 positions | | | | | | 30 positions | | | |
| Protein | CFN | Heur. | MC | REMC | mutationsa | CFN | | Heur. | MC | REMC |
| 1A81 1 | gmec\* | 0. | 0.3275 | 0.3851 | 0 | 1.2074 | | 0. | 0.6353 |  |
| 1A81 2 | gmec\* | 0.1705 | 2.4355 | 1.0069 | 3 | 2.5520 | | 0. | 0.0578 |  |
| 1A81 3 | gmec | 0. | 0.4640 | 0.6186 | 0 | 43.5263 | | 0. | 2.4996 | 1.2025 |
| 1A81 4 | gmec | 0.3878 | 0.5748 | 0.6991 | 4 | 5.1300 | | 0. | 0.0305 |  |
| 1A81 5 | gmec | 0.0068 | 0.5088 | 0.1541 | 4 | 3.2417 | | 0. | 1.9586 | 0.5791 |
| 1ABO 1 | gmec | 0.1205 | 1.1159 | 0.2153 | 2 | 44.5504 | | 0. | 0. |  |
| 1ABO 2 | 13.8563 | 0. | 0. | 0. | 8 | 12.7303 | | 0. | 0. |  |
| 1ABO 3 | 1.2190 | 0. | 0. | 0. | 9 | 9.3870 | | 0. | 0.2630 |  |
| 1ABO 4 | 1.9940 | 0. | 0.0076 | 0. | 5 | 10.7691 | | 0. | 0. |  |
| 1ABO 5 | 3.5418 | 0. | 0.9483 | 0.9483 | 9 | 4.3907 | | 0. | 0. |  |
| 1BM2 1 | gmec | 0. | 0.0619 | 0.1584 | 0 | 22.5876 | | 0. | 1.7290 | 1.6013 |
| 1BM2 2 | 7.5304 | 0. | 0.0725 | 0.0143 | 8 | 22.1386 | | 0. | 1.9856 | 1.5876 |
| 1BM2 3 | gmec | 0.0229 | 0.4762 | 0.2897 | 0 | 22.5410 | | 0. | 1.9990 | 1.1541 |
| 1BM2 4 | 0.1186 | 0. | 2.5883 | 0.0789 | 2 | 15.2639 | | 0. | 2.2127 | 2.3854 |
| 1BM2 5 | gmec | 0.2396 | 0.3746 | 0.3746 | 3 | 15.9890 | | 0. | 2.8354 | 1.1937 |
| 1CKA 1 | gmec\* | 0. | 0. | 0. | 0 | 6.2700 | | 0. | 0. |  |
| 1CKA 2 | gmec | 0. | 0. | 0. | 0 | 2.0995 | | 0. | 0. |  |
| 1CKA 3 | gmec | 0. | 0. | 0. | 0 | 47.0217 | | 0. | 0. |  |
| 1CKA 4 | 4.3122 | 0. | 0. | 0. | 4 | 44.0830 | | 0. | 0. |  |
| 1CKA 5 | 4.2849 | 0. | 0. | 0. | 3 | 8.8608 | | 0. | 0. |  |
| 1G9O 1 | 2.0574 | 0. | 1.2525 | 1.2525 | 5 | 2.0816 | | 0. | 1.5942 | 0. |
| 1G9O 2 | 3.2106 | 0. | 0.2177 | 0.1915 | 1 | 0.3270 | | 0. | 0.3126 |  |
| 1G9O 3 | 1.9008 | 0. | 0.4417 | 0.1019 | 1 | 17.7150 | | 0. | 1.5667 | 1.5667 |
| 1G9O 4 | 0.5030 | 0. | 0.3855 | 0.1455 | 5 | 2.9758 | | 0. | 1.4284 | 1.6202 |
| 1G9O 5 | 0.4298 | 0. | 0.1495 | 0.5114 | 5 | 0. | | 1.6890 | 7.6985 | 2.3857 |
| 1M61 1 | gmec | 0. | 0. | 0. | 0 | 14.4935 | | 0.0097 | 0. | 0. |
| 1M61 2 | gmec | 0. | 0. | 0. | 0 | 5.0899 | | 0. | 1.8749 | 0.008 |
| 1M61 3 | gmec | 0. | 0. | 0. | 0 | 3.5795 | | 0. | 0.0154 |  |
| 1M61 4 | gmec | 0. | 0. | 0. | 0 | 16.1511 | | 0. | 0. |  |
| 1M61 5 | gmec | 0. | 0.2521 | 0.1345 | 0 | 23.0927 | | 0. | 0. |  |
| 1O4C 1 | 0. | 0.3465 | 0.0690 | 0.0587 | 6 | 14.9064 | | 0. | 0.3435 |  |
| 1O4C 2 | 6.4214 | 0. | 0.1963 | 0.3175 | 4 | 58.1558 | | 0. | 0.0795 |  |
| 1O4C 3 | gmec | 0. | 0.3461 | 0.0997 | 0 | 9.9221 | | 0. | 0.1789 |  |
| 1O4C 4 | gmec | 0. | 0.3640 | 0.1382 | 0 | 5.7790 | | 0. | 0.0423 |  |
| 1O4C 5 | 0. | 0. | 0.1131 | 0.2206 | 0 | 9.9221 | | 0. | 0.1789 |  |
| 1R6J 1 | gmec | 0. | 0.2604 | 0.2002 | 0 | gmec\* | | 0. | 0.0246 |  |
| 1R6J 2 | gmec | 0. | 0.0071 | 0.0183 | 0 | 14.9800 | | 0. | 0.0957 |  |
| 1R6J 3 | gmec | 0. | 0.0537 | 0.0732 | 0 | 0. | | 0. | 0.0440 |  |
| 1R6J 4 | gmec | 0. | 0.0639 | 0.0601 | 0 | 0. | | 0. | 0.0957 |  |
| 1R6J 5 | gmec | 0. | 0.0735 | 0.0244 | 0 | 0. | | 0.7036 | 1.8823 | 0.0781 |
| 2BYG 1 | gmec | 0. | 3.1878 | 0.0257 | 0 | 17.9752 | | 0. | 0.1592 |  |
| 2BYG 2 | gmec | 0. | 0.0524 | 0.0831 | 0 | 0.3832 | | 0. | 0.1502 |  |
| 2BYG 3 | gmec\* | 0. | 1.3564 | 0.0826 | 0 | 0.1442 | | 0. | 0.1593 |  |
| 2BYG 4 | gmec | 0. | 0.1968 | 0.6022 | 0 | 0. | | 0.0958 | 0.0050 |  |
| 2BYG 5 | 1.8604 | 0. | 0.0933 | 0.0386 | 2 | 0.5003 | | 0. | 0.6876 |  |

Format as in Table [4](#tab:10act). gmec\* indicates the more aggressive protocol. aBetween CFN/Heur.

Table 6: Designed and Pfam sequence entropies

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Top 10,000 | Top 10,000 | Pfam | Pfam |
| Protein | structures | sequences | seed | full |
| 1ABO | 1.36 | 1.58 | 2.79 | 3.01 |
| 1CKA | 1.20 | 1.41 | 2.84 | 3.03 |
| 1R6J | 1.33 | 1.48 | 3.11 | 3.66 |
| 1G9O | 1.21 | 1.53 | 3.29 | 3.81 |
| 2BYG | 1.57 | 1.63 | 3.31 | 3.67 |
| 1BM2 | 1.08 | 1.26 | 2.90 | 3.50 |
| 1O4C | 1.36 | 1.68 | 2.94 | 3.47 |
| 1M61 | 1.31 | 1.41 | 2.91 | 3.51 |
| 1A81 | 1.13 | 1.29 | 2.91 | 3.51 |

The entropies are exponentiated, then averaged over all positions. The designed entropies correspond to REMC runs where all positions are designed (except Gly/Pro).