

Table S4 The most 50 important AAindex mRMR features scores, amino acids with window positions and corresponding properties are presented.

| Ranks | Scores | Amino acids with window positions | Properties | Descriptions |
|-------|--------|---|------------|---|
| 1 | 0.025 | (T, -3) | RADA880107 | Energy transfer from out to in(95%buried) |
| 2 | 0.015 | (P, +11) | YUTK870102 | Unfolding Gibbs energy in water, pH9.0 |
| 3 | 0.009 | (I, +9) | RACS820102 | Average relative fractional occurrence in AR |
| 4 | -0.002 | (W, -9) | VHEG790101 | Transfer free energy to lipophilic phase |
| 5 | -0.007 | (E, -3) | RICJ880117 | Relative preference value at C |
| 6 | -0.022 | (A, +3) | ZASB820101 | Dependence of partition coefficient on ionic strength |
| 7 | -0.017 | (L, +10) | RICJ880107 | Relative preference value at N4 |
| 8 | -0.028 | (H,+12) | RADA880103 | Transfer free energy from vap to chx |
| 9 | -0.024 | (K, -3) | VASM830102 | Relative population of conformational state C |
| 10 | -0.022 | (Q, +12) | RICJ880101 | Relative preference value at N |
| 11 | -0.029 | (N, -7) | WERD780103 | Free energy change of alpha(Ri) to alpha(Rh) |
| 12 | -0.029 | (I, +5) | ROBB760105 | Information measure for extended |
| 13 | -0.03 | (F, +13) | VASM830102 | Relative population of conformational state C |
| 14 | -0.027 | (D, -1) | RACS820111 | Average relative fractional occurrence in E0(i-1) |
| 15 | -0.032 | (T, -13) | RICJ880116 | Relative preference value at C' |
| 16 | -0.032 | (H, +9) | QIAN880139 | Weights for coil at the window position of 6 |
| 17 | -0.031 | (M, +10) | ZIMJ680101 | Hydrophobicity |
| 18 | -0.032 | (E, +6) | ROBB760102 | Information measure for N-terminal helix |
| 19 | -0.035 | (E, -8) | YUTK870102 | Unfolding Gibbs energy in water, pH9.0 |
| 20 | -0.035 | (V, -13) | RICJ880102 | Relative preference value at N' |
| 21 | -0.036 | (G, -13) | WARP780101 | Average interactions per side chain atom |
| 22 | -0.037 | (D, -10) | TANS770109 | Normalized frequency of coil |
| 23 | -0.036 | (S, -2) | WOLS870103 | Principal property value z3 |
| 24 | -0.035 | (V, +1) | RADA880106 | Accessible surface area |
| 25 | -0.037 | (E, -11) | RICJ880107 | Relative preference value at N4 |
| 26 | -0.038 | (G, +6) | VHEG790101 | Transfer free energy to lipophilic phase |
| 27 | -0.037 | (D, -1) | RADA880106 | Accessible surface area |
| 28 | -0.039 | (S, +10) | VHEG790101 | Transfer free energy to lipophilic phase |
| 29 | -0.038 | (N,+4) | RACS820111 | Average relative fractional occurrence in E0(i-1) |
| 30 | -0.037 | (N, +13) | RICJ880116 | Relative preference value at C' |
| 31 | -0.037 | (D, -8) | RICJ880107 | Relative preference value at N4 |
| 32 | -0.037 | (Q, +5) | WERD780103 | Free energy change of alpha(Ri) to alpha(Rh) |
| 33 | -0.039 | (G, -12) | WERD780103 | Free energy change of alpha(Ri) to alpha(Rh) |
| 34 | -0.04 | (Q, -4) | RICJ880117 | Relative preference value at C'' |
| 35 | -0.039 | (R, -10) | RACS820113 | Value of theta(i) |
| 36 | -0.04 | (W, -4) | TANS770108 | Normalized frequency of zeta R |
| 37 | -0.041 | (Q, +3) | RICJ880114 | Relative preference value at C1 |
| 38 | -0.041 | (V, -2) | RICJ880115 | Relative preference value at C-cap |
| 39 | -0.043 | (W, -11) | RACS770103 | Side chain orientational preference |

| | | | | |
|----|--------|----------|------------|---|
| 40 | -0.043 | (P, +2) | TANS770110 | Normalized frequency of chain reversal |
| 41 | -0.043 | (N, -10) | WOEC730101 | Polar requirement |
| 42 | -0.043 | (L, +2) | RICJ880111 | Relative preference value at C4 |
| 43 | -0.045 | (E, +7) | RICJ880102 | Relative preference value at N' |
| 44 | -0.044 | (L, +13) | WOEC730101 | Polar requirement |
| 45 | -0.045 | (M, +8) | ROBB760105 | Information measure for extended |
| 46 | -0.045 | (N, +2) | RACS820101 | Average relative fractional occurrence in A0(i) |
| 47 | -0.044 | (M, -2) | WERD780102 | Free energy change of epsilon(i) to epsilon(ex) |
| 48 | -0.043 | (I, +13) | RICJ880113 | Relative preference value at C2 |
| 49 | -0.044 | (Q, +6) | RICJ880102 | Relative preference value at N' |
| 50 | -0.043 | (D, +6) | TANS770108 | Normalized frequency of zeta R |
