**Supplementary Table T1**

|  |
| --- |
| 1)Signal sequence helical potential |
| 2)Membrane-buried preference parameters |
| 3)Average flexibility indices |
| 4)Information value for accessibility; average fraction 23% |
| 5)Retention coefficient in TFA |
| 6) alpha-NH chemical shifts |
| 7) alpha-CH chemical shifts |
| 8)Normalized frequency of extended structure |
| 9)A parameter defined from the residuals obtained from the best correlation of theChou-Fasman parameter of beta-sheet |
| 10) Average volume of buried residue |
| 11) Frequency of the 2nd residue in turn |
| 12) Frequency of the 4th residue in turn |
| 13) Normalized frequency of the 2nd and 3rd residues in turn |
| 14) Normalized hydrophobicity scales for alpha-proteins |
| 15) Membrane preference for cytochrome b: MPH89 |
| 16) Average membrane preference: AMP07 |
| 17) Consensus normalized hydrophobicity scale |
| 18) Hydrophobic parameter pi |
| 19) Aperiodic indices |
| 20) Aperiodic indices for beta-proteins |
| 21) Aperiodic indices for alpha/beta-proteins |
| 22) Partition energy |
| 23) Heat capacity |
| 24) Absolute entropy |
| 25) Entropy of formation |
| 26) Normalized relative frequency of alpha-helix |
| 27) Average accessible surface area |
| 28) Sequence frequency |
| 29) Average relative probability of helix |
| 30) Average relative probability of inner helix |
| 31) Flexibility parameter for one rigid neighbor |
| 32) Side chain interaction parameter |
| 33) Side chain interaction parameter |
| 34) Distance between C-alpha and centroid of side chain |
| 35) Radius of gyration of side chain |
| 36) Normalized frequency of beta-sheet |
| 37) Normalized frequency of reverse turn |
| 38) Average surrounding hydrophobicity |
| 39) Normalized frequency of extended structure |
| 40) Refractivity |
| 41) Retention coefficient in HPLC |
| 42) Retention coefficient in NaClO4 |
| 43) Effective partition energy |
| 44) AA composition of total proteins |
| 45) Normalized composition from animal |
| 46) AA composition of mt-proteins from fungi and plant |
| 47) AA composition of membrane proteins |
| 48) Transmembrane regions of non-mt-proteins |
| 49) Transmembrane regions of mt-proteins |
| 50) AA composition of CYT of single-spanning proteins |
| 51) AA composition of CYT2 of single-spanning proteins |
| 52) AA composition of EXT of single-spanning proteins |
| 53) AA composition of EXT2 of single-spanning proteins |
| 54) AA composition of MEM of single-spanning proteins |
| 55) AA composition of CYT of multi-spanning proteins |
| 56) AA composition of EXT of multi-spanning proteins |
| 57) 8 A contact number |
| 58) 14 A contact number |
| 59) Average non-bonded energy per atom |
| 60) Long range non-bonded energy per atom |
| 61) Average non-bonded energy per residue |
| 62) Short and medium range non-bonded energy per residue |
| 63) Optimized beta-structure-coil equilibrium constant |
| 64) Optimized propensity to form reverse turn |
| 65) Optimized transfer energy parameter |
| 66) Optimized average non-bonded energy per atom |
| 67) Optimized side chain interaction parameter |
| 68) Normalized frequency of alpha-helix from CF |
| 69) Normalized frequency of alpha-helix in alpha+beta class |
| 70) Average gain in surrounding hydrophobicity |
| 71) Surrounding hydrophobicity in alpha-helix |
| 72) Average number of surrounding residues |
| 73) Hydrophobicity |
| 74) Relative frequency in reverse-turn |
| 75) Weights for alpha-helix at the window position of -3 |
| 76) Weights for alpha-helix at the window position of 0 |
| 77) Weights for alpha-helix at the window position of 3 |
| 78) Weights for beta-sheet at the window position of 0 |
| 79) Weights for coil at the window position of -1 |
| 80) Weights for coil at the window position of 0 |
| 81) Weights for coil at the window position of 1 |
| 82) Weights for coil at the window position of 5 |
| 83) Average reduced distance for side chain |
| 84) Average relative fractional occurrence in AR (i) |
| 85) Average relative fractional occurrence in A0 (i-1) |
| 86) Value of theta (i) |
| 87) Transfer free energy from vap to chx |
| 88) Accessible surface area |
| 89) Information measure for middle helix |
| 90) Mean area buried on transfer |
| 91) Principal component I |
| 92) Principal component III |
| 93) Principal component IV |
| 94) Normalized frequency of isolated helix |
| 95) Normalized frequency of extended structure |
| 96) Normalized frequency of chain reversal R |
| 97) Normalized frequency of coil |
| 98) Relative population of conformational state A |
| 99) Relative population of conformational state C |
| 100)Transfer free energy to lipophilic phase |
| 101)Average interactions per side chain atom |
| 102)Hydration potential |
| 103)Principal property value z1 |
| 104)Principal property value z2 |
| 105)Principal property value z3 |
| 106)Activation Gibbs energy of unfolding |
| 107)Normalized positional residue frequency at helix termini C3 |
| 108)Normalized flexibility parameters |
| 109)Free energy in alpha-helical region |
| 110)Free energy in beta-strand conformation |
| 111) p-Values of thermophilic proteins based on the distributions of B values |
| 112)Distribution of amino acid residues in the 18 non-redundant families of thermophilicproteins |
| 113)Distribution of amino acid residues in the alpha-helices in thermophilic mesophilicproteins |
| 114)Hydropathy scale based on self-information values in the two-state model (16%accessibility) |
| 115)Hydropathy scale based on self-information values in the two-state model (25%accessibility) |
| 116)Hydropathy scale based on self-information values in the two-state model (50%accessibility) |
| 117)Alpha-helix propensity derived from designed sequences |
| 118)Surface composition of amino acids in intracellular proteins of thermophiles |
| 119)Surface composition of amino acids in intracellular proteins of mesophiles |
| 120)Surface composition of amino acids in nuclear proteins |
| 121)Interior composition of amino acids in intracellular proteins of thermophiles |
| 122)Interior composition of amino acids in intracellular proteins of mesophiles |
| 123)Interior composition of amino acids in extracellular proteins of mesophiles |
| 124)Entire chain composition of amino acids in intracellular proteins of thermophiles |
| 125)Entire chain composition of amino acids in extracellular proteins of mesophiles |
| 126)Entire chain compositino of amino acids in nuclear proteins |
| 127)Volumes including the crystallographic waters using the ProtOr |
| 128)Volumes not including the crystallographic waters using the ProtOr |
| 129)Hydrophobicity scales |
| 130)Hydrophobicity coefficient in RP-HPLC |
| 131)Hydrophobicity coefficient in RP-HPLC |
| 132)Interactivity scale obtained from the contact matrix |
| 133)Interactivity scale obtained by maximizing the mean of correlation coefficient over |
| single-domain globular proteins |
| 134)Interactivity scale obtained by maximizing the mean of correlation coefficient over |
| pairs of sequences sharing the TIM barrel fold |
| 135)Linker propensity index |
| 136)Linker propensity from 2-linker dataset |
| 137)Linker propensity from medium dataset |
| 138)Linker propensity from long dataset |
| 139)Linker propensity from helical (annotated by DSSP) dataset |
| 140)Linker propensity from non-helical (annotated by DSSP) dataset |
| 141)Linker index |
| 142)Mean volumes of residues buried in protein interiors |
| 143)Average volumes of residues |
| 144)Hydrostatic pressure asymmetry index |
| 145)Average internal preferences |
| 146)Apparent partition energies calculated from Janin index |
| 147)Apparent partition energies calculated from Chothia index |
| 148)Weights from the IFH scale |
| 149)Hydrophobicity index |
| 150)NNEIG index |
| 151)SWEIG index |
| 152)PRIFT index |
| 153)ALTFT index |
| 154)ALTLS index |
| 155)TOTFT index |
| 156)TOTLS index |
| 157)Relative partition energies derived by the Bethe approximation |
| 158)Hydrophobicity index |

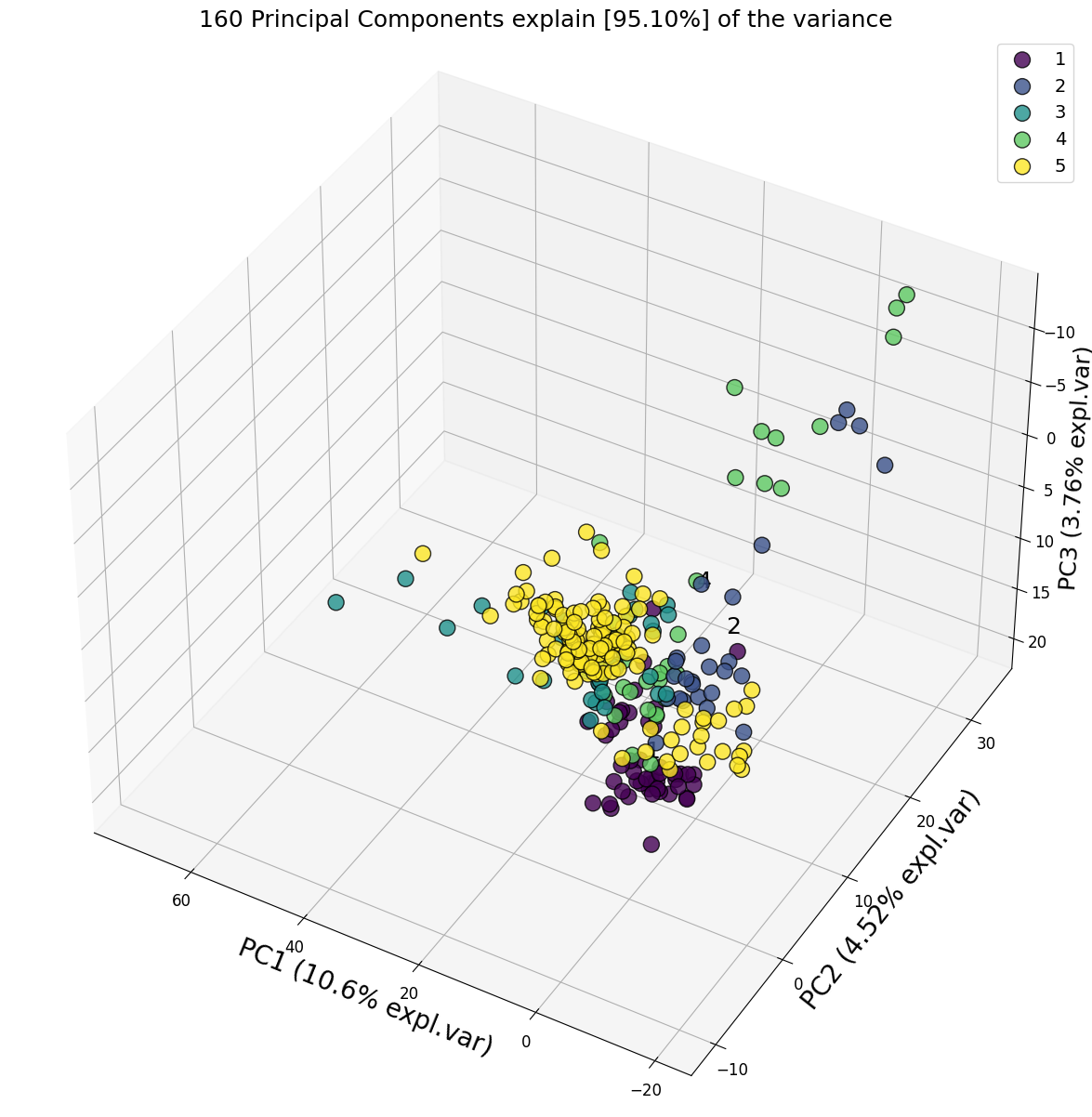
**Supplementary Table T2**

|  |  |
| --- | --- |
| ***Feature*** | **Type** |
| FEGS\_70 | FEGS |
| FEGS\_2 | FEGS |
| FEGS\_5 | FEGS |
| FEGS\_431 | FEGS |
| FEGS\_546 | FEGS |
| FEGS\_440 | FEGS |
| FEGS\_435 | FEGS |
| FEGS\_412 | FEGS |
| FEGS\_353 | FEGS |
| FEGS\_529 | FEGS |
| FEGS\_235 | FEGS |
| FEGS\_273 | FEGS |
| FEGS\_518 | FEGS |
| FEGS\_403 | FEGS |
| FEGS\_572 | FEGS |
| FEGS\_357 | FEGS |
| FEGS\_360 | FEGS |
| FEGS\_174 | FEGS |
| FEGS\_471 | FEGS |
| FEGS\_355 | FEGS |
| FEGS\_331 | FEGS |
| FEGS\_438 | FEGS |
| FEGS\_372 | FEGS |
| FEGS\_507 | FEGS |
| FEGS\_284 | FEGS |
| FEGS\_516 | FEGS |
| FEGS\_320 | FEGS |
| FEGS\_551 | FEGS |
| FEGS\_324 | FEGS |
| FEGS\_355 | FEGS |
| FEGS\_195 | FEGS |
| FEGS\_233 | FEGS |
| FEGS\_493 | FEGS |
| FEGS\_488 | FEGS |
| FEGS\_421 | FEGS |
| FEGS\_327 | FEGS |
| FEGS\_195 | FEGS |
| FEGS\_465 | FEGS |
| FEGS\_296 | FEGS |
| FEGS\_377 | FEGS |
| FEGS\_332 | FEGS |
| FEGS\_577 | FEGS |
| FEGS\_267 | FEGS |
| FEGS\_329 | FEGS |
| FEGS\_186 | FEGS |
| FEGS\_292 | FEGS |
| FEGS\_560 | FEGS |
| FEGS\_321 | FEGS |
| FEGS\_486 | FEGS |
| FEGS\_362 | FEGS |
| FEGS\_259 | FEGS |
| FEGS\_571 | FEGS |
| FEGS\_312 | FEGS |
| FEGS\_477 | FEGS |
| FEGS\_437 | FEGS |
| FEGS\_325 | FEGS |
| FEGS\_415 | FEGS |
| FEGS\_475 | FEGS |
| FEGS\_401 | FEGS |
| FEGS\_565 | FEGS |
| FEGS\_453 | FEGS |
| FEGS\_447 | FEGS |
| FEGS\_500 | FEGS |
| FEGS\_469 | FEGS |
| FEGS\_480 | FEGS |
| FEGS\_196 | FEGS |
| FEGS\_318 | FEGS |
| FEGS\_471 | FEGS |
| FEGS\_502 | FEGS |
| FEGS\_331 | FEGS |
| FEGS\_409 | FEGS |
| FEGS\_293 | FEGS |
| FEGS\_502 | FEGS |
| FEGS\_375 | FEGS |
| FEGS\_190 | FEGS |
| FEGS\_294 | FEGS |
| FEGS\_399 | FEGS |
| FEGS\_482 | FEGS |
| FEGS\_307 | FEGS |
| FEGS\_260 | FEGS |
| FEGS\_210 | FEGS |
| FEGS\_309 | FEGS |
| FEGS\_508 | FEGS |
| FEGS\_390 | FEGS |
| FEGS\_408 | FEGS |
| FEGS\_358 | FEGS |
| FEGS\_299 | FEGS |
| FEGS\_174 | FEGS |
| FEGS\_379 | FEGS |
| FEGS\_450 | FEGS |
| FEGS\_511 | FEGS |
| FEGS\_405 | FEGS |
| FEGS\_509 | FEGS |
| FEGS\_521 | FEGS |
| FEGS\_390 | FEGS |
| FEGS\_241 | FEGS |
| FEGS\_473 | FEGS |
| FEGS\_189 | FEGS |
| PSSM\_HOG92 | PSSM |
| PSSM\_HOG105 | PSSM |
| PSSM\_HOG149 | PSSM |
| PSSM\_HOG28 | PSSM |
| PSSM\_HOG188 | PSSM |
| PSSM\_HOG157 | PSSM |
| PSSM\_HOG52 | PSSM |
| PSSM\_HOG81 | PSSM |
| PSSM\_HOG106 | PSSM |
| PSSM\_HOG18 | PSSM |
| PSSM\_HOG88 | PSSM |
| PSSM\_HOG15 | PSSM |
| PSSM\_HOG215 | PSSM |
| PSSM\_HOG96 | PSSM |
| PSSM\_HOG31 | PSSM |
| PSSM\_HOG106 | PSSM |
| PSSM\_HOG93 | PSSM |
| PSSM\_HOG136 | PSSM |
| PSSM\_HOG17 | PSSM |
| PSSM\_HOG187 | PSSM |
| PSSM\_HOG147 | PSSM |
| PSSM\_HOG48 | PSSM |
| PSSM\_HOG15 | PSSM |
| PSSM\_HOG62 | PSSM |
| PSSM\_HOG132 | PSSM |
| PSSM\_HOG78 | PSSM |
| PSSM\_HOG2 | PSSM |
| PSSM\_HOG186 | PSSM |
| PSSMSeg\_CS41 | PSSMSeg |
| PSSMSeg\_CS16 | PSSMSeg |
| PSSMSeg\_CS50 | PSSMSeg |
| PSSMSeg\_CS44 | PSSMSeg |
| PSSMSeg\_CS26 | PSSMSeg |
| PSSMSeg\_CS8 | PSSMSeg |
| PSSMSeg\_CS30 | PSSMSeg |
| SAAC\_29 | SAAC |
| SAAC\_57 | SAAC |
| SAAC\_54 | SAAC |
| SAAC\_14 | SAAC |
| SAAC\_53 | SAAC |
| SAAC\_60 | SAAC |
| SAAC\_12 | SAAC |
| SAAC\_16 | SAAC |
| SAAC\_47 | SAAC |
| SAAC\_47 | SAAC |
| SAAC\_13 | SAAC |
| SAAC\_14 | SAAC |
| SAAC\_20 | SAAC |
| SAAC\_50 | SAAC |
| SAAC\_41 | SAAC |
| SAAC\_25 | SAAC |
| SAAC\_48 | SAAC |
| SAAC\_51 | SAAC |
| SAAC\_48 | SAAC |
| SAAC\_48 | SAAC |
| SAAC\_47 | SAAC |
| SAAC\_14 | SAAC |
| SAAC\_56 | SAAC |
| SAAC\_50 | SAAC |
| SAAC\_59 | SAAC |
| SAAC\_38 | SAAC |
| SAAC\_43 | SAAC |

**Supplementary Table T3**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Model/Class | Precision | Recall | Acc | Specificity | F1 score |
| **CatBooPt** |  |  |  |  |  |
| P1 | 0.87 | 0.92 | 0.96 | 0.97 | 0.89 |
| P2 | 0.88 | 0.58 | 0.95 | 0.99 | 0.70 |
| P3 | 0.94 | 0.97 | 0.99 | 0.99 | 0.96 |
| P4 | 0.87 | 0.68 | 0.95 | 0.99 | 0.76 |
| P5 | 0.90 | 0.98 | 0.94 | 0.90 | 0.94 |
| **DT** |  |  |  |  |  |
| P1 | 0.70 | 0.80 | 0.90 | 0.92 | 0.75 |
| P2 | 0.33 | 0.31 | 0.88 | 0.94 | 0.32 |
| P3 | 0.69 | 0.63 | 0.92 | 0.96 | 0.66 |
| P4 | 0.56 | 0.65 | 0.90 | 0.93 | 0.60 |
| P5 | 0.85 | 0.81 | 0.84 | 0.87 | 0.83 |
| **GNB** |  |  |  |  |  |
| P1 | 0.87 | 0.80 | 0.94 | 0.97 | 0.83 |
| P2 | 0.48 | 0.58 | 0.90 | 0.94 | 0.53 |
| P3 | 0.97 | 0.83 | 0.97 | 1.00 | 0.89 |
| P4 | 0.60 | 0.48 | 0.90 | 0.96 | 0.54 |
| P5 | 0.87 | 0.93 | 0.90 | 0.87 | 0.90 |
| **MLP** |  |  |  |  |  |
| P1 | 0.88 | 0.86 | 0.95 | 0.97 | 0.87 |
| P2 | 0.82 | 0.69 | 0.96 | 0.98 | 0.75 |
| P3 | 0.94 | 0.94 | 0.99 | 0.99 | 0.94 |
| P4 | 0.79 | 0.84 | 0.96 | 0.97 | 0.81 |
| P5 | 0.95 | 0.97 | 0.96 | 0.95 | 0.96 |
| **xGB** |  |  |  |  |  |
| P1 | 0.85 | 0.94 | 0.96 | 0.96 | 0.90 |
| P2 | 0.87 | 0.50 | 0.95 | 0.99 | 0.63 |
| P3 | 0.94 | 0.91 | 0.98 | 0.99 | 0.93 |
| P4 | 0.81 | 0.68 | 0.95 | 0.98 | 0.74 |
| P5 | 0.90 | 0.98 | 0.94 | 0.90 | 0.94 |

**Supplementary Figure S1**



**Supplementary Figure S2**

A graph with a curved line

Description automatically generated