Indices—Features in the Clusters

Cluster 1

1 — Frequency of the 2nd residue in turn (Chou-Fasman, 1978b) Feature 50

2 — Normalized relative frequency of bend R (Isogai et al., 1980) Feature 122

3 — Normalized frequency of chain reversal R (Tanaka-Scheraga, 1977) Feature 369

Cluster 2

1 — alpha-NH chemical shifts (Bundi-Wuthrich, 1979) Feature 16

2 — Helix initiation parameter at posision i,i+1,i+2 (Finkelstein et al., 1991) Feature 93

3 — Delta G values for the peptides extrapolated to 0 M urea (O'Neil-DeGrado, Feature 423

4 — Helix formation parameters (delta delta G) (O'Neil-DeGrado, 1990) Feature 424

5 — Alpha helix propensity of position 44 in T4 lysozyme (Blaber et al., 1993) Feature 437

6 — Linker propensity from non-helical (annotated by DSSP) dataset Feature 499

Cluster 3

1 — Hydrophobicity index (Argos et al., 1982) Feature 2

2 — Hydrophobicity factor (Goldsack-Chalifoux, 1973) Feature 108

3 — Hydrophobicity (Jones, 1975) Feature 132

4 — Transfer free energy (Simon, 1976), Cited by Charton-Charton (1982) Feature 358

5 — Side-chain contribution to protein stability (kJ/mol) (Takano-Yutani, 2001) Feature 444

Cluster 4

1 — Transfer free energy to surface (Bull-Breese, 1974) Feature 14

2 — Retention coefficient in NaClO4 (Meek-Rossetti, 1981) Feature 180

3 — Retention coefficient in NaH2PO4 (Meek-Rossetti, 1981) Feature 181

4 — Transfer energy, organic solvent/water (Nozaki-Tanford, 1971) Feature 212

5 — HPLC parameter (Parker et al., 1986) Feature 239

6 — Partition coefficient (Pliska et al., 1981) Feature 240

7 — Transfer free energy from oct to wat (Radzicka-Wolfenden, 1988) Feature 315

8 — Optimal matching hydrophobicity (Sweet-Eisenberg, 1983) Feature 365

9 — Bitterness (Venanzi, 1984) Feature 380

10 — Principal property value z1 (Wold et al., 1987) Feature 390

11 — Retention coefficient at pH 2 (Guo et al., 1986) Feature 483

12 — SWEIG index (Cornette et al., 1987) Feature 518

Cluster 5

1 — Conformational parameter of inner helix (Beghin-Dirkx, 1975) Feature 5

2 — Weights for alpha-helix at the window position of -2 (Qian-Sejnowski, 1988) Feature 262

Cluster 6

1 — Relative preference value at Mid (Richardson-Richardson, 1988) Feature 330

2 — Normalized positional residue frequency at helix termini N3 (Aurora-Rose, Feature 410

3 — Normalized positional residue frequency at helix termini C4 (Aurora-Rose, Feature 414

Cluster 7

1 — Hydrophobicity (Prabhakaran, 1990) Feature 252

2 — Transfer free energy from chx to wat (Radzicka-Wolfenden, 1988) Feature 314

3 — Side chain hydropathy, uncorrected for solvation (Roseman, 1988) Feature 355

4 — Transfer free energy to lipophilic phase (von Heijne-Blomberg, 1979) Feature 381

5 — Hydrophilicity scale (Kuhn et al., 1995) Feature 482

6 — Weights from the IFH scale (Jacobs-White, 1989) Feature 513

7 — Hydrophobicity index (Engelman et al., 1986) Feature 530

Cluster 8

1 — Proportion of residues 951 — pK-N (Fasman, 1976) Feature 75

Cluster 10

1 — Average relative fractional occurrence in ER(i-1) (Rackovsky-Scheraga, 1982) Feature 311

2 — Normalized positional residue frequency at helix termini C"' (Aurora-Rose, Feature 421

Cluster 11

1 — Normalized frequency of extended structure (Burgess et al., 1974) Feature 20

Cluster 12

1 — Normalized positional residue frequency at helix termini N" (Aurora-Rose, Feature 405

Cluster 13

1 — Weights for alpha-helix at the window position of -4 (Qian-Sejnowski, 1988) Feature 260

Cluster 14

1 — Weights for coil at the window position of -4 (Qian-Sejnowski, 1988) Feature 286

Cluster 15

1 — Normalized positional residue frequency at helix termini N1 (Aurora-Rose, Feature 408

Cluster 16

1 — Normalized positional residue frequency at helix termini N2 (Aurora-Rose, Feature 409

Cluster 17

1 — Distribution of amino acid residues in the alpha-helices in thermophilic Feature 442

2 — Distribution of amino acid residues in the alpha-helices in mesophilic Feature 443

Cluster 18

1 — Distribution of amino acid residues in the 18 non-redundant families of Feature 440

2 — Distribution of amino acid residues in the 18 non-redundant families of Feature 441

3 — Entire chain composition of amino acids in intracellular proteins of Feature 469

4 — Entire chain composition of amino acids in intracellular proteins of Feature 470

Cluster 19

1 — Signal sequence helical potential (Argos et al., 1982) Feature 3

2 — Membrane-buried preference parameters (Argos et al., 1982) Feature 4

Cluster 20

1 — Turn propensity scale for transmembrane helices (Monne et al., 1999) Feature 436

2 — Averaged turn propensities in a transmembrane helix (Monne et al., 1999) Feature 453

3 — Knowledge-based membrane-propensity scale from 1D_Helix in MPtopo databases Feature 489

4 — Knowledge-based membrane-propensity scale from 3D_Helix in MPtopo databases Feature 490

Cluster 21

1 — Normalized frequency of N-terminal beta-sheet (Chou-Fasman, 1978b) Feature 45

2 — Normalized frequency of bata-structure (Nagano, 1973) Feature 187

Cluster 22

1 — Beta-strand indices for beta-proteins (Geisow-Roberts, 1980) Feature 102

2 — Aperiodic indices for beta-proteins (Geisow-Roberts, 1980) Feature 106

3 — Normalized frequency of beta-sheet in all-beta class (Palau et al., 1981) Feature 232

Cluster 23

1 — Relative preference value at N4 (Richardson-Richardson, 1988) Feature 328

Cluster 24

1 — van der Waals parameter R0 (Levitt, 1976) Feature 158

2 — Mean area buried on transfer (Rose et al., 1985) Feature 353

3 — The relative stability scale extracted from mutation experiments (Zhou-Zhou, Feature 501

Cluster 25

1 — Normalized frequency of beta-sheet (Chou-Fasman, 1978b) Feature 39

2 — Beta-strand indices (Geisow-Roberts, 1980) Feature 101

3 — Beta-strand indices for alpha/beta-proteins (Geisow-Roberts, 1980) Feature 103

4 — Average relative probability of beta-sheet (Kanehisa-Tsong, 1980) Feature 139

5 — Average relative probability of inner beta-sheet (Kanehisa-Tsong, 1980) Feature 141

6 — Normalized frequency of beta-sheet, unweighted (Levitt, 1978) Feature 164

7 — Normalized frequency of beta-sheet from LG (Palau et al., 1981) Feature 225

8 — Normalized frequency of beta-sheet from CF (Palau et al., 1981) Feature 226

9 — Beta-coil equilibrium constant (Ptitsyn-Finkelstein, 1983) Feature 257

10 — Information measure for extended (Robson-Suzuki, 1976) Feature 343

11 — Information measure for pleated-sheet (Robson-Suzuki, 1976) Feature 344

Cluster 26
1 — Normalized frequency of beta-sheet, with weights (Levitt, 1978) Feature 161

2 — Conformational preference for all beta-strands (Lifson-Sander, 1979) Feature 167

3 — Conformational preference for antiparallel beta-strands (Lifson-Sander, 1979) Feature 169

4 — Normalized frequency of beta-sheet in alpha/beta class (Palau et al., 1981) Feature 234

5 — Relative frequency in beta-sheet (Prabhakaran, 1990) Feature 254

6 — Weights for beta-sheet at the window position of -1 (Qian-Sejnowski, 1988) Feature 276

7 — Weights for beta-sheet at the window position of 0 (Qian-Sejnowski, 1988) Feature 277

8 — Weights for beta-sheet at the window position of 1 (Qian-Sejnowski, 1988) Feature 278

9 — Free energy in beta-strand conformation (Munoz-Serrano, 1994) Feature 431

Cluster 27
1 — Relative preference value at N2 (Richardson-Richardson, 1988) Feature 326

Cluster 28
1 — Negative charge (Fauchere et al., 1988) Feature 89

2 — Relative preference value at N3 (Richardson-Richardson, 1988) Feature 327

Cluster 29
1 — Normalized frequency of beta-sheet in alpha+beta class (Palau et al., 1981) Feature 233

Cluster 30
1 — Average relative fractional occurrence in E0(i) (Rackovsky-Scheraga, 1982) Feature 304

Cluster 31
1 — Normalized frequency of zeta L (Maxfield-Scheraga, 1976) Feature 175

2 — Normalized frequency of coil (Tanaka-Scheraga, 1977) Feature 374

Cluster 32
1 — STERIMOL minimum width of the side chain (Fauchere et al., 1988) Feature 82

2 — N.m.r. chemical shift of alpha-carbon (Fauchere et al., 1988) Feature 84

Cluster 33
1 — Helix-coil equilibrium constant (Ptitsyn-Finkelstein, 1983) Feature 256

2 — Weights for coil at the window position of -2 (Qian-Sejnowski, 1988) Feature 288

3 — Weights for coil at the window position of 1 (Qian-Sejnowski, 1988) Feature 291

4 — Weights for coil at the window position of 2 (Qian-Sejnowski, 1988) Feature 292

Cluster 34
1 — The Chou-Fasman parameter of the coil conformation (Charton-Charton, 1983) Feature 24

2 — Normalized frequency of beta-turn (Chou-Fasman, 1978a) Feature 37

3 — Normalized frequency of beta-turn (Chou-Fasman, 1978b) Feature 40

4 — Normalized frequency of N-terminal non beta region (Chou-Fasman, 1978b) Feature 47

5 — Normalized frequency of the 2nd and 3rd residues in turn (Chou-Fasman, 1978b) Feature 53

6 — Aperiodic indices (Geisow-Roberts, 1980) Feature 104

7 — Aperiodic indices for alpha/beta-proteins (Geisow-Roberts, 1980) Feature 107

8 — Normalized relative frequency of bend (Isogai et al., 1980) Feature 121

9 — Normalized frequency of reverse turn, with weights (Levitt, 1978) Feature 162

10 — Normalized frequency of reverse turn, unweighted (Levitt, 1978) Feature 165

11 — Normalized frequency of turn from CF (Palau et al., 1981) Feature 228

12 — Relative frequency in reverse-turn (Prabhakaran, 1990) Feature 255

13 — Weights for coil at the window position of -1 (Qian-Sejnowski, 1988) Feature 289

14 — Weights for coil at the window position of 0 (Qian-Sejnowski, 1988) Feature 290

15 — Normalized frequency of chain reversal (Tanaka-Scheraga, 1977) Feature 375

Cluster 35
1 — Weights for beta-sheet at the window position of -2 (Qian-Sejnowski, 1988) Feature 275

Cluster 36
1 — Relative preference value at C2 (Richardson-Richardson, 1988) Feature 334

Cluster 37
1 — Relative preference value at C1 (Richardson-Richardson, 1988) Feature 335

Cluster 38
1 — Information value for accessibility; average fraction 351 — Information value for accessibility; average fraction 231 — Partition coefficient (Garel et al., 1973) Feature 96

2 — van der Waals parameter epsilon (Levitt, 1976) Feature 159

3 — Optimized propensity to form reverse turn (Oobatake et al., 1985) Feature 219

4 — RF value in high salt chromatography (Weber-Lacey, 1978) Feature 383

Cluster 41
1 — Principal component III (Sneath, 1966) Feature 361

2 — Dependence of partition coefficient on ionic strength (Zaslavsky et al., Feature 397

3 — Free energies of transfer of AcWl-X-LL peptides from bilayer interface to Feature 434

Cluster 42
1 — Normalized frequency of turn in all-alpha class (Palau et al., 1981) Feature 235

Cluster 43
1 — Normalized frequency of turn from LG (Palau et al., 1981) Feature 227

2 — Normalized frequency of turn in alpha/beta class (Palau et al., 1981) Feature 238

Cluster 44
1 — Average relative fractional occurrence in EL(i-1) (Rackovsky-Scheraga, 1982) Feature 309

2 — Free energy in beta-strand region (Munoz-Serrano, 1994) Feature 432

3 — Free energy in beta-strand region (Munoz-Serrano, 1994) Feature 433

Cluster 45
1 — Weights for alpha-helix at the window position of 1 (Qian-Sejnowski, 1988) Feature 265

2 — Value of theta(i-1) (Rackovsky-Scheraga, 1982) Feature 313

3 — Information measure for middle helix (Robson-Suzuki, 1976) Feature 341

4 — Free energy in alpha-helical conformation (Munoz-Serrano, 1994) Feature 429

5 — Free energy in alpha-helical region (Munoz-Serrano, 1994) Feature 430

Cluster 46
1 — Positive charge (Fauchere et al., 1988) Feature 88

Cluster 47
1 — Absolute entropy (Hutchens, 1970) Feature 117

2 — Entropy of formation (Hutchens, 1970) Feature 118

Cluster 48
1 — Weights for alpha-helix at the window position of -6 (Qian-Sejnowski, 1988) Feature 258

Cluster 49
1 — Hydrophobicity coefficient in RP-HPLC, C4 with 0.11 — A parameter of charge transfer donor capability (Charton-Charton, 1983) Feature 31

Cluster 51
1 — Amino acid composition (Dayhoff et al., 1978a) Feature 64

2 — Amino acid distribution (Jukes et al., 1975) Feature 136

3 — Sequence frequency (Jungck, 1978) Feature 137

4 — Refractivity (McMeekin et al., 1964), Cited by Jones (1975) Feature 177

5 — Composition of amino acids in extracellular proteins (percent) (Cedano et Feature 456

6 — Entire chain composition of amino

acids in extracellular proteins of Feature 471

Cluster 52
1 — Weights for beta-sheet at the window position of 3 (Qian-Sejnowski, 1988) Feature 280

Cluster 53
1 — Average relative fractional occurrence in AR(i) (Rackovsky-Scheraga, 1982) Feature 301

Cluster 54
1 — Weights for coil at the window position of -3 (Qian-Sejnowski, 1988) Feature 287

Cluster 55
1 — Average relative fractional occurrence in EL(i) (Rackovsky-Scheraga, 1982) Feature 303

Cluster 56
1 — Heat capacity (Hutchens, 1970) Feature 116

Cluster 57
1 — Normalized frequency of turn in all-beta class (Palau et al., 1981) Feature 236

Cluster 58
1 — Hydrophobicity coefficient in RP-HPLC, C8 with 0.11 — Retention coefficient in TFA (Browne et al., 1982) Feature 12
2 — Retention coefficient in HFBA (Browne et al., 1982) Feature 13
3 — Retention coefficient in HPLC, pH7.4 (Meek, 1980) Feature 178
4 — Retention coefficient in HPLC, pH2.1 (Meek, 1980) Feature 179
5 — RF rank (Zimmerman et al., 1968) Feature 402

Cluster 60
1 — Normalized frequency of alpha-helix (Burgess et al., 1974) Feature 19
2 — Normalized frequency of alpha-helix (Chou-Fasman, 1978b) Feature 38
3 — Normalized frequency of middle helix (Crawford et al., 1973) Feature 60
4 — Alpha-helix indices (Geisow-Roberts, 1980) Feature 97
5 — Normalized relative frequency of alpha-helix (Isogai et al., 1980) Feature 119
6 — Average relative probability of helix (Kanehisa-Tsong, 1980) Feature 138
7 — Normalized frequency of alpha-helix, with weights (Levitt, 1978) Feature 160
8 — Normalized frequency of alpha-helix, unweighted (Levitt, 1978) Feature 163
9 — Normalized frequency of alpha-helix (Maxfield-Scheraga, 1976) Feature 171
10 — Normalized frequency of alpha-helix (Nagano, 1973) Feature 186
11 — Normalized frequency of coil (Nagano, 1973) Feature 188
12 — Normalized frequency of alpha-helix from LG (Palau et al., 1981) Feature 223
13 — Normalized frequency of alpha-helix from CF (Palau et al., 1981) Feature 224
14 — Relative frequency in alpha-helix (Prabhakaran, 1990) Feature 253
15 — Average relative fractional

occurrence in AR(i-1) (Rackovsky-Scheraga, 1982) Feature 307
16 — Information measure for alpha-helix (Robson-Suzuki, 1976) Feature 339
17 — Normalized frequency of alpha-helix (Tanaka-Scheraga, 1977) Feature 366

Cluster 61
1 — Alpha-helix indices for alpha/beta-proteins (Geisow-Roberts, 1980) Feature 100
2 — Normalized frequency of alpha-helix in alpha/beta class (Palau et al., 1981) Feature 231
3 — Weights for alpha-helix at the window position of -3 (Qian-Sejnowski, 1988) Feature 261

Cluster 62
1 — Flexibility parameter for one rigid neighbor (Karplus-Schulz, 1985) Feature 143
2 — Side chain interaction parameter (Krigbaum-Rubin, 1971) Feature 147
3 — Surrounding hydrophobicity in turn (Ponnuswamy et al., 1980) Feature 246

Cluster 63
1 — Normalized frequency of N-terminal non helical region (Chou-Fasman, 1978b) Feature 43

Cluster 64
1 — Weights for beta-sheet at the window position of -6 (Qian-Sejnowski, 1988) Feature 271
2 — Weights for beta-sheet at the window position of -5 (Qian-Sejnowski, 1988) Feature 272

Cluster 65
1 — Helix termination parameter at posision j+1 (Finkelstein et al., 1991) Feature 95

Cluster 66
1 — Normalized frequency of N-terminal helix (Chou-Fasman, 1978b) Feature 41
2 — Information measure for N-terminal helix (Robson-Suzuki, 1976) Feature 340

Cluster 67
1 — Normalized positional residue frequency at helix termini N" (Aurora-Rose, Feature 404

Cluster 68
1 — Alpha-helix indices for alpha-proteins (Geisow-Roberts, 1980) Feature 98
2 — Aperiodic indices for alpha-proteins (Geisow-Roberts, 1980) Feature 105
3 — Normalized frequency of alpha-helix in all-alpha class (Palau et al., 1981) Feature 229

Cluster 69
1 — Relative frequency of occurrence (Jones et al., 1992) Feature 134
2 — AA composition of total proteins (Nakashima et al., 1990) Feature 189
3 — AA composition of EXT of single-spanning proteins (Nakashima-Nishikawa, 1992) Feature 204
4 — AA composition of EXT2 of single-spanning proteins (Nakashima-Nishikawa, Feature 205
5 — AA composition of EXT of multi-spanning proteins (Nakashima-Nishikawa, 1992) Feature 208

6 — Composition of amino acids in anchored proteins (percent) (Cedano et al., Feature 457
7 — Composition of amino acids in intracellular proteins (percent) (Cedano et Feature 459
8 — Entire chain compositino of amino acids in nuclear proteins (percent) Feature 472

Cluster 70
1 — SD of AA composition of total proteins (Nakashima et al., 1990) Feature 190
2 — AA composition of CYT of single-spanning proteins (Nakashima-Nishikawa, 1992) Feature 202
3 — AA composition of CYT2 of single-spanning proteins (Nakashima-Nishikawa, Feature 203
4 — AA composition of CYT of multi-spanning proteins (Nakashima-Nishikawa, 1992) Feature 207
5 — Composition of amino acids in nuclear proteins (percent) (Cedano et al., Feature 460

Cluster 71
1 — Relative preference value at C5 (Richardson-Richardson, 1988) Feature 331

Cluster 72
1 — Normalized positional residue frequency at helix termini C" (Aurora-Rose, Feature 420

Cluster 73
1 — Zimm-Bragg parameter sigma x 1.0E4 (Sueki et al., 1984) Feature 364

Cluster 74
1 — Optimized side chain interaction parameter (Oobatake et al., 1985) Feature 222
2 — Weights for coil at the window position of -6 (Qian-Sejnowski, 1988) Feature 284

Cluster 75
1 — Helix-coil equilibrium constant (Finkelstein-Ptitsyn, 1977) Feature 91
2 — Average relative probability of inner helix (Kanehisa-Tsong, 1980) Feature 140
3 — Weights for alpha-helix at the window position of -1 (Qian-Sejnowski, 1988) Feature 263
4 — Weights for alpha-helix at the window position of 0 (Qian-Sejnowski, 1988) Feature 264
5 — Zimm-Bragg parameter s at 20 C (Sueki et al., 1984) Feature 363
6 — Normalized positional residue frequency at helix termini N4 (Aurora-Rose, Feature 411
7 — Normalized positional residue frequency at helix termini C3 (Aurora-Rose, Feature 415
8 — Normalized positional residue frequency at helix termini C2 (Aurora-Rose, Feature 416

Cluster 76
1 — Information measure for N-terminal turn (Robson-Suzuki, 1976) Feature 347

Cluster 77
1 — AA composition of mt-proteins (Nakashima et al., 1990) Feature 191
2 — AA composition of mt-proteins from animal (Nakashima et al., 1990) Feature 193
3 — AA composition of mt-proteins

from fungi and plant (Nakashima et al., 1990) Feature 195
4 — Transmembrane regions of mt-proteins (Nakashima et al., 1990) Feature 200

Cluster 78
1 — AA composition of membrane proteins (Nakashima et al., 1990) Feature 197
2 — Transmembrane regions of non-mt-proteins (Nakashima et al., 1990) Feature 199
3 — AA composition of MEM of single-spanning proteins (Nakashima-Nishikawa, 1992) Feature 206
4 — AA composition of MEM of multi-spanning proteins (Nakashima-Nishikawa, 1992) Feature 209
5 — Composition of amino acids in membrane proteins (percent) (Cedano et al., Feature 458
6 — Interior composition of amino acids in intracellular proteins of thermophiles Feature 465
7 — Interior composition of amino acids in intracellular proteins of mesophiles Feature 466
8 — Interior composition of amino acids in extracellular proteins of mesophiles Feature 467
9 — Interior composition of amino acids in nuclear proteins (percent) Feature 468

Cluster 79
1 — Weights for alpha-helix at the window position of 2 (Qian-Sejnowski, 1988) Feature 266
2 — Weights for alpha-helix at the window position of 3 (Qian-Sejnowski, 1988) Feature 267
3 — Weights for alpha-helix at the window position of 4 (Qian-Sejnowski, 1988) Feature 268
4 — Weights for coil at the window position of 3 (Qian-Sejnowski, 1988) Feature 293

Cluster 80
1 — Normalized frequency of C-terminal helix (Chou-Fasman, 1978b) Feature 42
2 — Information measure for C-terminal helix (Robson-Suzuki, 1976) Feature 342

Cluster 81
1 — Relative preference value at C-cap (Richardson-Richardson, 1988) Feature 336
2 — Normalized frequency of left-handed helix (Tanaka-Scheraga, 1977) Feature 372
3 — Normalized positional residue frequency at helix termini C' (Aurora-Rose, Feature 419

Cluster 82
1 — Normalized relative frequency of coil (Isogai et al., 1980) Feature 126
2 — Normalized frequency of left-handed alpha-helix (Maxfield-Scheraga, 1976) Feature 174
3 — Average relative fractional occurrence in ER(i) (Rackovsky-Scheraga, 1982) Feature 305
4 — Average relative fractional occurrence in AL(i-1) (Rackovsky-Scheraga, 1982) Feature 308

Cluster 83
1 — Beta-sheet propensity derived

from designed sequences (Koehl-Levitt, 1999) Feature 455

Cluster 84
1 — Normalized frequency of beta-sheet (Crawford et al., 1973) Feature 61
2 — Conformational preference for parallel beta-strands (Lifson-Sander, 1979) Feature 168

Cluster 85
1 — Principal property value z3 (Wold et al., 1987) Feature 392

Cluster 86
1 — Amphiphilicity index (Mitaku et al., 2002) Feature 473

Cluster 87
1 — Residue volume (Bigelow, 1967) Feature 9
2 — Polarizability parameter (Charton-Charton, 1982) Feature 22
3 — Average volume of buried residue (Chothia, 1975) Feature 32
4 — Molecular weight (Fasman, 1976) Feature 72
5 — Normalized van der Waals volume (Fauchere et al., 1988) Feature 80
6 — Residue volume (Goldsack-Chalifoux, 1973) Feature 109
7 — Volume (Grantham, 1974) Feature 112
8 — Side chain volume (Krigbaum-Komoriya, 1979) Feature 150
9 — Average non-bonded energy per residue (Oobatake-Ooi, 1977) Feature 216
10 — Short and medium range non-bonded energy per residue (Oobatake-Ooi, 1977) Feature 217
11 — Volumes including the crystallographic waters using the ProtOr (Tsai et al., Feature 474
12 — Volumes not including the crystallographic waters using the ProtOr (Tsai et Feature 475
13 — Mean volumes of residues buried in protein interiors (Harpaz et al., 1994) Feature 504
14 — Average volumes of residues (Pontius et al., 1996) Feature 505
15 — Number of vertices (order of the graph) (Karkbara-Knisley, 2016) Feature 532
16 — Number of edges (size of the graph) (Karkbara-Knisley, 2016) Feature 533
17 — Total weighted degree of the graph (obtained by adding all the weights of Feature 534
18 — Weighted domination number (Karkbara-Knisley, 2016) Feature 535
19 — Average eccentricity (Karkbara-Knisley, 2016) Feature 536
20 — Radius (minimum eccentricity) (Karkbara-Knisley, 2016) Feature 537
21 — Diameter (maximum eccentricity) (Karkbara-Knisley, 2016) Feature 538
22 — Weighted domination number using the atomic number (Karkbara-Knisley, 2016) Feature 544
23 — Total weighted atomic number of the graph (obtained by summing all the atomic Feature 548

Cluster 88
1 — Second smallest eigenvalue of the Laplacian matrix of the graph Feature 543

Cluster 89
1 — Weighted second smallest eigenvalue of the weighted Laplacian matrix Feature 553

Cluster 90
1 — Relative preference value at C" (Richardson-Richardson, 1988) Feature 338

Cluster 91
1 — Normalized frequency of isolated helix (Tanaka-Scheraga, 1977) Feature 367

Cluster 92
1 — Conformational parameter of beta-structure (Beghin-Dirkx, 1975) Feature 6
2 — Average reduced distance for C-alpha (Meirovitch et al., 1980) Feature 182
3 — Effective partition energy (Miyazawa-Jernigan, 1985) Feature 185
4 — Accessibility reduction ratio (Ponnuswamy et al., 1980) Feature 247
5 — Average reduced distance for C-alpha (Rackovsky-Scheraga, 1977) Feature 297
6 — Relative partition energies derived by the Bethe approximation Feature 525
7 — Optimized relative partition energies - method A (Miyazawa-Jernigan, 1999) Feature 526

Cluster 93
1 — Normalized hydrophobicity scales for alpha-proteins (Cid et al., 1992) Feature 54
2 — Normalized hydrophobicity scales for beta-proteins (Cid et al., 1992) Feature 55
3 — Normalized hydrophobicity scales for alpha+beta-proteins (Cid et al., 1992) Feature 56
4 — Normalized hydrophobicity scales for alpha/beta-proteins (Cid et al., 1992) Feature 57
5 — Normalized average hydrophobicity scales (Cid et al., 1992) Feature 58
6 — Flexibility parameter for no rigid neighbors (Karplus-Schulz, 1985) Feature 142
7 — Hydration free energy (Robson-Osguthorpe, 1979) Feature 352
8 — Normalized flexibility parameters (B-values) for each residue surrounded by Feature 426
9 — p-Values of mesophilic proteins based on the distributions of B values Feature 438
10 — Surface composition of amino acids in extracellular proteins of mesophiles Feature 463
11 — Interactivity scale obtained by maximizing the mean of correlation Feature 486
12 — The stability scale from the knowledge-based atom-atom potential (Zhou-Zhou, Feature 500
13 — Buriability (Zhou-Zhou, 2004) Feature 502

Cluster 94
1 — Relative preference value at C3 (Richardson-Richardson, 1988) Feature 333

Cluster 95
1 — Value of theta(i) (Rackovsky-Scheraga, 1982) Feature 312

4

2 — Alpha-helix propensity derived from designed sequences (Koehl-Levitt, 1999) Feature 454

Cluster 96
1 — Solvation free energy (Eisenberg-McLachlan, 1986) Feature 69
2 — Hydrophobic parameter pi (Fauchere-Pliska, 1983) Feature 77
3 — Polarity (Grantham, 1974) Feature 111
4 — Hydrophilicity value (Hopp-Woods, 1981) Feature 115
5 — Hydrophobic parameter (Levitt, 1976) Feature 153
6 — Polar requirement (Woese, 1973) Feature 388
7 — Hydrophobicity-related index (Kidera et al., 1985) Feature 509
8 — Scaled side chain hydrophobicity values (Black-Mould, 1991) Feature 515

Cluster 97
1 — Normalized composition of membrane proteins (Nakashima et al., 1990) Feature 198
2 — Hydrophobicity coefficient in RP-HPLC, C18 with 0.11 — Frequency of occurrence in beta-bends (Lewis et al., 1971) Feature 166

Cluster 99
1 — Normalized frequency of C-terminal non beta region (Chou-Fasman, 1978b) Feature 48
2 — Normalized frequency of turn in alpha+beta class (Palau et al., 1981) Feature 237
3 — Information measure for coil (Robson-Suzuki, 1976) Feature 350

Cluster 100
1 — Weights for beta-sheet at the window position of 2 (Qian-Sejnowski, 1988) Feature 279
2 — Normalized flexibility parameters (B-values) for each residue surrounded by Feature 428

Cluster 101
1 — Residue accessible surface area in folded protein (Chothia, 1976) Feature 34
2 — Proportion of residues 1001 — Normalized frequency of alpha-helix in alpha+beta class (Palau et al., 1981) Feature 230

Cluster 103
1 — Weights for alpha-helix at the window position of -5 (Qian-Sejnowski, 1988) Feature 259

Cluster 104
1 — A parameter of charge transfer capability (Charton-Charton, 1983) Feature 30

Cluster 105
1 — Normalized relative frequency of extended structure (Isogai et al., 1980) Feature 120
2 — Normalized frequency of extended structure (Maxfield-Scheraga, 1976) Feature 172
3 — Average relative fractional occurrence in E0(i-1) (Rackovsky-Scheraga, 1982) Feature 310
4 — Normalized frequency of extended structure (Tanaka-Scheraga, 1977) Feature 368

Cluster 106
1 — Frequency of the 3rd residue in turn (Chou-Fasman, 1978b) Feature 51
2 — Normalized relative frequency of bend S (Isogai et al., 1980) Feature 123
3 — Normalized frequency of chain reversal S (Tanaka-Scheraga, 1977) Feature 370

Cluster 107
1 — Frequency of the 1st residue in turn (Chou-Fasman, 1978b) Feature 49
2 — Relative preference value at N-cap (Richardson-Richardson, 1988) Feature 324

Cluster 108
1 — Relative preference value at C' (Richardson-Richardson, 1988) Feature 337

Cluster 109
1 — Normalized flexibility parameters (B-values) for each residue surrounded by Feature 427
2 — Hydropathy scale based on self-information values in the two-state model (251 — Weighted maximum eigenvalue based on the atomic numbers Feature 550
2 — Weighted average eigenvalue based on the atomic numbers Feature 552

Cluster 111
1 — Size (Dawson, 1972) Feature 63
2 — Average weighted degree (total degree, divided by the number of vertices) Feature 539
3 — Maximum eigenvalue of the weighted Laplacian matrix of the graph Feature 540
4 — Average eigenvalue of the Laplacian matrix of the the graph Feature 542

Cluster 112
1 — Surrounding hydrophobicity in beta-sheet (Ponnuswamy et al., 1980) Feature 245

Cluster 113
1 — The number of atoms in the side chain labelled 3+1 (Charton-Charton, 1983) Feature 28
2 — The number of bonds in the longest chain (Charton-Charton, 1983) Feature 29
3 — Residue accessible surface area in tripeptide (Chothia, 1976) Feature 33
4 — STERIMOL length of the side chain (Fauchere et al., 1988) Feature 81
5 — Distance between C-alpha and centroid of side chain (Levitt, 1976) Feature 154
6 — Radius of gyration of side chain (Levitt, 1976) Feature 157
7 — Transfer free energy from vap to chx (Radzicka-Wolfenden, 1988) Feature 316
8 — Accessible surface area (Radzicka-Wolfenden, 1988) Feature 319
9 — Average weighted eccentricity based on the atomic number Feature 545
10 — Weighted diameter based on the atomic number (maximum eccentricity) Feature 547

Cluster 114
1 — STERIMOL maximum width of the side chain (Fauchere et al., 1988) Feature 83

2 — Short and medium range non-bonded energy per atom (Oobatake-Ooi, 1977) Feature 214
3 — Principal property value z2 (Wold et al., 1987) Feature 391
4 — Weighted radius based on the atomic number (minimum eccentricity) Feature 546

Cluster 115
1 — Free energy of solution in water, kcal/mole (Charton-Charton, 1982) Feature 23

Cluster 116
1 — Linker propensity index (Suyama-Ohara, 2003) Feature 488

Cluster 117
1 — Normalized positional residue frequency at helix termini N4'(Aurora-Rose, Feature 403

Cluster 118
1 — Normalized positional residue frequency at helix termini C4' (Aurora-Rose, Feature 422

Cluster 119
1 — Atom-based hydrophobic moment (Eisenberg-McLachlan, 1986) Feature 70
2 — Activation Gibbs energy of unfolding, pH7.0 (Yutani et al., 1987) Feature 395
3 — Activation Gibbs energy of unfolding, pH9.0 (Yutani et al., 1987) Feature 396

Cluster 120
1 — Consensus normalized hydrophobicity scale (Eisenberg, 1984) Feature 68
2 — Number of hydrogen bond donors (Fauchere et al., 1988) Feature 86
3 — Transfer free energy from chx to oct (Radzicka-Wolfenden, 1988) Feature 317
4 — Transfer free energy from vap to oct (Radzicka-Wolfenden, 1988) Feature 318
5 — Energy transfer from out to in(951 — The Kerr-constant increments (Khanarian-Moore, 1980) Feature 145

Cluster 122
1 — Average weighted atomic number or degree based on atomic number in the graph Feature 549

Cluster 123
1 — Optimized average non-bonded energy per atom (Oobatake et al., 1985) Feature 221

Cluster 124
1 — Hydrophobicity coefficient in RP-HPLC, C18 with 0.11 — Weights for coil at the window position of -5 (Qian-Sejnowski, 1988) Feature 285

Cluster 126
1 — Free energy change of epsilon(i) to epsilon(ex) (Wertz-Scheraga, 1978) Feature 385

Cluster 127
1 — Conformational parameter of beta-turn (Beghin-Dirkx, 1975) Feature 7
2 — Normalized frequency of turn (Crawford et al., 1973) Feature 62
3 — Information measure for turn (Robson-Suzuki, 1976) Feature 346
4 — Information measure for middle turn (Robson-Suzuki, 1976) Feature 348

5

5 — Information measure for loop (Robson-Suzuki, 1976) Feature 351

Cluster 128
1 — Normalized positional residue frequency at helix termini Cc (Aurora-Rose, Feature 418

Cluster 129
1 — Ratio of average and computed composition (Nakashima et al., 1990) Feature 201

Cluster 130
1 — pK-C (Fasman, 1976) Feature 76
2 — pK (-COOH) (Jones, 1975) Feature 133

Cluster 131
1 — Unfolding Gibbs energy in water, pH7.0 (Yutani et al., 1987) Feature 393
2 — Unfolding Gibbs energy in water, pH9.0 (Yutani et al., 1987) Feature 394

Cluster 132
1 — A parameter defined from the residuals obtained from the best correlation of Feature 25

Cluster 133
1 — Flexibility parameter for two rigid neighbors (Karplus-Schulz, 1985) Feature 144

Cluster 134
1 — Weights for beta-sheet at the window position of -4 (Qian-Sejnowski, 1988) Feature 273
2 — Weights for beta-sheet at the window position of 6 (Qian-Sejnowski, 1988) Feature 283

Cluster 135
1 — Weights for beta-sheet at the window position of 4 (Qian-Sejnowski, 1988) Feature 281
2 — Weights for beta-sheet at the window position of 5 (Qian-Sejnowski, 1988) Feature 282

Cluster 136
1 — Relative preference value at N1 (Richardson-Richardson, 1988) Feature 325

Cluster 137
1 — Loss of Side chain hydropathy by helix formation (Roseman, 1988) Feature 357

Cluster 138
1 — Average flexibility indices (Bhaskaran-Ponnuswamy, 1988) Feature 8
2 — Relative preference value at C4 (Richardson-Richardson, 1988) Feature 332

Cluster 139
1 — Average relative fractional occurrence in AL(i) (Rackovsky-Scheraga, 1982) Feature 302

Cluster 140
1 — Linker propensity from small dataset (linker length is less than six Feature 495

Cluster 141
1 — Relative population of conformational state C (Vasquez et al., 1983) Feature 377
2 — Relative population of conformational state E (Vasquez et al., 1983) Feature 378

Cluster 142
1 — Average relative fractional occurrence in A0(i-1) (Rackovsky-Scheraga, 1982) Feature 306

Cluster 143
1 — Normalized positional residue frequency at helix termini N5 (Aurora-Rose, Feature 412
2 — Normalized positional residue frequency at helix termini C5 (Aurora-Rose, Feature 413
3 — Normalized positional residue frequency at helix termini C1 (Aurora-Rose, Feature 417
4 — Linker propensity from helical (annotated by DSSP) dataset (George-Heringa, Feature 498

Cluster 144
1 — Average relative fractional occurrence in A0(i) (Rackovsky-Scheraga, 1982) Feature 300

Cluster 145
1 — Optimized transfer energy parameter (Oobatake et al., 1985) Feature 220
2 — Normalized frequency of chain reversal D (Tanaka-Scheraga, 1977) Feature 371

Cluster 146
1 — Hydration number (Hopfinger, 1971), Cited by Charton-Charton (1982) Feature 114
2 — Polarity (Zimmerman et al., 1968) Feature 400

Cluster 147
1 — Number of full nonbonding orbitals (Fauchere et al., 1988) Feature 87
2 — Principal component II (Sneath, 1966) Feature 360

Cluster 148
1 — Information measure for extended without H-bond (Robson-Suzuki, 1976) Feature 345

Cluster 149
1 — Weights for alpha-helix at the window position of 5 (Qian-Sejnowski, 1988) Feature 269
2 — Weights for alpha-helix at the window position of 6 (Qian-Sejnowski, 1988) Feature 270

Cluster 150
1 — Principal component IV (Sneath, 1966) Feature 362

Cluster 151
1 — Normalized relative frequency of helix end (Isogai et al., 1980) Feature 124
2 — Normalized frequency of alpha region (Maxfield-Scheraga, 1976) Feature 176

Cluster 152
1 — Linker propensity from 1-linker dataset (George-Heringa, 2003) Feature 492
2 — Linker propensity from long dataset (linker length is greater than 14 Feature 497

Cluster 153
1 — Apparent partial specific volume (Bull-Breese, 1974) Feature 15
2 — Partial specific volume (Cohn-Edsall, 1943) Feature 59
3 — Optimized beta-structure-coil equilibrium constant (Oobatake et al., 1985) Feature 218

Cluster 154
1 — The number of atoms in the side chain labelled 1+1 (Charton-Charton, 1983) Feature 26

Cluster 155

1 — Weighted minimum eigenvalue based on the atomic numbers Feature 551

Cluster 156
1 — Normalized frequency of zeta R (Maxfield-Scheraga, 1976) Feature 173

Cluster 157
1 — Relative population of conformational state A (Vasquez et al., 1983) Feature 376

Cluster 158
1 — Free energy change of alpha(Ri) to alpha(Rh) (Wertz-Scheraga, 1978) Feature 386

Cluster 159
1 — Surrounding hydrophobicity in alpha-helix (Ponnuswamy et al., 1980) Feature 244
2 — Free energy change of epsilon(i) to alpha(Rh) (Wertz-Scheraga, 1978) Feature 387

Cluster 160
1 — Relative preference value at N" (Richardson-Richardson, 1988) Feature 322
2 — Relative preference value at N' (Richardson-Richardson, 1988) Feature 323

Cluster 161
1 — Normalized frequency of C-terminal beta-sheet (Chou-Fasman, 1978b) Feature 46
2 — Propensity of amino acids within pi-helices (Fodje-Al-Karadaghi, 2002) Feature 445

Cluster 162
1 — Hydropathy scale based on self-information values in the two-state model (501 — Slope in regression analysis x 1.0E1 (Prabhakaran-Ponnuswamy, 1982) Feature 250
2 — Correlation coefficient in regression analysis (Prabhakaran-Ponnuswamy, 1982) Feature 251

Cluster 164
1 — Graph shape index (Fauchere et al., 1988) Feature 78
2 — Bulkiness (Zimmerman et al., 1968) Feature 399
3 — Minimum eigenvalue of the weighted Laplacian matrix of the graph Feature 541

Cluster 165
1 — Alpha-helix indices for beta-proteins (Geisow-Roberts, 1980) Feature 99

Cluster 166
1 — Weights for beta-sheet at the window position of -3 (Qian-Sejnowski, 1988) Feature 274

Cluster 167
1 — The number of atoms in the side chain labelled 2+1 (Charton-Charton, 1983) Feature 27

Cluster 168
1 — Principal component I (Sneath, 1966) Feature 359

Cluster 169
1 — Hydrostatic pressure asymmetry index, PAI (Di Giulio, 2005) Feature 506

Cluster 170
1 — Electron-ion interaction potential (Veljkovic et al., 1985) Feature 379
2 — Electron-ion interaction potential values (Cosic, 1994) Feature 476

Cluster 171

1 — alpha-CH chemical shifts (Andersen et al., 1992) Feature 1

2 — alpha-CH chemical shifts (Bundi-Wuthrich, 1979) Feature 17

Cluster 172

1 — Optical rotation (Fasman, 1976) Feature 74

2 — pK-a(RCOOH) (Fauchere et al., 1988) Feature 90

Cluster 173

1 — Normalized frequency of C-terminal non helical region (Chou-Fasman, 1978b) Feature 44

Cluster 174

1 — p-Values of thermophilic proteins based on the distributions of B values Feature 439

2 — Surface composition of amino acids in intracellular proteins of thermophiles Feature 461

3 — Surface composition of amino acids in intracellular proteins of mesophiles Feature 462

4 — Surface composition of amino acids in nuclear proteins (percent) Feature 464

Cluster 175

1 — Normalized composition of mt-proteins (Nakashima et al., 1990) Feature 192

2 — Normalized composition from animal (Nakashima et al., 1990) Feature 194

3 — Normalized composition from fungi and plant (Nakashima et al., 1990) Feature 196

Cluster 176

1 — Transfer free energy, CHP/water (Lawson et al., 1984) Feature 152

2 — Hydrophobicity (Zimmerman et al., 1968) Feature 398

Cluster 177

1 — Steric parameter (Charton, 1981) Feature 21

2 — Smoothed upsilon steric parameter (Fauchere et al., 1988) Feature 79

3 — Side chain angle theta(AAR) (Levitt, 1976) Feature 155

4 — Side chain torsion angle phi(AAAR) (Levitt, 1976) Feature 156

5 — Thermodynamic beta sheet propensity (Kim-Berg, 1993) Feature 435

Cluster 178

1 — Melting point (Fasman, 1976) Feature 73

Cluster 179

1 — Relative mutability (Dayhoff et al., 1978b) Feature 65

2 — Relative mutability (Jones et al., 1992) Feature 135

Cluster 180

1 — Relative preference value at N5 (Richardson-Richardson, 1988) Feature 329

Cluster 181

1 — Helix termination parameter at posision j-2,j-1,j (Finkelstein et al., 1991) Feature 94

2 — Net charge (Klein et al., 1984) Feature 146

3 — Isoelectric point (Zimmerman et al., 1968) Feature 401

Cluster 182

1 — Weights for coil at the window position of 4 (Qian-Sejnowski, 1988) Feature 294

2 — Weights for coil at the window position of 5 (Qian-Sejnowski, 1988) Feature 295

Cluster 183

1 — Normalized relative frequency of double bend (Isogai et al., 1980) Feature 125

Cluster 184

1 — Localized electrical effect (Fauchere et al., 1988) Feature 85

2 — Composition (Grantham, 1974) Feature 110

Cluster 185

1 — Frequency of the 4th residue in turn (Chou-Fasman, 1978b) Feature 52

2 — Information measure for C-terminal turn (Robson-Suzuki, 1976) Feature 349

Cluster 186

1 — Linker propensity from all dataset (George-Heringa, 2003) Feature 491

2 — Linker propensity from 2-linker dataset (George-Heringa, 2003) Feature 493

3 — Linker propensity from 3-linker dataset (George-Heringa, 2003) Feature 494

4 — Linker propensity from medium dataset (linker length is between six and 14 Feature 496

Cluster 187

1 — Weights for coil at the window position of 6 (Qian-Sejnowski, 1988) Feature 296

Cluster 188

1 — Normalized frequency of zeta R (Tanaka-Scheraga, 1977) Feature 373

Cluster 189

1 — Helix initiation parameter at posision i-1 (Finkelstein et al., 1991) Feature 92

2 — Normalized positional residue frequency at helix termini N'(Aurora-Rose, Feature 406

3 — Normalized positional residue frequency at helix termini Nc (Aurora-Rose, Feature 407

Cluster 190

1 — Intercept in regression analysis (Prabhakaran-Ponnuswamy, 1982) Feature 249

Cluster 191

1 — Spin-spin coupling constants 3JHalpha-NH (Bundi-Wuthrich, 1979) Feature 18