Supporting Information

Rackovsky 10.1073/pnas.0903433106

Table S1. The 10 property factors

1. Alpha-helix/bend preference
2. Side-chain size
3. Extended structure preference
4. Hydrophobicity
5. Double-bend preference
6. Amino acid composition
7. Flat extended preference
8. Occurrence in α region
9. pK
10. Surrounding hydrophobicity

The first 4 property factors correspond essentially to single amino acid physical properties. The remaining 6 factors are superpositions of several physical properties and for convenience are identified by the property making the greatest contribution to the factor.

Table S2. The 59 CAT groups

C	A	T	Number of sequences	CATH group name
1	20	5	39	Single alpha-helices involved in coiled coils or other helix–helix interfaces
1	50	10	31	Glycosyltransferase
1	20	58	38	Methane monooxygenase Hydroxylase, chain G, domain 1
1	10	238	67	Recoverin, domain 1
1	10	490	32	Globin-like
1	10	287	53	Helix hairpins
1	10	8	40	Helicase, ruva protein, domain 3
1	20	1,050	29	Glutathione S-transferase Yfyf (class Pi), chain A, domain 2
1	20	1,250	27	Growth hormone, chain A
1	10	472	20	Cyclin A, domain 1
1	10	760	33	Cytochrome Bc1 complex, chain D, domain 2
1	20	120	54	Four-helix bundle [hemerythrin (Met), subunit A]
1	10	10	150	Arc repressor mutant, subunit A
1	10	150	54	DNA polymerase, domain 1
1	25	40	67	Serine threonine protein phosphatase 5, tetratricopeptide repeat
1	10	510	28	Transferase (phosphotransferase), domain 1
2	40	50	125	OB fold (dihydrolipoamide acetyltransferase, E2P)
2	130	10	30	Methylamine dehydrogenase, chain H
2	40	30	36	Elongation factor Tu (Ef-tu), domain 3
2	120	10	20	Neuraminidase
2	40	128	36	Lipocalin
2	160	20	22	Pectate lyase C–like
2	30	29	38	PH domain–like
2	30	42	37	Pdz3 domain
2	40	70	30	Cathepsin D, subunit A, domain 1
2	80	10	36	Trefoil (acidic fibroblast growth factor, subunit A)
2	60	40	423	Immunoglobulin-like
2	60	120	209	Jelly rolls
2	40	10	95	Thrombin, subunit H
2	30	30	83	SH3-type barrels
3	90	190	22	Protein-tyrosine phosphatase; chain A
3	30	390	28	Enolase-like, domain 1
3	30	360	20	Dihydrodipicolinate Reductase, domain 2
3	10	50	20	Chitinase A, domain 3
3	40	630	47	Aminopeptidase
3	40	1,190	21	UDP-N-acetylmuramoyl-L-alanine:D-glutamate ligase
3	10	180	22	2,3-Dihydroxybiphenyl 1,2-dioxygenase, domain 1
3	50	50	46	FAD/NAD(P)-binding domain
3	30	1,330	22	60 s ribosomal protein L30, chain A
3	60	20	30	Glutamine phosphoribosylpyrophosphate, subunit 1, domain 1
3	90	550	21	Spore coat polysaccharide biosynthesis protein SpsA, chain A
3	30	230	24	Ribosomal protein S5, domain 2
3	10	450	47	Nuclear transport factor 2, chain A
3	40	30	71	Glutaredoxin
3	30	505	20	SHC adaptor protein
3	30	470	23	D-amino acid aminotransferase, chain A, domain 1
3	30	200	48	Phosphorylase kinase, domain 1
3	40	50	944	Rossmann fold
3	30	1,490	27	DNA ligase, domain 1
3	20	20	247	TIM barrel
3	30	70	193	Alpha-beta plaits
3	30	300	28	GMP synthetase, chain A, domain 3
3	30	420	52	Nucleotidyltransferase, domain 5
3	10	100	29	Mannose-binding protein A, chain A
3	40	640	40	Aspartate aminotransferase, domain 2
3	10	20	83	Ubiquitin-like (UB roll)
3	40	190	83	D-maltodextrin–binding protein, domain 2
3	90	1,150	42	Aspartate aminotransferase, domain 1
3	30	450	37	Beta-lactamase

Table S3. The separating hyperplanes

	H(1,3)	H(2,3)
w_0	−0.048	−0.021
w_1	0.025	0.118
w_2	−0.031	−0.014
w_3	−0.222	0.064
w_4	0.091	0.135
w_5	0.052	0.059
w_6	−0.116	−0.039
w_7	0.049	−0.119
w_8	−0.197	−0.125
w_9	−0.001	−0.128
w_{10}	−0.067	0.017

The hyperplane H(*i*,*j*) separates regions occupied predominantly by CAT groups with C values *i* and *j*. The hyperplanes have the general form $H(i,j) = w_0 + \sum_{n=1}^{10} w_n \langle f^{(n)} \rangle$

Table S4. The 60 CAT groups

C	A	T	Number of sequences	CATH group name
1	10	220	14	Annexin V, domain 1
1	10	565	16	Retinoid X receptor
1	10	418	13	Actin-binding protein, T-fimbrin, domain 1
1	10	260	14	434 repressor (amino-terminal domain)
1	20	140	11	Butyryl-CoA dehydrogenase, subunit A, domain 3
1	20	1,260	14	Ferritin
1	10	630	15	Cytochrome P450
1	10	1,040	12	N-(1-d-carboxylethyl)-l-norvaline dehydrogenase, domain 2
1	25	10	14	Leucine-rich repeat variant
1	10	20	18	Histone, subunit A
1	10	1,200	14	Nonribosomal peptide synthetase peptidyl carrier protein, chain A
1	10	530	11	Lysozyme
1	10	246	14	Serum albumin, chain A, domain 1
1	10	533	15	Death domain, Fas
2	60	20	12	Gamma-B crystallin, domain 1
2	30	110	10	Pnp oxidase, chain A
2	60	200	13	Tumor suppressor Smad4
2	30	40	12	Urease, subunit C, domain 1
2	10	70	16	Complement module, domain 1
2	70	98	16	Beta-galactosidase, chain A, domain 5
2	10	60	12	CD59
2	30	39	15	Alpha-1-antitrypsin, domain 1
2	160	10	12	UDP N-acetylglucosamine acyltransferase, domain 1
2	40	40	14	Barwin-like endoglucanases
2	10	90	13	Cystine knot cytokines, subunit B
3	10	310	10	Diaminopimelate epimerase, chain A, domain 1
3	60	15	13	Metallo-beta-lactamase, chain A
3	40	1,280	13	Alpha/beta knot
3	30	110	12	Translation initiation factor IF3
3	30	1,370	16	Ribosomal protein S8, chain A, domain 1
3	40	250	13	Oxidized rhodanese, domain 1
3	90	79	14	Nucleoside triphosphate pyrophosphohydrolase
3	30	310	13	TATA-binding protein
3	90	180	16	Quinine oxidoreductase, chain A, domain 1
3	30	365	13	Aldehyde oxidoreductase, domain 4
3	90	110	12	L-2-hydroxyisocaproate dehydrogenase, chain A, domain 2
3	30	830	11	Cytochrome Bc1 complex, chain A, domain 1
3	30	497	11	Antithrombin, chain I, domain 2
3	40	430	10	Dihydrofolate reductase, subunit A
3	40	710	19	Beta-lactamase
3	60	21	12	Purple acid phosphatase, chain A, domain 2
3	30	160	14	Double-stranded RNA-binding domain
3	10	28	13	Endonuclease I-crel
3	40	47	15	Peroxisomal thiolase, chain A, domain 1
3	40	192	10	Leucine dehydrogenase, chain A, domain 1
3	30	930	18	BirA bifunctional protein, domain 2
3	30	500	11	Murine class I major histocompatibility complex, H2-DB, chain A, domain 1
3	30	1,360	18	Gyrase A, domain 2
3	80	10	15	Leucine-rich repeat, LRR (right-handed beta-alpha superhelix)
3	40	140	11	Cytidine deaminase, domain 2
3	90	70	15	Cathepsin B, chain A
3	90	10	15	Cytochrome C3
3	90	226	16	2-enoyl-CoA hydratase, chain A, domain 1
3	40	20	12	Severin
3	30	40	18	Herpes virus 1
3	30	565	15	Heat shock protein 90
3	90	1,170	12	Aldehyde oxidoreductase, domain 3
3	90	25	10	UDP-galactose 4-epimerase, domain 1
3	10	129	19	Thiol ester dehydrase, chain A
3	40	390	16	Collagenase (C)