Features of MKLoc predictor

Table S1 list the features of the MKLoc predictor. The features description has been taken from the work of Briesemeister et al. [36].

Table S1. List of 30 features of MKLoc predictor

Feature No.	Biological interpretation	Detailed feature description
1	protein size	number of amino acids
2	ER retention signal	number of weak ER retention signals (number of K,D,E,L in the very C-terminus) times a factor that depends on the fact that no transmembrane helix and no glycolization signal is present. If a strong ER retention signal (KDEL, KRHQSADENQEL, DEL, EL) at C-terminus is present it is 100.
3	peroxisomal targeting signal (PTS)	weighted PTS sum = SKL*0.83+SKF*0.5+[SAGCN][RKH][LIVMAF]*0.25 (presented by Nakai et al.)
4	number of leucine and leucine clusters in the N-terminus	product of scores for leucine clusters (1.5 for LL, 2.5 for LLL, 5 for LLLL, 12 for LLLLL, 20 for LLLLLL) in the first 50 amino acids in the N-terminus
5	alternative secretory pathway sorting signal	sum of hydrophobicity of 10 amino acids before the N-terminal cleavage
6	length of longest very hydrophobic region	maximal length of very hydrophobic region where the iterative sum of hydrophobicity normed to mean zero drops below zero in at most two cases
7	putative mitochondrial sorting signal	weighted sum of typical amino acids for mitochondrial proteins in N-terminus
8	secretory pathway sorting	autocorrelation of every second hydrophobic amino acid within the first 20 amino acids in the N-terminus
9	mitochondrial targeting peptide	maximal autocorrelation of every sixth charged amino acid within the first 30 amino acids in the N-terminus
10	number of methionine in the N-terminus	number of methionine within the first 70 amino acids in the N-terminus
11	number of asparagine in the N-terminus	number of asparagine within the first 70 amino acids in the N-terminus
12	number of cysteine in the N-terminus	maximal pseudo amino acid count of cysteine within the first 120 amino acids in the N-terminus
13	number of lysine in the N-terminus	maximal normalized pseudo amino acid count of lysine within the first 120 amino acids in the N-terminus
14	number of tryptophane in the N-terminus	number of tryptophane within the first 120 amino acids in the N-terminus
15	hydrophobic N-terminus	pseudo amino acid count of very hydrophobic residues in a distance of two within the first 130 amino acids in the N-terminus
16	hydrophobic C-terminus	maximal normalized pseudo amino acid count of very hydrophobic residues in the last 40 amino acids in the C-terminus
17	negatively charged N-terminus	maximal pseudo amino acid count of negatively charged residues within the first 30 amino acids in the N-terminus
18	number of alpha-helix preferred residues AFGHKLMNR	minimal overall pseudo amino acid count of alpha-helix preferred residues
19	number of small amino acids in the N-terminus	number of small amino acids within the first 20 amino acids in the N-terminus
20	number of aromatic residues	maximal normalized overall pseudo amino acid count of

		aromatic residues
21	number of potentially hydroxylated residues in the C-terminus	pseudo amino acid count of potentially hydroxylated residues in a distance of three within the last 100 amino acids in the C-terminus
22	uncharged and hydrophobic protein	overall pseudo amino acid count of uncharged hydrophobic residues (I,L,V,M,F,Y,W,C,T,A,G) in a distant of two
23	typical plasma membrane prosite p.	cluster of typical plasma membrane prosite patterns
24	typical nuclear prosite pattern	cluster of typical nuclear prosite patterns
25	GO:0005783 (endoplasmic reticulum)	
26	GO:0005739 (mitochondrion)	
27	GO:0005576 (extracellular region)	
28	GO:0042025 (host cell nucleus)	
29	GO:0005778 (peroxisomal membrane)	
30	GO term cluster	