

Feature	Description	Number
$C(r)$	Single-residue composition (occurrence over sequence length); $r = A, C, D, E, F, G, H, [I], K, L, M, N, P, Q, R, S, [T], V, W, [Y]$	20
$C(c)$	Combined amino acid compositions; $c = [KR], NQ, [DE], ST, LM, [FWY], HKR, AVILM, [DENQ], GAVL, SCTM$	11
$C(a)$	Predicted secondary structure composition; $a = [\alpha], \beta, [coil]$	3
[Signal]	Presence of signal sequence	1
[Length]	Amino acid sequence length	1
[CPLX(x)]	Number of amino acids in low-complexity regions; $x = s$ (short), l (long)	2
[CPLXn(x)]	Normalized low-complexity value (CPLX over sequence length); $x = s$ (short), l (long)	2
[Hphobe]	Minimum GES hydrophobicity score calculated over all amino acids in a 20-residue sequence window	1
HP-AA	Number of amino acids within a hydrophobic stretch below a threshold of -1.0 kcal/mol	1
Total		42