Description of Features

In this section, we have elaborate and compared the features calculated by Pfeature and other available resources. Pfeature is able to calculate more than 70,000 composition features from the primary sequence of protein or peptide. In the table 3, we have described the group and type of features, kinds of sub-sequences, their dimension vectors, and methods which support the respective features.

Table: Brief description of features calculated by Pfeature

Type of Features	Description	Features	Dimension Vectors	Supported By
COMPOSITION: SIMPLE				
AAC	Amino acid Composition	Whole	20	{a,b,c,d,e}
		N-Terminal	20	{a}
		C-Terminal	20	{a}
		Rest	20	{a}
		Split	20*N	{a}
DPC	Dipeptide Composition	Whole	400	{a,b,c,d,e}
		N-Terminal	400	{a}
		C-Terminal	400	{a}
		Rest	400	{a}
		Split	400*N	{a}
TPC	Tripeptide Composition	Whole	8000	{a,b,c,d}
		N-Terminal	8000	{a}
		C-Terminal	8000	{a}
		Rest	8000	{a}
		Split	8000*N	{a}
ABC	Atom and Bond Composition	Whole	9	{a}
		N-Terminal	9	{a}
		C-Terminal	9	{a}
		Rest	9	{a}
		Split	9*N	{a}
COMPOSITION: PHYSICO-CHEMICAL PROPERTIES				
PCP	Physico-Chemical properties composition	Whole	19	{a,b,c,d,e}
		N-Terminal	19	{a}
		C-Terminal	19	{a}
		Rest	19	{a}

		Split	19*N	{a}
AAI	Amino Acid Index Composition	Whole	553	{a,b,c}
		N-Terminal	553	{a}
		C-Terminal	553	{a}
		Rest	553	{a}
		Split	553*N	{a}
PCP_adv	Advanced Physico-Chemical properties composition	Whole	5	{a,b,c,d,e}
		N-Terminal	5	{a}
		C-Terminal	5	{a}
		Rest	5	{a}
		Split	5*N	{a}
PCP_str	Structural Physico-Chemical properties composition	Whole	6	{a,b,c,d,e}
		N-Terminal	6	{a}
		C-Terminal	6	{a}
		Rest	6	{a}
		Split	6	{a}
COMPOSITION: REPEATS & DISTRIBUTION				
RRI	Repetitive Residue Information	Whole	20	{a}
		N-Terminal	20	{a}
		C-Terminal	20	{a}
		Rest	20	{a}
		Split	20*N	{a}
PRI	Repeat of Physico-chemical Properties	Whole	19	{a}
		N-Terminal	19	{a}
		C-Terminal	19	{a}
		Rest	19	{a}
		Split	19*N	{a}
DDR	Distance Distribution of Residues	Whole	20	{a}
		N-Terminal	20	{a}
		C-Terminal	20	{a}
		Rest	20	{a}
		Split	20*N	{a}
COMPOSITION: SHANNON ENTROPY				

SEP	Shannon Entropy at Protein Level	Whole	1	{a}
		N-Terminal	1	{a}
		C-Terminal	1	{a}
		Rest	1	{a}
		Split	1*N	{a}
SER	Shannon Entropy at Residue Level	Whole	20	{a}
		N-Terminal	20	{a}
		C-Terminal	20	{a}
		Rest	20	{a}
		Split	20*N	{a}
SPC	Shannon Entropy at Property Level	Whole	19	{a}
		N-Terminal	19	{a}
		C-Terminal	19	{a}
		Rest	19	{a}
		Split	19*N	{a}
COMPOSITION: MISCELLANEOUS				
ACR	Autocorrelation Descriptors	Whole	1659	{a,b,c,d,e}
		N-Terminal	1659	{a}
		C-Terminal	1659	{a}
		Rest	1659	{a}
		Split	1659*N	{a}
CTC	Conjoint Triad Descriptors	Whole	343	{a,b,c,d,e}
		N-Terminal	343	{a}
		C-Terminal	343	{a}
		Rest	343	{a}
		Split	343*N	{a}
CeTD	Composition enhanced Transition Distribution	Whole	189	{a,b,c,d,e}
		N-Terminal	189	{a}
		C-Terminal	189	{a}
		Rest	189	{a}
		Split	189*N	{a}
PAAC	Pseudo Amino Acid Composition	Whole	20 + λ	{a,b,c,d,e}
		N-Terminal	20 + λ	{a}
		C-Terminal	20 + λ	{a}
		Rest	20 + λ	{a}
		Split	N*(20 + λ)	{a}
APAAC	Amphiphilic Pseudo Amino Acid Composition	Whole	20 + (λ*3)	{a,b,c,d,e}
		N-Terminal	20 + (λ*3)	{a}
		C-Terminal	20 + (λ*3)	{a}
		Rest	20 + (λ*3)	{a}

		Split	N*(20 + (λ*3))	{a}
QSO	Quasi-Sequence Order	Whole	40 + (λ*2)	{a,b,c,d,e}
		N-Terminal	40 + (λ*2)	{a}
		C-Terminal	40 + (λ*2)	{a}
		Rest	40 + (λ*2)	{a}
		Split	N*(40 + (λ*2))	{a}
(SOCN)	Sequence Order Coupling Number	Whole	λ*2	{a,b,c,d,e}
		N-Terminal	λ*2	{a}
		C-Terminal	λ*2	{a}
		Rest	λ*2	{a}
		Split	N*λ*2	{a}
BINARY PROFILES				
AAB	Amino Acid Binary Profile	Whole	20*L	{a,b}
		N-Terminal	20*L	{a}
		C-Terminal	20*L	{a}
		Rest	20*L	{a}
		Split	N*(20*L)	{a}
DPB	Dipeptide Binary Profile	Whole	400*L	{a}
		N-Terminal	400*L	{a}
		C-Terminal	400*L	{a}
		Rest	400*L	{a}
		Split	N*(400*L)	{a}
ABB	Atom and Bond Binary Profile	Whole	(5*η)+(4*ε)	{a}
		N-Terminal	(5*η)+(4*ε)	{a}
		C-Terminal	(5*η)+(4*ε)	{a}
		Rest	(5*η)+(4*ε)	{a}
		Split	N*((5*η)+(4*ε))	{a}
PCB	Physico-Chemical Properties Binary Profile	Whole	25*L	{a}
		N-Terminal	25*L	{a}
		C-Terminal	25*L	{a}
		Rest	25*L	{a}
		Split	N*25*L	{a}
AIB	Amino Acid Index Binary Profile	Whole	553*L	{a}
		N-Terminal	553*L	{a}
		C-Terminal	553*L	{a}
		Rest	553*L	{a}
		Split	N*553*L	{a}
EVOLUTIONARY INFORMATION				
G_PSSM	Generation of PSSM	Whole	L X 21	{a}
N_PSSM	Normalization of PSSM	Whole	L X 21	{a}
C_PSSM	Composition of PSSM	Whole	400	{a}
P_PSSM	Profile of PSSM	Whole	L X 21	{a}

		N-Terminal	L X 21	{a}
		C-Terminal	L X 21	{a}
		Rest	L X 21	{a}
STRUCTURE				
FIN	Fingerprints	Whole	14532	{a}
SMI	SMILES	Whole	1	{a}
SA	Surface Accessibility	Whole	9	{a}
SS	Secondary Structure	Whole	3	{a}
PATTERN				
Binary Profile	Binary Profile generated using patterns of window length (ω)	Whole	L X (21*ω)	{a}
PSSM Profile	PSSM Profile generated using patterns of window length (ω)	Whole	L X (21*ω)	{a}
Physico-Chemical Properties	Physico-Chemical Properties, calculated using patterns of window length (ω)	Whole	L X (30*ω)	{a}
AA Index	Amino acid index composition, calculated using patterns of window length (ω)	Whole	L X 1	{a}
Universal	Generation of patterns of window length (ω)	Whole	L X ω	{a}
MODEL BUILDING				
Merging Features	Merge the two files into single file	2 CSV files	R X M	{a}
Feature Relevance	Mean based method to get the relevance of each feature	Positive and Negative Dataset	F X 9	{a}

a: Pfeature, b: ifeatue, c: PyBioMed, d: PyDPI, e: PROFEAT; L: length of protein; N: Number of splits; λ: The number depends upon the choice of maxlag; η: Number of atoms; ε: Number of bonds; R: Number of Rows; M: Total number of features in two files; F: Total number of features