Rcpi Quick Reference Card

Nan Xiao, Dongsheng Cao, Qingsong Xu

Package Version: Release 3
2014-09-10

COMPUTATIONAL BIOLOGY &
DRUG DESIGN GROUP
CENTRAL SOUTH UNIV., CHINA

Table 1: Retrieving protein sequence data from various online databases

Function name	Function description
`getProt()`	Retrieve protein sequence in FASTA format or PDB format from various online databases
`getFASTAFromUniProt()`	Retrieve protein sequence in FASTA format from UniProt
`getFASTAFromKEGG()`	Retrieve protein sequence in FASTA format from KEGG
`getPDBFromRCSBPDB()`	Retrieve protein sequence in PDB Format from RCSB PDB
`getSeqFromUniProt()`	Retrieve protein sequence from UniProt
`getSeqFromKEGG()`	Retrieve protein sequence from KEGG
`getSeqFromRCSBPDB()`	Retrieve protein sequence from RCSB PDB

Table 2: Retrieving drug molecular data from various online databases

Function name	Function description
`getDrug()`	Retrieve drug molecules in MOL format and SMILES format from various online databases
`getMolFromDrugBank()`	Retrieve drug molecules in MOL format from DrugBank
`getMolFromPubChem()`	Retrieve drug molecules in MOL format from PubChem
`getMolFromChEMBL()`	Retrieve drug molecules in MOL format from ChEMBL
`getMolFromKEGG()`	Retrieve drug molecules in MOL format from the KEGG
`getMolFromCAS()`	Retrieve drug molecules in InChI format from CAS
`getSmiFromDrugBank()`	Retrieve drug molecules in SMILES format from DrugBank
`getSmiFromPubChem()`	Retrieve drug molecules in SMILES format from PubChem
`getSmiFromChEMBL()`	Retrieve drug molecules in SMILES format from ChEMBL
`getSmiFromKEGG()`	Retrieve drug molecules in SMILES format from KEGG

Table 3: Calculating commonly used protein sequence derived descriptors

Function name	Descriptor name	Descriptor group
`extractProtAAC()`	Amino acid composition	Amino acid composition
`extractProtDC()`	Dipeptide composition	
`extractProtTC()`	Tripeptide composition	
`extractProtMoreauBroto()`	Normalized Moreau-Broto autocorrelation	Autocorrelation
`extractProtMoran()`	Moran autocorrelation	
`extractProtGeary()`	Geary autocorrelation	
`extractProtCTDC()`	Composition	CTD
`extractProtCTDT()`	Transition	
`extractProtCTDD()`	Distribution	
`extractProtCTriad()`	Conjoint Triad	Conjoint Triad
`extractProtSOCN()`	Sequence-order-coupling number	Quasi-sequence-order
`extractProtQSO()`	Quasi-sequence-order descriptors	
`extractProtPAAC()`	Pseudo-amino acid composition	Pseudo-amino acid composition
`extractProtAPAAC()`	Amphiphilic pseudo-amino acid composition	
`AAindex`	AAindex data of 544 physicochemical and biological properties for 20 amino acids	Dataset

Table 4: Generating profile-based protein representations

Function name	Function description
`extractProtPSSM()`	Compute PSSM (Position-Specific Scoring Matrix) for given protein sequence or peptides
`extractProtPSSMFeature()`	Profile-based protein representation derived by PSSM
`extractProtPSSMAcc()`	Profile-based protein representation derived by PSSM and auto cross covariance (ACC)

Table 5: Generating scales-based descriptors for proteochemometrics modeling

Function name	Descriptor class	Derived by
`extractPCMScales()`	Generalized scales-based descriptors derived by principal components analysis (PCA)	Principal components analysis
`extractPCMPropScales()`	Generalized scales-based descriptors derived by amino acid properties (AAindex)	
`extractPCMDescScales()`	Generalized scales-based descriptors derived by 2D and 3D molecular descriptors (Topological, WHIM, VHSE, etc.)	
`extractPCMFAScales()`	Generalized scales-based descriptors derived by factor analysis	Factor analysis
`extractPCMMDSScales()`	Generalized scales-based descriptors derived by multidimensional scaling (MDS)	Multidimensional scaling
`extractPCMBLOSUM()`	Generalized BLOSUM and PAM matrix-derived descriptors	Substitution matrix
`acc()`	Auto cross covariance (ACC) for generating scales-based descriptors of the same length	

Table 6: Pre-calculated molecular descriptor sets of the 20 amino acids in **Rcpi** for generating scales-based descriptors for proteochemometrics modeling. Note that the non-informative descriptors (like the descriptors have only one value across all the 20 amino acids) in these datasets have already been filtered out.

Dataset name	Dataset description	Dimensionality	Calculated by
`OptAA3d`	Optimized 20 amino acids	–	MOE
`AA2DACOR`	2D autocorrelations descriptors	92	Dragon
`AA3DMoRSE`	3D-MoRSE descriptors	160	Dragon
`AAACF`	Atom-centred fragments descriptors	6	Dragon
`AABurden`	Burden Eigenvalues descriptors	62	Dragon
`AAConn`	Connectivity indices descriptors	33	Dragon
`AAConst`	Constitutional descriptors	23	Dragon
`AAEdgeAdj`	Edge adjacency indices descriptors	97	Dragon
`AAEigIdx`	Eigenvalue-based indices descriptors	44	Dragon
`AAFGC`	Functional group counts descriptors	5	Dragon
`AAGeom`	Geometrical descriptors	41	Dragon
`AAGETAWAY`	GETAWAY descriptors	194	Dragon
`AAInfo`	Information indices descriptors	47	Dragon
`AAMolProp`	Molecular properties descriptors	12	Dragon
`AARandic`	Randic molecular profiles descriptors	41	Dragon
`AARDF`	RDF descriptors	82	Dragon
`AATopo`	Topological descriptors	78	Dragon
`AATopoChg`	Topological charge indices descriptors	15	Dragon
`AAWalk`	Walk and path counts descriptors	40	Dragon
`AAWHIM`	WHIM descriptors	99	Dragon
`AACPSA`	CPSA descriptors	41	Accelrys Discovery Studio
`AADescAll`	All the 2D descriptors calculated by Dragon	1171	Dragon
`AAMOE2D`	All the 2D descriptors calculated by MOE	148	MOE
`AAMOE3D`	All the 3D descriptors calculated by MOE	143	MOE
`AABLOSUM45`	BLOSUM45 matrix for 20 amino acids	20×20	Biostrings
`AABLOSUM50`	BLOSUM50 matrix for 20 amino acids	20×20	Biostrings
`AABLOSUM62`	BLOSUM62 matrix for 20 amino acids	20×20	Biostrings
`AABLOSUM80`	BLOSUM80 matrix for 20 amino acids	20×20	Biostrings
`AABLOSUM100`	BLOSUM100 matrix for 20 amino acids	20×20	Biostrings
`AAPAM30`	PAM30 matrix for 20 amino acids	20×20	Biostrings
`AAPAM40`	PAM40 matrix for 20 amino acids	20×20	Biostrings
`AAPAM70`	PAM70 matrix for 20 amino acids	20×20	Biostrings
`AAPAM120`	PAM120 matrix for 20 amino acids	20×20	Biostrings
`AAPAM250`	PAM250 matrix for 20 amino acids	20×20	Biostrings

Table 7: Molecular descriptors

Function name	Descriptor name
`extractDrugAIO()`	All the molecular descriptors in the **Rcpi** package
`extractDrugALOGP()`	Atom additive logP and molar refractivity values descriptor
`extractDrugAminoAcidCount()`	Number of amino acids
`extractDrugApol()`	Sum of the atomic polarizabilities
`extractDrugAromaticAtomsCount()`	Number of aromatic atoms
`extractDrugAromaticBondsCount()`	Number of aromatic bonds
`extractDrugAtomCount()`	Number of atom descriptor
`extractDrugAutocorrelationCharge()`	Moreau-Broto autocorrelation descriptors using partial charges
`extractDrugAutocorrelationMass()`	Moreau-Broto autocorrelation descriptors using atomic weight
`extractDrugAutocorrelationPolarizability()`	Moreau-Broto autocorrelation descriptors using polarizability
`extractDrugBCUT()`	BCUT, the eigenvalue based descriptor
`extractDrugBondCount()`	Number of bonds of a certain bond order
`extractDrugBPol()`	Sum of the absolute value of the difference between atomic polarizabilities of all bonded atoms in the molecule
`extractDrugCarbonTypes()`	Topological descriptor characterizing the carbon connectivity in terms of hybridization
`extractDrugChiChain()`	Kier & Hall Chi chain indices of orders 3, 4, 5, 6 and 7
`extractDrugChiCluster()`	Kier & Hall Chi cluster indices of orders 3, 4, 5 and 6
`extractDrugChiPath()`	Kier & Hall Chi path indices of orders 0 to 7
`extractDrugChiPathCluster()`	Kier & Hall Chi path cluster indices of orders 4, 5 and 6
`extractDrugCPSA()`	Descriptors combining surface area and partial charge information
`extractDrugDescOB()`	Molecular descriptors provided by OpenBabel
`extractDrugECI()`	Eccentric connectivity index descriptor
`extractDrugFMF()`	FMF descriptor
`extractDrugFragmentComplexity()`	Complexity of a system
`extractDrugGravitationalIndex()`	Mass distribution of the molecule
`extractDrugHBondAcceptorCount()`	Number of hydrogen bond acceptors
`extractDrugHBondDonorCount()`	Number of hydrogen bond donors
`extractDrugHybridizationRatio()`	Molecular complexity in terms of carbon hybridization states
`extractDrugIPMolecularLearning()`	Ionization potential
`extractDrugKappaShapeIndices()`	Kier & Hall Kappa molecular shape indices
`extractDrugKierHallSmarts()`	Number of occurrences of the E-State fragments
`extractDrugLargestChain()`	Number of atoms in the largest chain
`extractDrugLargestPiSystem()`	Number of atoms in the largest Pi chain
`extractDrugLengthOverBreadth()`	Ratio of length to breadth descriptor
`extractDrugLongestAliphaticChain()`	Number of atoms in the longest aliphatic chain
`extractDrugMannholdLogP()`	LogP based on the number of carbons and hetero atoms
`extractDrugMDE()`	Molecular Distance Edge (MDE) descriptors for C, N and O
`extractDrugMomentOfInertia()`	Principal moments of inertia and ratios of the principal moments
`extractDrugPetitjeanNumber()`	Petitjean number of a molecule
`extractDrugPetitjeanShapeIndex()`	Petitjean shape indices
`extractDrugRotatableBondsCount()`	Number of non-rotatable bonds on a molecule
`extractDrugRuleOfFive()`	Number failures of the Lipinski's Rule Of Five
`extractDrugTPSA()`	Topological Polar Surface Area (TPSA)
`extractDrugVABC()`	Volume of a molecule
`extractDrugVAdjMa()`	Vertex adjacency information of a molecule
`extractDrugWeight()`	Total weight of atoms
`extractDrugWeightedPath()`	Weighted path (Molecular ID)
`extractDrugWHIM()`	Holistic descriptors described by Todeschini et al.
`extractDrugWienerNumbers()`	Wiener path number and wiener polarity number
`extractDrugXLogP()`	Prediction of logP based on the atom-type method called XLogP
`extractDrugZagrebIndex()`	Sum of the squared atom degrees of all heavy atoms

Table 8: Molecular fingerprints

Function name	Fingerprint type
`extractDrugStandard()`	Standard molecular fingerprints (in compact format)
`extractDrugStandardComplete()`	Standard molecular fingerprints (in complete format)
`extractDrugExtended()`	Extended molecular fingerprints (in compact format)
`extractDrugExtendedComplete()`	Extended molecular fingerprints (in complete format)
`extractDrugGraph()`	Graph molecular fingerprints (in compact format)
`extractDrugGraphComplete()`	Graph molecular fingerprints (in complete format)
`extractDrugHybridization()`	Hybridization molecular fingerprints (in compact format)
`extractDrugHybridizationComplete()`	Hybridization molecular fingerprints (in complete format)
`extractDrugMACCS()`	MACCS molecular fingerprints (in compact format)
`extractDrugMACCSComplete()`	MACCS molecular fingerprints (in complete format)
`extractDrugEstate()`	E-State molecular fingerprints (in compact format)
`extractDrugEstateComplete()`	E-State molecular fingerprints (in complete format)
`extractDrugPubChem()`	PubChem molecular fingerprints (in compact format)
`extractDrugPubChemComplete()`	PubChem molecular fingerprints (in complete format)
`extractDrugKR()`	KR (Klekota and Roth) molecular fingerprints (in compact format)
`extractDrugKRComplete()`	KR (Klekota and Roth) molecular fingerprints (in complete format)
`extractDrugShortestPath()`	Shortest Path molecular fingerprints (in compact format)
`extractDrugShortestPathComplete()`	Shortest Path molecular fingerprints (in complete format)
`extractDrugOBFP2()`	FP2 molecular fingerprints
`extractDrugOBFP3()`	FP3 molecular fingerprints
`extractDrugOBFP4()`	FP4 molecular fingerprints
`extractDrugOBMACCS()`	MACCS molecular fingerprints

Table 9: Protein-protein and compound-protein interation descriptors

Function name	Function description
`getPPI()`	Generating protein-protein interaction descriptors
`getCPI()`	Generating compound-protein interaction descriptors

Table 10: Similarity and similarity searching

Function name	Function description
`calcDrugFPSim()`	Calculate drug molecule similarity derived by molecular fingerprints
`calcDrugMCSSim()`	Calculate drug molecule similarity derived by maximum common substructure search
`searchDrug()`	Parallelized drug molecule similarity search by molecular fingerprints similarity or maximum common substructure search
`calcTwoProtSeqSim()`	Similarity calculation based on sequence alignment for a pair of protein sequences
`calcParProtSeqSim()`	Parallellized protein sequence similarity calculation based on sequence alignment
`calcTwoProtGOSim()`	Similarity calculation based on Gene Ontology (GO) similarity between two proteins
`calcParProtGOSim()`	Protein similarity calculation based on Gene Ontology (GO) similarity

Table 11: Protein sequence data manipulation

Function name	Function description
`readFASTA()`	Read protein sequences in FASTA format
`readPDB()`	Read protein sequences in PDB format
`segProt()`	Protein sequence segmentation
`checkProt()`	Check if the protein sequence's amino acid types are the 20 default types

Table 12: Molecular data manipulation

Function name	Function description
`readMolFromSDF()`	Read molecules from SDF files and return parsed Java molecular object
`readMolFromSmi()`	Read molecules from SMILES files and return parsed Java molecular object or plain text list
`convMolFormat()`	Chemical file formats conversion

Affiliation:

Nan Xiao
School of Mathematics and Statistics
Central South University
Changsha, Hunan, P. R. China
E-mail: road2stat@gmail.com
URL: http://r2s.name

Dongsheng Cao
School of Pharmaceutical Sciences
Central South University
Changsha, Hunan, P. R. China
E-mail: oriental-cds@163.com
URL: http://cbdd.csu.edu.cn

Qingsong Xu
School of Mathematics and Statistics
Central South University
Changsha, Hunan, P. R. China
E-mail: dasongxu@gmail.com