Turkish Journal of Agriculture - Food Science and Technology, 5(1): 65-70, 2017

Turkish Journal of Agriculture - Food Science and Technology

Available online, ISSN: 2148-127X
www.agrifoodscience.com,
Turkish Science and Technology

Working with Proteins *in silico*: A Review of Online Available Tools for Basic Identification of Proteins

Caner Yavuz*, Zahide Neslihan Öztürk

Department of Agricultural Genetic Engineering, Ayhan Şahenk Faculty of Agricultural Science and Technology, Ömer Halisdemir University, Central Campus, 51240 Niğde, Turkey

ARTICLE INFO

Review articles
Received 21 July 2016
Accepted 02 January 2016

Keywords:

Protein
Annotation
Tool
Bioinformatics
In silico

*Corresponding Author:
E-mail: caneryavuznnn@gmail.com

ABSTRACT

Increase in online available bioinformatics tools for protein research creates an important opportunity for scientists to reveal characteristics of the protein of interest by only starting from the predicted or known amino acid sequence without fully depending on experimental approaches. There are many sophisticated tools used for diverse purposes. However, there are not enough reviews covering the tips and tricks in selecting and using the correct tools as the literature mainly states the promotion of the new ones. In this review, with the aim of providing young scientists with no specific experience on protein work a reliable starting point for *in silico* analysis of the protein of interest, we summarized tools for annotation of proteins. Annotation has included identification of motifs and domains, determination isoelectric point, molecular weight, subcellular localization, and post-translational modifications by focusing on the important points to be considered while selecting from online available tools.

Introduction

Development of new generation sequencing technologies in transcriptomics and mass spectrometry-based proteomics enabled generation of massive amount of data in DNA, RNA and protein levels. Identifying the function of a protein *in vivo*, on the other hand, still requires a functional genomics approach involving detection of tissue-specific promoter activities, subcellular localization, possible post-translation modifications, and defining knock-outs or over-expressing mutants (Salzano and Crescenzi, 2005; Free et al., 2009). There are quite number of bioinformatics tools that can be used for *in silico* analysis of a protein to design a functional genomics approach (Table 1), thereby, in this manuscript we aimed to summarize online available tools for characterization of the protein of interest through *in silico* calculation of molecular weight and isoelectric point, detection of protein motifs and domains, estimation of subcellular localization and possible post-translational modifications.

Annotation and Prediction of the Structure

The first step in a protein work is the annotation, i.e. naming, thus, identifying the protein coded by DNA or amino acid sequence at hand (Wilson et al., 2000). The establishment of Gene Ontology Consortium made it possible to tackle massive amount of sequence data by converting it into more structured and publically available form (Ashburner et al., 2000). There are three important ontologies that the Consortium proposed for a well-prepared study: prediction of the function of the gene product *per se*, i.e. molecular function, estimation of the role of its protein in a metabolic process, and the location of its product in the cell (Ashburner et al., 2000). *In silico* annotation of a protein, therefore, mainly involves prediction of its function by means of homology, possible interactions and post-translational modifications, as well as, its subcellular localization.

It is important to note that availability of high-throughput sequencing technologies enabled the scientists to work with predicted proteins from the genes encoded by the genome. There are many protein specific databases including Protein Data Bank (Berman et al., 2002), UniProt (Universal Protein Resource) Database (Apweiler et al., 2004), and NCBI (The National Center for Biotechnology Information) Protein Database (Gish and States, 1993). Apart from these common databases, there are also species-specific resources such as The Arabidopsis Information Resource (TAIR) (Huala et al., 2001), *Solanum tuberosum* genome database (PGSC et

al., 2011), and *Medicago truncatula* Genome Database (MTGD) (Krishnakumar et al., 2015). Gramene Protein Database with diverse species opportunity (barley, Brachypodium, foxtail millet, maize, oat, pearl millet, rye, sorghum, wheat, rice, wild rice and other Oryza), for example, offers a very informative data (Stein et al., 2002). The most widely used database for protein studies is UniProt which was established on 2002 with the decision to create one-hand information center by combining European Bioinformatics Institute (EBI) and Swiss Institute of Bioinformatics (SIB) (The UniProt Consortium, 2008). Currently, UniProt is serving in two ways; Swiss-Prot and TrEMBL, where both basically provide users with, if available, post-translational modification of the protein of interest, family or domain detection, prediction of its subcellular localization, and possible interactions with other proteins (Apweiler et al., 2004). As of July 2016, TrEMBL provides over 65,000,000 entries waiting to be fully annotated, and SwissProt mostly functions as a final destination to which annotated sequences from TrEMBL or other locations can be transferred and stored.

Prediction of the annotation or function of a protein is also possible through identification of conserved motifs and/or domains in protein families. Most of the bioinformatics tools for detecting motifs and domains mainly benefit from the fact that closely related species share common regulatory or functional regions. Motifs and domains are two terms that are often confused; a motif is a structural unit including conserved sequence motif, whereas domain is the conserved functional motif. In other terms, while sequence motifs concern the primary sequence, functional motif is mostly about the secondary structure. There are many tools available that provide motif search including MEME (Multiple Em for Motif Elicitation) (Bailey et al., 2009) and MiniMotif Miner (Balla et al., 2006). Existing tools perform motif search both in DNA and protein sequences; however the results of motif search in DNA sequence are mostly uncreditable due to the presence of short and degenerate sequences. Conversely, motif search through protein sequence is more reliable, and there is less likelihood to get incorrect results because of the insufficiency of the program (Bailey et al., 2006).

Prediction of secondary and/or tertiary structure of a protein through electron microscopy, X-Ray and NMR (Nuclear Magnetic Resonance) can clearly help to reveal the main function by enabling the prediction of active and/or ligand binding sites. Protein Data Bank was generated for this purpose; to create an open community macromolecular structure database and, as of July 2015, it contains 120,388 protein structures both in secondary and tertiary levels. This database enables the users to search existing protein structures based on homology and ligand binding capacity. The database actively involves protein structures of *Homo sapiens*, *Escherichia coli*, *Mus musculus*, *Bos taurus*, *Saccharomyces cerevisiae* and *Rattus norvegicus*, but there is a lack of structure predictions for the most plant proteins (Berman et al., 2000).

Calculation of Isoelectric Point and Molecular Weight

For the determination of molecular weight (MW) and isoelectric point (pI) via mass spectroscopy, the most critical point is the isolation of the pure protein, which, in most cases, cannot be performed due to, for example, low stability *in vitro*, or requirement of ions for structural stability. Prediction of MW and pI is also critical for the design of 2D-gel electrophoresis in separation of proteins in a proteomics approach. Knowing the pI of a protein, i.e. the pH in which the total charge of a protein is accepted to be zero, is also important to adjust its solubility, especially for isolation and storage purposes. Keep in mind that a protein in water with a pH very close to the protein's pI will mostly lead to minimal solubilization.

The most widely used approach for MW and pI determination based on average pKa (acid disassociation constant) value of the protein sequence at hand is by ExPASy (Expert Protein Analysis System) supported Compute pI/MW tool (Bjellqvist et al., 1993; Bjellqvist et al., 1994; Gasteiger et al., 2005; Henriksson et al., 1995). There are other tools available for *in silico* prediction of pI and MW such as TagIdent (Gasteiger et al., 2005; Wilkins et al., 1998a) and Multildent (Wilkins et al., 1998b); however they only accept sequences already available in SwissProt or UniProt databases. Despite the advantage of *in silico* predicted pI, it should be cautiously accepted as bioinformatically calculated pI may not always be reliable (Kiraga et al., 2007; Garcia-Moreno, 2009). There are some important shortcomings present in those calculations such that pH range capacity is restricted with acidity, therefore pI for highly basic proteins cannot be detected (Bjellqvist et al., 1993; Bjellqvist et al., 1994), and post-translational modifications are excluded in all calculations (Hoogland et al., 2000).

Prediction of Subcellular Localization

Determination of the subcellular localization of a protein is important to correctly define its function, especially for receptor proteins that reside in membrane systems (Geda et al., 2008). Experimental prediction of a subcellular localization of the protein of interest requires labeling with a dye, mostly EGFPs (Green Fluorescent Protein), and approaches like transient expression in a model plant system (Llopis et al., 1998). Such approaches, however, do not always lead to successful outcomes.

There are several online available bioinformatics tools that can provide a reliable subcellular localization in a various types of organisms. Current bioinformatics tools commonly used for the prediction of subcellular localization are divided into three in terms of their search criteria: (i) amino acid composition, (ii) signatures on the protein, and (iii) homology-based (Scott et al., 2004). SignalP (Nielsen et al., 1997), TargetP (Emanuelsson et al., 2000) and Predotar (Small et al., 2004) use some basic machine-learning methods, including neural networks (Reinhardt and Hubbard, 1998) and support vector

Table 1 Shows the commonly used online bioinformatics tools used to annotate protein sequences

Tool Name	Web Adress	Reference
Database		
Gramene Protein Database	http://archive.gramene.org/protein/	Stein et al., 2012
Medicago truncatula Genome Database	http://medicago.jcvi.org/MTGD/?q=home	Krishnakumar et al., 2014
NCBI Protein Database	http://www.ncbi.nlm.nih.gov/protein	Gish and States, 1993
Protein Data Bank	http://www.rcsb.org/pdb/home/home.do	Berman et al., 2002
Solanum tuberosum Genome Database	http://www.plantgdb.org/StGDB/	PGSC et al., 2011
The Arabidopsis Information Resource	https://www.arabidopsis.org	Huala et al., 2001
UniProt Database	http://www.uniprot.org	Apweiler et al., 2004
Conserved Domain and Motif		
MEME	http://meme-suite.org	Bailey et al., 2009
MiniMotif Miner	http://mnm.engr.uconn.edu/MNM/SMSSearch Servlet	Balla et al., 2006
Molecular Weight and Isoelectric Point		
Compute pI/MW	http://web.expasy.org/compute_pi/	Bjellqvist et al., 1993 Bjellqvist et al., 1994 Gasteiger et al., 2005 Henriksson et al., 1995
Multident	http://web.expasy.org/multiident/	Wilkins et al., 1998b
TagIdent	http://web.expasy.org/tagident/	Gasteiger et al., 2005 Wilkins et al., 1998a
Subcellular Localization		
comPPI	http://comppi.linkgroup.hu	Veres and Gyurko, 2014
Predotar	https://urgi.versailles.inra.fr/predotar/predotar.html	Small et al., 2004
PSLPred	http://www.imtech.res.in/raghava/pslpred/	Bhasin et al., 2005
PSORTb	http://www.psort.org/psortb/	Yu et al., 2010
SignalP	http://www.cbs.dtu.dk/services/SignalP/	Nielsen et al., 1997
TargetP	http://www.cbs.dtu.dk/services/TargetP/	Emanuelsson et al., 2000
Compartments	http://compartments.jensenlab.org/Search	Binder and Pletsher-Frankild, 2014
Post-translational modifications		
Big-PI	http://mendel.imp.ac.at/gpi/plant_server.html	Eisenhaber et al., 2003
BSPAT	http://cbc.case.edu/BSPAT/about.jsp	Li et al. 2015
CSS-Palm	http://csspalm.biocuckoo.org	Zhou et al., 2006
GPI-SOM	http://gpi.unibe.ch	Frankhauser and Maser, 2005
GPP	http://comp.chem.nottingham.ac.uk/glyco/	Hamby and Hirst, 2008
iDNA-Methyl	http://www.jci-bioinfo.cn/iDNA-Methyl	Xiao and Qiu, 2015
iHyd-PseAAC	http://app.aporc.org/iHyd-PseAAC/	Chou, 2011
NBA-Palm	http://nbapalm.biocuckoo.org	Xue et al., 2006b
Net-Acet	http://www.cbs.dtu.dk/services/NetAcet/	Kiemer et al., 2005
NetPhosK 1.0	http://www.cbs.dtu.dk/services/NetPhosK/	Blom et al., 1999
PHOSIDA	http://141.61.102.18/phosida/index.aspx	Gunawerdana et al., 2011
Phospho.ELM	http://phospho.elm.eu.org	Diella et al., 2004
PhosphoSite	http://www.phosphosite.org/homeAction.action	Hornbeck et al., 2012
PredGPI	http://gpcr.biocomp.unibo.it/predgpi/	Pierloni et al., 2008
PredHydroxy	http://bioinfo.ncu.edu.cn/PredHydroxy.aspx	Qiu et al., 2015
PRENbase	http://mendel.imp.ac.at/PrePS/PRENbase/	Maurer-Stroh et al., 2007
PrePS	http://mendel.imp.ac.at/PrePS/	Maurer-Stroh and Eisenhaber, 2005
PPSP	http://www.phosphosite.org/homeAction.action	Xue et al.,2006
ScanSite	http://scansite.mit.edu	Obenauer et al., 2003
Sulfinator	http://web.expasy.org/sulfinator/	Monigatti et al., 2002
Sulfosite	http://sulfosite.mbc.nctu.edu.tw	Chang et al., 2009
SUMOsp	http://sumosp.biocuckoo.org	Xue et al., 2006a

*The web site addresses were retrieved as of 28.12.2016

machines (Hua and Sun, 2001), to predict localization on the amino acid composition level. These tools predict the proteins in chloroplast, mitochondria and endoplasmic reticulum, but incapable of predicting the proteins localized in other organelles or places. Another commonly used approach is to predict the localization through the protein characteristics including signal peptides and transmembrane motifs with the benefit of k-nearest neighbor and Bayesian network methods (Dubey and Chouhan, 2011). Nearly each protein carries a signal peptide signature specific to the location that it resides. For example, nuclear proteins have signal peptides that are not cleaved, proteins targeted from cytosol to mitochondria have 20-60 long amino acid sequence in their N-terminus, and transmembrane proteins have hydrophobic side chains. PSORTb (Yu et al., 2010) and PSLpred (Bhasin et al., 2005) uses this type of approach and the benefit of machine learning to improve the prediction of localization in different compartments for the proteins that have an ability to reside in different locations at different times. The last method for the prediction is mainly based on homology through phylogenetic relationship (Dubey and Chouhan, 2011).

The most important criteria of using subcellular localization prediction tools is being specific to organism itself through their signal signatures. Hence, every program is not suitable for all organisms. Therefore, the specificity of the tools to prokaryotes, fungi, plants or animals, should be identified before use to get proper match with signal. Another limitation is based on the design of the tool; some tools like comPPI (Veres and Gyurko, 2014), Compartments (Binder and Pletscher-Frankild, 2014) require accession number of the protein, which limits the search on already identified proteins.

Computation of Possible Post-Translational Modifications

Post-translational modification (PTM) is a process of the addition or removal of some chemical groups to amino acid structures to manage cellular organization in a controlled manner in several developmental stages. PTMs after translation increase the diversity of proteins in terms of their function on dedicated regions to modify tertiary structure of the protein in either reversible or irreversible way. For this reason, investigation of PTMs on proteins is important to reveal the process in which protein function and to predict the localization. There are more than 100 possible PTMs known, but just a few of them are properly studied and understood (Williams and Stone, 1995).

Experimental determination of PTMs on the protein of interest not just requires effective and functional isolation of protein from the organism, but also traditional and laborious techniques like Edman degradation, immunochemistry and several mass spectrometry applications. Some established databases and tools have really profound contributions for the detection of possible PTMs in a simpler manner as supported by many publications (Gunawardena et al., 2011; Bendtsen et al., 2004). PTMs manage their function in three different types of processes that takes place including enzymatic-addition or -removal of chemical groups to the amino acid skeleton, chemical change in the surrounding environment, and physically destined for degradation (Cordwell et al., 2013). Most commonly encountered PTMs include phosphorylation, glycosylation, methylation, sulfation, acetylation, hydroxylation and prenylation, so we mainly focused on the prediction of these PTMs.

Phosporylation: Phosphorylation holds the position of being the most prevalent PTM and two types of enzymes, kinases and phosphatases, are involved in this type of modification. Due to ability to remove formerly added phosphate group (by kinase) by phosphatases, this type of modification is considered to be reversible (Ubersax and Ferrell, 2007). Phosphorylation mostly occurs on serine, threonine and tyrosine amino acids (Sefton and Hunter, 1998). Tools searching for phosphorylation in proteins use this strategy to facilitate the prediction of possible phosphorylation sites include PhoshoSite (Hornbeck et al., 2012), Phospho.ELM (Diella et al., 2004), PPSP (Xue et al., 2006), ScanSite (Obenauer et al., 2003), and NetPhosK 1.0 (Blom et al., 1999).

Glycosylation: According to SWISS-PROT, more than half of the proteins were destined to be glycosylated to be functional (Apweiler et al., 1999). Membrane proteins and secretory proteins need glycosylation through ER-Golgi pathway to benefit from solubility and hydrophobicity of glycosylated proteins to avoid undesired intermolecular interactions and to define the functional shape of the protein. There are five different types of glycosylation which are N-linked, O-linked, glypiation, C-linked and phosphoglycosylation. Possible glycosylation sites search for a consensus sequence like Asparagine-XXX (any type of amino acid other than proline) Serine/Threonine/Cysteine in N-linked glycosylation. However in O-linked glycosylation, there are no known conserved sequence pattern, instead in this situation they have preference for the serine and threonine residues (Varki et al., 2009).

Present bioinformatics tools specifically search for any kind of glycosylation types. There are several other tools which are online accessible to search for different types of modifications at one position, like GPP (Hamby and Hirst, 2008). Tools for the GPI (glycophosphatidylinositol)-anchor search includes Big-PI (Eisenhaber et al., 2003), GPI-SOM (Frankhauser and Maser, 2005) and PredGPI (Pierleoni et al., 2008).

Acetylation: Acetylation effectively takes place in histone modifications which is very important for epigenetic control of gene expression, and is thought to be responsible from various functions of proteins which makes this PTM prominent topic for bioinformaticians (Valdes-Mora et al., 2012). Acetylation can occur in two ways: co-translationally, which commonly includes the transfer of acetyl groups to the N-alpha-terminal group of proteins, and post-translationally in which acetylation occur in the N-terminal epsilon-lysine (Polevoda and Sherman, 2000). The latter type of acetylation modification was shown to have an essential role in the regulation of protein-DNA interaction (Vuzman et al., 2012). Today, very limited number of acetylation search tools are available. One of those tools is NetAcet (Kiemer et al., 2005), which is able to find only the proteins which

are undergone N-alpha-terminal acetylation, but not N-epsilon-lysine acetylation. With the development of PHOSIDA (Gunawardena et al., 2011), an important gap in lysine acetylation search has become possible. PHOSIDA can make multiple searches for different post-translational modifications including acetylation, phosphorylation and glycosylation, but if we evaluate the tool in terms of organism type, for the acetylation, search is only accessible for human proteins for now.

Beside these very commonly encountered modifications, there are tools currently being developed for the other post-translational modifications involving SUMOsp (Xue et al., 2006a) for sumoylation, CSS-Palm (Zhou et al., 2006) and NBA-Palm (Xue et al., 2006b) for the palmitoylation, and Sulfinator (Monigatti et al., 2002) and SulfoSite (Chang et al., 2009) for the sulfation, BSPAT (Li et al., 2015), iDNA-Methyl (Xiao and Qui 2015) for methylation, iHyd-PseAAC (Chou, 2011) and PredHydroxy (Qiu et al., 2015) for hydroxylation, PrePS (Maurer-Stroh and Eisenhaber, 2005) and PRENbase (Maurer-Stroh et al., 2007) for prenylation.

Conclusions

The main aim of this review was to provide a baseline for students with no experience on protein studies to create a reliable background to design their functional genomics approach for proper identification of their protein of interest. For that, we reviewed commonly used online available bioinformatics tools in Table 1 and the important factors to be considered while choosing and using these tools.

References

Apweiler R, Bairoch A, Wu HC, Barker WC, Boeckmann B, Ferro S, Gasteiger E, Huang H, Lopez R, Magrane M, Martin MJ, Natale DA, O'Donovan C, Redaschi N, Yeh LSL. 2004. UniProt: the Universal Protein knowledgebase. Nucleic Acids Res, 32: D115-119.

Apweiler R, Hermjakob H, Sharon N. 1999. On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. Biochim. Biophys. Acta, Gen. Subj., 1473: 4–8.

Ashburner M, Ball AC, Blake AJ, Botstein D, Butler H, Cherry M, David PA, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G. 2000. Gene Ontology: tool for the unification of biology, The Gene Ontology Consortium. Nat. Genet., 25(1): 25-29.

Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS. 2009. MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res., 37: 202-208.

Bailey TL, Williams N, Misleh C, Li WW. 2006. MEME: discovering and analyzing DNA and protein sequence motifs . Nucleic Acids Res., 34: 369-373.

Balla S, Thapar V, Verma S, Luong T, Faghri T, Huang CH, Rajasekaran S, del Campo JJ, Shinn JH, Mohler WA, Maciejewski MW, Gryk MR, Piccirillo B, Schiller SR, Schiller MR. 2006. Minimotif Miner: a tool for investigating protein function. Nat Methods. 3(3): 175-7.

Bendtsen JD, Jensen LJ, Blom N, von Heijne G, Brunak S. 2004. Feature-based prediction of non-classical and leaderless protein secretion. Protein Eng. Des. Sel., 17(4): 349–56.

Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. 2000. The Protein Data Bank. Nucleic Acids Res, 28: 235-242.

Bhasin M, Garg A, Raghava GP. 2005. PSLpred: prediction of subcellular localization of bacterial proteins. Bioinformatics, 21: 2522–2524.

Binder JX, Pletscher-Frankild S, Tsafou K, Stolte C, O'Donoghue SI, Schneider R, Jensen LJ. 2014. COMPARTMENTS: unification and visualization of protein subcellular localization evidence. Database, 2014, 1-9.

Bjellqvist B, Basse B, Olsen E, Celis JE. 1994. Reference points for comparisons of two-dimensional maps of proteins from different human cell types defined in a pH scale where isoelectric points correlate with polypeptide compositions. Electrophoresis. 15: 529-539.

Bjellqvist B, Hughes G, JPasquali Ch, Paquet N, Ravier F, Sanchez JCh,., Frutiger S., Hochstrasser D.F. 1993. The focusing positions of polypeptides in immobilized pH gradients can be predicted from their amino acid sequences. Electrophoresis. 14: 1023-1031.

Blom N, Gammeltoft S, Brunak S. 1999. Sequence- and structure-based prediction of eukaryotic protein phosphorylation sites. J. Mol. Biol., 294(5): 1351-1362.

Chang WC, Lee TY, Shien DM, Hsu JBK, Horng JT, Hsu PC, Wang TY, Huang HD, Pan RL. 2009. Incorporating support vector machine for identifying protein tyrosine sulfation sites. J. Comput. Chem., 30(15): 2526-37.

Chou KC. 2011. Some remarks on protein attribute prediction and pseudo amino acid composition (50^{th} Anniversary Year Review). Journal of Theorotical Biology, 273: 236-247.

Diella F, Cameron S, Gemünd C, Linding R, Via A, Kuster B, Sicheritz-Ponten T, Blom N, Gibson TJ. 2004. Phospho.ELM: a database of experimentally verified phosphorylation sites in eukaryotic proteins. BMC Bioinformatics, 5(79): 1-5.

Dubey A, Chouhan U. 2011. Subcellular localization of proteins. Arch. Appl. Sci. Res., 3(6): 392-401.

Eisenhaber B, Wildpaner M, Schultz CJ, Borner GHH, Dupree P, Eisenhaber F. 2003. Glycosylphosphatidylinositol lipid anchoring of plant proteins. Sensitive prediction from sequence- and genome- wide studies for Arabidopsis and rice. Plant Physiol., 133(4): 1691-1701.

Emanuelsson O, Nielsen H, Brunak S, von Heijne G. 2000. Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. J. Mol. Biol., 3000: 1015-1016.

Frankhauser N, Maser P. 2005. Identification of GPI anchor attachment signals and by a Kohonen self-organizing map. Bioinformatics, 21 (9): 1846-52.

Free RB, Hazelwood LA, Sibley DR. 2009. Identifying novel protein-protein interactions using co-immunoprecipitation and mass spectrometry. In: Free RB, Hazelwood LA, Sibley DR. Current Protocols in Neuroscience. Place of publication: Hoboken, New Jersey. Current Protocols Editorial Office, John Wiley and Sons, Inc. 28. 9780471142300

Garcia-Moreno B. 2009. Adaptations of proteins to cellular and subcellular pH. J. Biol., 8: 98.

Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, Bairoch A. 2005. Protein Identification and Analysis Tools on the ExPASy Server. In: Walker JM. The Proteomics Protocols Handbook. Place of publication: Hertfordshire. Humana Press. 571-607. 978-1-59259-890-8.

Geda P, Patury S, Ma J, Bharucha N, Dobry CJ, Lawson SK, Gestwicki JE, Kumar A. 2008. A small molecule- directed approach to control protein localization and function. Yeast, 25: 577-594.

Gish W, States DJ. 1993. Identification of protein coding regions by database similarity search. Nature Genet., 3: 266-272.

Gnad F, Gunawardena J, Mann M. 2011. PHOSIDA 2011: the posttranslational modification database. Nucleic Acids Res., 39: D253-60.

Hamby SE, Hirst JD. 2008. Prediction of glycosylation sites using random forests. BMC Bioinformatics, 9(500): 1-13.

Henriksson G, Englund AK, Johansson G, Lundahl P. 1995. Calculation of the isoelectric points of native proteins with spreading of pKa values. Electrophoresis, 16(8): 1377-1380.

Hoogland C, Sanchez JC, Tonella L. 2000. The 1999 SWISS-2DPAGE database update. Nucleic Acids Res., 28: 286–288.

Hornbeck PV, Kornhauser JM, Tkachev S, Zhang B, Skrzypek E, Murray B, Latham V, Sullivan, M. 2012. PhosphoSitePlus: a comprehensive resource for investigating the structure and function of experimentally determined post-translational modifications in man and mouse. Nucleic Acids Res., 40: 261-270.

Hu K, Ting AH, and Li, J. 2015. BSPAT: a fast online tool for DNA methylation co-occurrence pattern analysis based on high-throughput bisulfite sequencing data. BMC Bioinformatics, 16(1): 1.

Hua S, Sun Z. 2001. Support vector machine approach for protein sub-cellular localization prediction. Bioinformatics, 17: 721–728.

Huala E, Dickerman A, Garcia-Hernandez M, Weems D, Reiser L, LaFond F, Hanley D, Kiphart D, Zhuang J, Huang W, Mueller L, Bhattacharyya D, Bhaya D, Sobral B, Beavis B, Somerville C, Rhee SY. 2001. The Arabidopsis Information Resource (TAIR): A comprehensive database and web-based information retrieval, analysis, and visualization system for a model plant. Nucleic Acids Res, 29 (1): 102-5.

Kiemer L, Bendtsen JD, Blom N. 2005. NetAcet: Prediction of N-terminal acetylation. Bioinformatics, 21(7): 1269-1270.

Kiraga J, Mackiewicz P, Mackiewicz D, Kowalczuk M, Biecek P, Polak N, Smolarczyk K, Dudek MR, Cebrat S. 2007. The relationship between the isoelectric point and: length of proteins, taxonomy and ecology of organisms. BMC Genomics, 8: 163.

Krishnakumar V, Kim M, Rosen BD, Karamycheva S, Bidwell SL, Tang H, Town CD. 2015. MTGD: The Medicago truncatula Genome Database. Plant Cell Pysiol., 56 (1): 1-9.

Liddy KA, White MY, Cordwell SJ. 2013. Function decorations: post-translational modifications and heart disease delineated by targeted proteomics. Genome Med., 5(2): 1-12.

Liu Zi, Xiao X, Qiu WR, Chou KC. 2015. iDNA- Methyl: Identifying DNA methylation sites via pseudo trinucleotide composition. Analytical Biochemistry, 474: 69-77.

Llopis J, McCaffery M, Miyawaki A, Farquhar MG, Tsien RY. 1998. Measurement of cytosolic, mitochondrial, and Golgi pH in single living cells with green fluorescent proteins. Proc. Natl. Acad. Sci., 95(12): 6803-6808.

Maurer-Stroh S, Eisenhaber F. 2005. Refinement and prediction of protein prenylation motifs. Genome Biology, 6: R55.

Maurer-Stroh S, Koranda M, Benetka W, Schneider G, Sirota FL, Eisenhaber F. 2007. Towards complete sets of farnesylated and geranylgeranylated proteins. PLoS Computational Biology, 3(4): e66.

Monigatti F, Gasteiger E, Bairoch A, Jung E. 2002. The Sulfinator: predictig tyrosine sulfation sites in protein sequences. Bioinformatics, 18: 769-770.

Nielsen H, Engelbrecht J, Brunak S, von Heijne G. 1997. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavege site. Protein Eng., 10: 1-6.

Obenauer JC, Cantley LC, Yaffe MB. 2003. ScanSite 2.0: proteome- wide prediction of cell signaling interactions using short sequence motifs. Nucleic Acids Res., 31(13): 3635-3641.

Pierleoni A, Martelli PL, Casadio R. 2008. PredGPI: a GPI-anchor predictor. BMC Bioinformatics, 9: 1-11.

Polevoda B, Sherman F. 2000. N-alpha-terminal acetylation of eukaryotic proteins. J. Biol. Chem. 275: 36479- 36482.

Reinhardt A, Hubbard T. 1998. Using neural networks for prediction of the subcellular location of proteins. Nucleic Acids Res., 26: 2230–2236.

Salzano AM, Crescenzi M. 2005. Mass spectrometry for protein identification and the study of post translational modifications. Ann. Ist. Super. Sanita, 41(4): 443-450.

Scott MS, Thomas DY, Hallett MT. 2004. Predicting Subcellular Localization via Protein Motif Co-occurence. Genome Res., 14: 1957-1966.

Sefton BM, Hunter T. 1998. Protein Phosphorylation. In: Abelson JN, Simon MI. Methods in Enzimology. 1st edition. Place of publication: San Diego. Academic Press. 978-0121821029.

Shi S, Chen X, Xu H, Qiu J. 2015. PredHydroxy: computational prediction of protein hydroxylation site locations based on the primary structure. Molecular BioSystems. 11: 819-825.

Small I, Peeters N, Legeai F, Lurin C. 2004. Predotar: A tool for rapidly screening proteomes for N-terminal targeting sequences. Proteomics, 4: 1581–1590.

The Potato Genome Sequencing Consortium (PGSC). 2011. Genome sequence and analysis of the tuber crop potato. Nature, 475: 189-195.

The UniProt Consortium. 2008. The Universal Protein Resource (UniProt). Nucleic Acids Res., 36: D190-D195.

Ubersax JA, Ferrell JE. 2007. Mechanism of spesifity in protein phosphorylation. Nat. Rev. Mol. Cell Biol., 8(7): 530-541.

Valdes-Mora F, Song JZ, Statham AL, Strbenac D, Robinson MD, Nair SS, Patterson KI, Tremethick DH, Stirzaker C, Clark SJ. 2012. Acetylation of H2A.Z is a key epigenetic modification associated with gene deregulation and epigenetic remodelling in cancer. Genome Res., 22(2): 307-321.

Varki A, Esko JD, Colley KJ. 2009. Cellular Organization of Glycosylation. In: Varki A, Esko JD, Colley KJ, Freeze HH, Stanley P, Bertozzi CR, Hart GW, Etzler ME. Essentials of Glycobiology. 2nd edition. Place of publication: La Jolla, California. Cold Spring Harbor Laboratory Press. 9780879697709.

Veres DV, Gyurko DM, Thaler B, Szalay KZ, Fazekas D, Korcsmaros T, Csermely P. 2014. ComPPI: a cellular compartment-specific database for protein-protein interaction network analysis. Nucleic Acids Res., 43: D485-D493.

Vuzman D, Hoffman Y, Levy Y. 2012. Modulating protein –DNA interactions by post-translational modifications at disordered regions. In: Altman RB, Dunker AK, Hunter L, Murray TA, Klein TE. Pacific Symposium on Biocomputing. Hawai, USA. 3-7 January 2012. 188-189.

Ware D, Jaiswal P, Ni J,Pan X, Chang K, Clark K, Teytelman L, Schmidt S, Zhao W, Cartinhour S, McCouch S, Stein L. 2002. Gramene: a resource for comparative grass genomics. Nucleic Acids Res., 30(1): 103-105.

Wilkins MR, Gasteiger E, Tonella L, Keli O, Tyler M, Sanchez JC, Gooley AA, Walsh BJ, Bairoch A, Appel RD, Williams KL, Hochstrasser DF. 1998a. Protein Identification with N and C-terminal Sequence Tags in Proteome Projects. J. Mol. Biol. 278(3): 599-608.

Wilkins MR, Gasteiger E, Wheeler C, Lindskog I, Sanchez JC, Bairoch A, Appel RD, Dunn MD, Hochstrasser D.F. 1998b. Multiple parameter cross-species protein identification using MultiIdent - a world-wide web accessible tool. Electrophoresis. 19(18): 3199-206.

Williams KR, Stone KL. 1995. Identifying sites of post-translational modifications in proteins via HPLC peptide mapping. In: Shirley BA. Methods in Molecular Biology- Protein Stability and Folding. Place of publication: New Haven. Humana Press. 157- 175. 978-1-59259-527-3.

Wilson CA, Kreychman J, Gerstein M. 2000. Assessing annotation transfer for genomics: quantifying the relations between protein sequence, structure and function through traditional and probabilistic scores. J. Mol. Biol., 297: 233-249.

Xue Y, Chen H, Jin C, Sun Z, Yao X. 2006b. NBA-Palm: prediction of palmitoylation site implemented in Naive Bayes Algorithm. BMC Bioinformatics, 7(458): 1-10.

Xue Y, Li A, Wang L, Feng H, Yao X. 2006. PPSP: prediction of PK-spesific phosphorylation site with Bayesian decision theory. BMC Bioinformatics, 7(163): 1-12.

Xue Y, Zhou F, Fu C, Xu Y, Yao X. 2006a. SUMOsp: a web sever for sumoylation site prediction. Nucleic Acids Res., 34: 254-257.

Yu NY, Wagner JR, Laird MR, Melli G, Rey S, Lo R, Dao P, Sahinalp CS, Ester M, Foster LJ, Brinkman FSL. 2010. PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilitives for all prokaryotes. Bioinformatics, 26(13): 1608-1615.

Zhou F, Xue Y, Yao X, Xu Y. 2006. CSS-Palm: palmitoylation site prediction with a clustering and scoring strategy (CSS). Bioinformatics, 22(7): 894-6.