BIOINFORMATICS APPLICATIONS NOTE Vol. 28 no. 17 2012, pages 2293–2295
doi:10.1093/bioinformatics/bts436

Databases and ontologies

Advance Access publication July 10, 2012

dbSNO: a database of cysteine *S*-nitrosylation

Tzong-Yi Lee[1,*], Yi-Ju Chen[2], Cheng-Tsung Lu[1], Wei-Chieh Ching[2], Yu-Chuan Teng[1], Hsien-Da Huang[3,4] and Yu-Ju Chen[2,*]

[1]Department of Computer Science and Engineering, Yuan Ze University, Taoyuan 320, [2]Institute of Chemistry, Academia Sinica, Taipei 115, [3]Institute of Bioinformatics and Systems Biology, National Chiao Tung University, Hsin-Chu 300 and [4]Department of Biological Science and Technology, National Chiao Tung University, Hsin-Chu 300, Taiwan

Associate Editor: Jonathan Wren

ABSTRACT

Summary: *S*-nitrosylation (SNO), a selective and reversible protein post-translational modification that involves the covalent attachment of nitric oxide (NO) to the sulfur atom of cysteine, critically regulates protein activity, localization and stability. Due to its importance in regulating protein functions and cell signaling, a mass spectrometry-based proteomics method rapidly evolved to increase the dataset of experimentally determined SNO sites. However, there is currently no database dedicated to the integration of all experimentally verified *S*-nitrosylation sites with their structural or functional information. Thus, the dbSNO database is created to integrate all available datasets and to provide their structural analysis. Up to April 15, 2012, the dbSNO has manually accumulated >3000 experimentally verified *S*-nitrosylated peptides from 219 research articles using a text mining approach. To solve the heterogeneity among the data collected from different sources, the sequence identity of these reported *S*-nitrosylated peptides are mapped to the UniProtKB protein entries. To delineate the structural correlation and consensus motif of these SNO sites, the dbSNO database also provides structural and functional analyses, including the motifs of substrate sites, solvent accessibility, protein secondary and tertiary structures, protein domains and gene ontology.

Availability: The dbSNO is now freely accessible via http://dbSNO.mbc.nctu.edu.tw. The database content is regularly updated upon collecting new data obtained from continuously surveying research articles.

Contacts: francis@saturn.yu.edu.tw or yujuchen@gate.sinica.edu.tw

Supplementary Information: Supplementary data are available at *Bioinformatics* online.

Received on May 15, 2012; revised on June 22, 2012; accepted on July 3, 2012

1 INTRODUCTION

Of the various post-translational modifications (PTMs), *S*-nitrosylation (SNO) is reversible and involves the covalent attachment of nitric oxide (NO) to the thiol group of cysteine (Cys) residues. An increasing number of studies suggest that protein *S*-nitrosylation plays an important role in the NO-related redox pathway, especially in immune, cardiovascular, neuronal and plant systems (Bogdan, 2001; Gaston *et al*., 2006; Jaffrey *et al*., 2001; Karpuzoglu and Ahmed, 2006; Lindermayr *et al*., 2005; Stamler *et al*., 2001). The various targets of *S*-nitrosylation and differential expression of those targets modulate the activity, localization and stability of proteins (Hess *et al*., 2005; Lam *et al*., 2010; Nakamura and Lipton, 2009) and further regulate patho-physiological events, such as neurodegenerative diseases and cancers (Cho *et al*., 2009; Hess *et al*., 2005; Yao *et al*., 2004). Due to the importance of *S*-nitrosylation in biological processes, numerous efforts have been directed toward mass spectrometry-based *S*-nitrosyl-proteomic studies using various biological systems to increase significantly the number of known *S*-nitrosylated peptides (Chen *et al*., 2010; Greco *et al*., 2006; Hao *et al*., 2006; Murray *et al*., 2011). Among these studies, the biotin-switch method is the most commonly adapted method for the site-specific identification of protein *S*-nitrosylation (Jaffrey *et al*., 2001). As the number of experimentally identified *S*-nitrosylated peptides grows, a structured database of protein *S*-nitrosylation is desirable for the further investigation of the biological functions of *S*-nitrosylated proteins and the substrate specificities of SNO sites. Although a number of databases have been developed for post-translational modifications, such as dbPTM (Lee *et al*., 2006), Phospho.ELM (Dinkel *et al*., 2011), PhosphoSite (Hornbeck *et al*., 2004), O-GLYCBASE (Gupta *et al*., 1999), UbiProt (Chernorudskiy *et al*., 2007), dbOGAP (Wang *et al*., 2011) and PupDB (Tung, 2012), there is no database dedicated to *S*-nitrosylated proteins and their corresponding substrate sites. Thus, a new database (dbSNO) containing the experimentally verified *S*-nitrosylated peptides from research articles is proposed to facilitate with the functional analysis of *S*-nitrosylated proteins and assist with the structural investigation of *S*-nitrosylation sites.

2 METHODS

The system flow of the construction of the dbSNO is presented in Figure 1. The dbSNO contains the experimentally verified data for *S*-nitrosylated proteins and *S*-nitrosylation sites. All the database entries are manually extracted from research articles through a literature survey. First, all fields in the PubMed database are searched based on the keywords '*S*-nitrosylation' or '*S*-nitrosylated' followed by downloading the full text of these research articles. For the various proteomic identification experiments (Derakhshan *et al*., 2007; Greco *et al*., 2006; Hao *et al*., 2006; Murray *et al*., 2011; Wang *et al*., 2008), a text-mining system is developed to survey the full-text literature that potentially describes the site-specific identification of *S*-nitrosylated sites. Approximately, 400 original and review articles associated with protein *S*-nitrosylation are retrieved from PubMed (April 2012). Next, the full-length articles are manually reviewed for the precise extraction of the *S*-nitrosylated peptides and the

* To whom correspondence should be addressed.

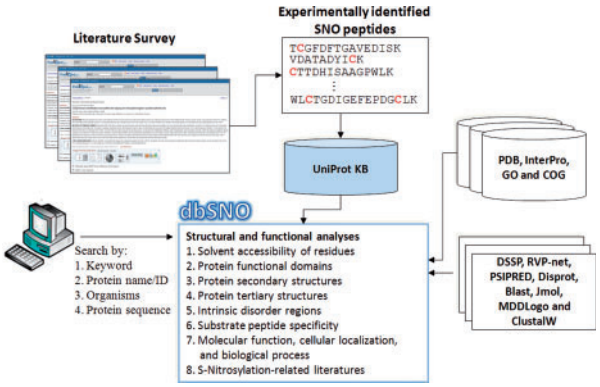

Fig. 1. System flow of dbSNO

modified cysteines. To determine the locations of *S*-nitrosylated cysteines within a full-length protein sequence, the experimentally verified *S*-nitrosylated peptides are then mapped to UniProtKB (Bairoch *et al.*, 2005) protein entries based on database identifier or sequence similarity. The *S*-nitrosylated peptides that cannot align exactly to a protein sequence are discarded. Finally, each mapped *S*-nitrosylation site is attributed to the corresponding literature (PubMed ID). In addition, a small number of protein *S*-nitrosylation sites in UniProtKB are integrated into dbSNO with the attributed literature. Up to April 15, 2012, a total of 219 *S*-nitrosylation-associated articles covering 18 organisms are retrieved from PubMed. In the dbSNO database, there are 3374 *S*-nitrosylated cysteines in 1757 *S*-nitrosylated proteins. Supplementary Table S1 shows the statistics for the *S*-nitrosylation data in each organism.

For a given protein, the biological function can be obtained from the annotation of UniProtKB. To provide more effective information about protein functional and structural annotations relevant to cysteine *S*-nitrosylation, a variety of biological databases, such as InterPro (Hunter *et al.*, 2011), Gene Ontology (GO), Protein Data Bank (PDB) (Rose *et al.*, 2011) and dbPTM (Lee *et al.*, 2006), are integrated. In this study, the InterPro BioMart, which is a web-service and provides the ability to build custom queries on the InterPro database, is utilized to extract the information about protein domains and functional sites for the *S*-nitrosylated proteins in dbSNO. The information regarding the molecular function, cellular components and biological process for a *S*-nitrosylated protein can be accessed by a crosslink that refers to the corresponding entry from QuickGO (Binns *et al.*, 2009) through a UniProtKB accession number.

Accumulating studies reveal that the cysteine residue, having a low pK_a or a thiol group exposed on the protein surface, is more accessible by NO modification (Derakhshan *et al.*, 2007; Hess *et al.*, 2005). To study the preference of the solvent accessible surface area (ASA) that surrounds SNO sites within the protein tertiary structure, the experimentally identified *S*-nitrosylation sites should be mapped to the corresponding positions of the protein entries in the Protein Data Bank (PDB). The preference of the secondary structure around the modification sites is also considered. However, due to the limited information of protein structures in PDB, only ~3% of the *S*-nitrosylation sites have corresponding tertiary structures. With respect to the previous studies investigating the structural characteristics of PTMs (Lee *et al.*, 2011; Shien *et al.*, 2009) in proteins without known tertiary structures, two effective tools, RVP-net (Ahmad *et al.*, 2003) and PSIPRED (McGuffin *et al.*, 2000), are used to predict the solvent accessibility and secondary structure, respectively.

3 UTILITY

To facilitate the use of the dbSNO resource, a web interface has been developed for users to browse and search efficiently for their *S*-nitrosylated proteins of interest. Supplementary Figure S1 shows the content of a typical dbSNO entry, including basic information, graphical visualization of SNO sites, table of SNO sites with the associated literature and visualization of tertiary structures by the Jmol program. Based on the integration of the GO and InterPro databases, the distributions of the

Table 1. The MDDLogo-identified motifs for the mouse *S*-nitrosylation data

Group	Number of data	Entropy plot of MDDLogo-identified motif
MC1	335	
MC2	333	
MC3	281	
MC4	218	
MC5	183	
MC6	150	
MC7	116	
MC8	108	
MC9	83	
MC10	81	
MC11	59	
MC12	59	

biological processes and functional domains of *S*-nitrosylated proteins are presented in Supplementary Tables S2 and S3, respectively. The analysis of the predicted protein domains shows that ~70% of the reported SNO sites are located within functional domains (Supplementary Table S4).

To investigate the substrate specificity of the SNO sites, MDDLogo (Lee *et al*., 2011), a maximal dependence decomposition analysis that clusters the aligned signal sequences into subgroups containing statistically significant motifs was performed to identify the substrate motifs for the *S*-nitrosylation data in each organism with a sufficient dataset. As shown in the example of the mouse dataset (Table 1), 12 substrate motifs, which were identified from 2216 mouse SNO sites with a 21-mer window length, contain a conserved motif of positively charged amino acids (K, R and H) surrounding the SNO sites. In particular, the 11th group contains negatively charged amino acids (D and E) accompanied by positively charged amino acids in conserved motifs at two specific positions. Consistent with previous studies (Chen *et al*., 2010; Greco *et al*., 2006; Greco, 2006; Hao *et al*., 2006; Lane *et al*., 2001), the *S*-nitrosylated cysteines may be located within an acid–base motif flanked by acidic and basic amino acids.

4 CONCLUSION

The dbSNO database is the first public resource to allow efficient access to curated SNO sites, functional annotations, structural characteristics and substrate motifs for *S*-nitrosylated proteins. With regard to the computational prediction of SNO sites, dbSNO could provide the benchmark dataset for the development and comparison of prediction tools.

Funding: National Science Council of the Republic of China (NSC-100-2221-E-155-079 to T.Y.L.) and (NSC-100-2911-I-009-101 and NSC-100-2627-B-009-002 to H.D.H.) and (NSC-100-2627-M-001-003 to Y.J.C.).

Conflict of Interest: none declared.

REFERENCES

Ahmad,S. *et al*. (2003) RVP-net: online prediction of real valued accessible surface area of proteins from single sequences. *Bioinformatics*, **19**, 1849–1851.

Bairoch,A. *et al*. (2005) The Universal Protein Resource (UniProt). *Nucleic Acids Res*., **33**, D154–D159.

Binns,D. *et al*. (2009) QuickGO: a web-based tool for Gene Ontology searching. *Bioinformatics*, **25**, 3045–3046.

Bogdan,C. (2001) Nitric oxide and the immune response. *Nat. Immun*., **2**, 907–916.

Chen,Y.J. *et al*. (2010) S-alkylating labeling strategy for site-specific identification of the s-nitrosoproteome. *J. Proteome Res*., **9**, 6417–6439.

Chernorudskiy,A.L. *et al*. (2007) UbiProt: a database of ubiquitylated proteins. *BMC Bioinformatics*, **8**, 126.

Cho,D.-H. *et al*. (2009) *S*-nitrosylation of Drp1 mediates β-Amyloid-related mitochondrial fission and neuronal injury. *Science*, **324**, 102–105.

Derakhshan,B. *et al*. (2007) Balancing reactivity against selectivity: the evolution of protein *S*-nitrosylation as an effector of cell signaling by nitric oxide. *Cardiovasc. Res*., **75**, 210–219.

Derakhshan,B. *et al*. (2007) Unbiased identification of cysteine *S*-nitrosylation sites on proteins. *Nat. Protoc*., **2**, 1685–1691.

Dinkel,H. *et al*. (2011) Phospho.ELM: a database of phosphorylation sites-update 2011. *Nucleic Acids Res*., **39**, D261–D267.

Gaston,B. *et al*. (2006) S-nitrosothiol signaling in respiratory biology. *Am. J. Respir. Crit. Care Med*., **173**, 1186–1193.

Greco,T.M. *et al*. (2006) Identification of *S*-nitrosylation motifs by site-specific mapping of the S-nitrosocysteine proteome in human vascular smooth muscle cells. *Proc. Natl. Acad. Sci. USA*, **103**, 7420–7425.

Gupta,R. *et al*. (1999) O-GLYCBASE version 4.0: a revised database of O-glycosylated proteins. *Nucleic Acids Res*., **27**, 370–372.

Hao,G. *et al*. (2006) SNOSID, a proteomic method for identification of cysteine *S*-nitrosylation sites in complex protein mixtures. *Proc. Natl. Acad. Sci. USA*, **103**, 1012–1017.

Hess,D.T. *et al*. (2005) Protein *S*-nitrosylation: purview and parameters. *Nat. Rev. Mol. Cell Biol*., **6**, 150–166.

Hornbeck,P.V. *et al*. (2004) PhosphoSite: a bioinformatics resource dedicated to physiological protein phosphorylation. *Proteomics*, **4**, 1551–1561.

Hunter,S. *et al*. (2011) InterPro in 2011: new developments in the family and domain prediction database. *Nucleic Acids Res*., **40**, D306–D312.

Jaffrey,S.R. *et al*. (2001) Protein *S*-nitrosylation: a physiological signal for neuronal nitric oxide. *Nat. Cell Biol*., **3**, 193–197.

Karpuzoglu,E. and Ahmed,S.A. (2006) Estrogen regulation of nitric oxide and inducible nitric oxide synthase (iNOS) in immune cells: implications for immunity, autoimmune diseases, and apoptosis. *Nitric Oxide*, **15**, 177–186.

Lam,Y.W. *et al*. (2010) Comprehensive identification and modified-site mapping of S-nitrosylated targets in prostate epithelial cells. *PLoS One*, **5**, e9075.

Lane,P. *et al*. (2001) *S*-nitrosylation is emerging as a specific and fundamental posttranslational protein modification: head-to-head comparison with O-phosphorylation. *Sci. STKE*, **2001**, re1.

Lee,T.Y. *et al*. (2011) RegPhos: a system to explore the protein kinase-substrate phosphorylation network in humans. *Nucleic Acids Res*., **39**, D777–D787.

Lee,T.Y. *et al*. (2006) dbPTM: an information repository of protein post-translational modification. *Nucleic Acids Res*., **34**, D622–D627.

Lee,T.Y. *et al*. (2011) Exploiting maximal dependence decomposition to identify conserved motifs from a group of aligned signal sequences. *Bioinformatics*, **27**, 1780–1787.

Lindermayr,C. *et al*. (2005) Proteomic identification of S-nitrosylated proteins in *Arabidopsis*.. *Plant Physiol*., **137**, 921–930.

McGuffin,L.J. *et al*. (2000) The PSIPRED protein structure prediction server. *Bioinformatics*, **16**, 404–405.

Murray,C.I. *et al*. (2011) Identification and quantification of *S*-nitrosylation by cysteine reactive tandem mass tag switch assay. *Mol. Cell Proteomics*, **11**, M111013441.

Nakamura,T. and Lipton,S. (2009) Cell death: protein misfolding and neurodegenerative diseases. *Apoptosis*, **14**, 455–468.

Rose,P.W. *et al*. (2011) The RCSB Protein Data Bank: redesigned web site and web services. *Nucleic Acids Res*., **39**, D392–D401.

Shien,D.M. *et al*. (2009) Incorporating structural characteristics for identification of protein methylation sites. *J. Comput. Chem*., **30**, 1532–1543.

Stamler,J.S. *et al*. (2001) Nitrosylation. The prototypic redox-based signaling mechanism. *Cell*, **106**, 675–683.

Stamler,J.S. *et al*. (1997) (S)NO signals: translocation, regulation, and a consensus motif. *Neuron*, **18**, 691–696.

Tung,C.W. (2012) PupDB: a database of pupylated proteins. *BMC Bioinformatics*, **13**, 40.

Wang,J. *et al*. (2011) dbOGAP—an integrated bioinformatics resource for protein O-GlcNAcylation. *BMC Bioinformatics*, **12**, 91.

Wang,Y. *et al*. (2008) A strategy for direct identification of protein *S*-nitrosylation sites by quadrupole time-of-flight mass spectrometry. *J. Am. Soc. Mass Spectrom*., **19**, 1353–1360.

Yao,D. *et al*. (2004) Nitrosative stress linked to sporadic Parkinson's disease: *S*-nitrosylation of parkin regulates its E3 ubiquitin ligase activity. *Proc. Nat. Acad. Sci. USA*, **101**, 10810–10814.