Prediction of thioredoxin and glutaredoxin target proteins by identifying reversibly oxidized cysteinyl residues

Hang-mao Lee[1*], Karl Josef Dietz[2], and Ralf Hofestädt[1]

[1] Technical Faculty, Bielefeld University, Germany

[2] Faculty of Biology, Bielefeld University, Germany

[*] Corresponding author - hmlee@techfak.uni-bielefeld.de

Summary

A significant part of cellular proteins undergo reversible thiol-dependent redox transitions which often control or switch protein functions. Thioredoxins and glutaredoxins constitute two key players in this redox regulatory protein network. Both interact with various categories of proteins containing reversibly oxidized cysteinyl residues. The identification of thioredoxin/glutaredoxin target proteins is a critical step in constructing the redox regulatory network of cells or subcellular compartments. Due to the scarcity of thioredoxin/glutaredoxin target protein records in the public database, a tool called Reversibly Oxidized Cysteine Detector (ROCD) is implemented here to identify potential thioredoxin/glutaredoxin target proteins computationally, so that the in silico construction of redox regulatory network may become feasible. ROCD was tested on 46 thioredoxin target proteins in plant mitochondrion, and the recall rate was 66.7% when 50% sequence identity was chosen for structural model selection. ROCD will be used to predict the thioredoxin/glutaredoxin target proteins in human liver mitochondrion for our redox regulatory network construction project. The ROCD will be developed further to provide prediction with more reliability and incorporated into biological network visualization tools as a node prediction component. This work will advance the capability of traditional database- or text mining-based method in the network construction.

1 Introduction

Changes in redox balance and development of oxidative stress are associated with many cell functions and life processes including aging, diseases, loss of fitness, and yield [1,2,3,4]. On the molecular scale, oxidation will change the structure of biomolecules and often switches or tunes enzyme activity or causes enzyme malfunction. To keep the cellular environment in a proper redox state, cells contain several antioxidants, such as vitamin C, vitamin E, and ubiquinol and also antioxidant enzymes [5]. By decomposing reactive oxygen species (ROS) and reactive nitrogen species (RNS) these antioxidants constitute the first line of defense to avoid damage to macromolecules by uncontrolled oxidation. Once ROS or RNS escape from the first defense line, lipids, nucleic acids and also proteins may get oxidized. A major oxidation reaction of proteins is the dithiol-disulfide transition. Cells have developed two rescue systems that involve thioredoxins and glutaredoxins [6,7], respectively, to re-reduce the oxidized proteins. Because these two proteins are not engaged in metabolite turnover but couple redox input elements to the redox state of target proteins and thereby modify the activity of metabolic enzymes, they are termed ‘transmitters’ in the redox regulatory network [4].

The mitochondrion is a subcellular compartment where respiratory electron transport proceeds at high rate and with strong reduction potential differences. If NADH availability is high, electron carriers such as ubiquinone may become over-reduced during electron transport

at the mitochondrial inner membrane. In this case, electrons may be transferred to the oxygen molecule to produce superoxide, which is a strong oxidant [8]. Thus the thioredoxin and glutaredoxin systems in the mitochondrion play an important role in preventing this organelle from over-oxidized. Several researches have shown that oxidized mitochondrial proteins accumulate with aging and neural degenerative disease [1-3]. In order to ultimately simulate electron flow through the redox regulatory system, the thiol-disulfide protein network of human mitochondrion shall be constructed. This will allow testing of its capability to maintain redox homeostasis and to infer the biological outcome by the downstream regulated metabolic network in case of oxidative stress. The relationship between the redox regulatory network and the affected metabolic network is depicted in Fig. 1.

The proteins connecting the upstream regulatory network and the downstream metabolic network are the target proteins of thioredoxin/glutaredoxin. The critical step in expanding the redox regulatory network is to identify the thioredoxin/glutaredoxin target proteins in order to complete the network. The thioredoxin/glutaredoxin target proteins contain reversibly oxidized cysteines which form disulfide bonds with another cysteine bearing the same property when oxidized and are present in free thiol form when reduced. But unfortunately only few thioredoxin/glutaredoxin target proteins could be found in the public protein-protein interaction databases when human liver mitochondrion was chosen as our modeling environment. The lack of database and experimental information thus becomes an obstacle in the network construction process.

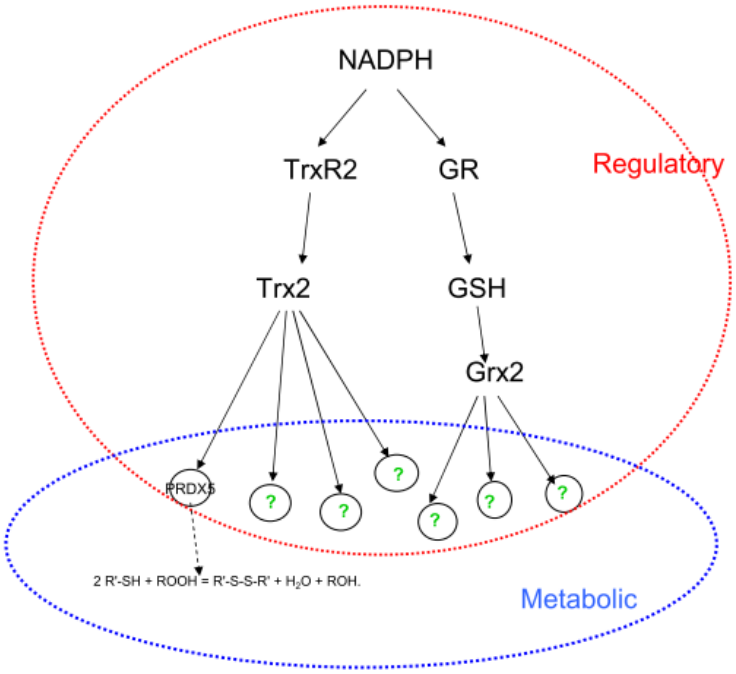

Fig. 1: Redox regulatory network and exemplarily affected metabolic network. Abbreviations: TrxR: thioredoxin reductase [9], GR: glutathione reductase, GSH: glutathione, Grx2: glutaredoxin.

With the advance of proteomics techniques, chromatography coupled with gel-based or mass spectrometry can experimentally identify thioredoxin/glutaredoxin target proteins or proteins that undergo thiol-disulfide transitions [10-13]. Due to technical limitation, further experiments are needed to eliminate the false positive proteins. In addition, specificity and sensitivity of the experimental techniques need to be improved in order to overcome the false positive problem and to identify target proteins with low abundance. Therefore this work is aimed at deploying the capability of bioinformatics to identify additional potential target protein candidates to complete the network and also to support the experimental approach.

This tool is intended to serve as a node prediction module for the automatic construction of redox regulatory network and also as an experimental candidate predictor for identifying thioredoxin/glutaredoxin target proteins.

2 Related works

The thioredoxin- and glutaredoxin-dependent regulatory networks are two redox systems in the cell sustaining the normal protein structure and function in an oxidizing environment. Both networks are composed of proteins with reversibly oxidized cysteine residues. According to the redox potential of each player in the network, electrons are transferred through specific donor-acceptor pairs [4-7]. Besides experimental strategies such as affinity chromatography and gel based methods [10-13], Marino and Gladyshev have adapted an integrative methodology in bioinformatics to detect thiol oxidoreductases and their catalytic redox-active cysteine residues [14]. Thioredoxins and glutaredoxins regulate various categories of proteins, such as proteins involved in photorespiration, citric acid cycle-associated reactions, lipid metabolism, electron transport, etc.[11,15]. Due to the functional diversity of the target proteins, simple sequence analysis usually fails in predicting novel target proteins. The parameters that characterize thiol-disulfide transition proteins are the vicinity of two cysteinyl residues, the pK_a value of the thiol, and the accessible surface area (ASA) of the cysteinyl residues on the target protein. Sanchez *et al.*[16] provided an algorithm based on the exact values of these three parameters, cysteine-cysteine distance, pK_a, and accessible surface area, for predicting the reversibly oxidized cysteines. Implementation of this approach as computational methodology for identification of thioredoxin or glutaredoxin target proteins is lacking.

Compared to redox regulatory networks, present day knowledge of metabolic networks is more advanced in the public domain. The physiological influence resulting from the activity change of the metabolic enzyme could be inferred by exploring the downstream reactions in the metabolic network. There are databases devoted to storing metabolic network information, such as KEGG [17], Reactome [18], MetaCyc [19], and Brenda [20]. Some web-based or stand-alone network visualization tools have been developed and are suitable for easy examination of the metabolic network [21-23]. The metabolic network content displayed by the visualization tools can be inputted by users or through the built-in network retrieval module which queries the metabolic network databases. Besides the network structure, visualization tools can depict the annotation for each node of enzyme and metabolite. Network visualization tools provide browsing, querying, editing, and analyzing functions for the metabolic network. Some visualization tools allow the third-party plug-in to interact with the host application, so that the functionality of such visualization tools may be extended.

Novel sequencing technologies have shortened the time needed to sequence complete genomes. However, subsequent to sequence acquisition, the genome must be annotated for its genes, proteins, and encoded biological pathway. Thus, the construction of biological pathways often relies on the quality of information stored in the database and the use of homology concept. Biological databases often store information only on general and well-known biological pathways. If the focus is placed on different and more specialized networks, the researcher must usually explore literature and molecular databases and construct the network from primary information. In addition to literature search, text mining tools are another choice. Construction of the network will be hindered if the necessary data is still lacking, or if the information from the database or text-mining tools proves to be too rich to handle, since reliable target selection may not be possible.

3 Methods

3.1 ROCD algorithm

ROCD follows the strategy of Sanchez *et al.* [16]. The algorithms implemented by ROCD are depicted in Fig. 2. Users are requested to input the file containing the SwissProt accession numbers, the criteria for the cysteine-cysteine distance, accessible surface area (ASA) and pK_a screening, and to define the file name to save the output. Users can also choose the desired tissue and organelle instead of providing the protein list for scanning tissue and organelle specific proteins. The generation of tissue and organelle specific protein set is relied on Human Protein Reference Database (HPRD) [29]. HPRD is a highly expert-curated human protein database.

Since the PDB file is essential for PropKa [26] and Naccess [25] prediction, the PDB ID is obtained by querying iProClass [27] database, and the yet structure-unsolved entry is queried against Swiss Model Repository. iProClass provides the ID mapping function which ROCD used to obtain PDB ID and the residing chains for each protein entry, and SWISS MODEL Repository [24] stores automated modeled protein structure by homology. To comply with the maximum 5000 atoms limitation from Naccess, the PDB file is tailored to contain the first model in the original PDB file.

The tailored PDB is used to calculate the distance between any pair of cysteines which are located on the chains specified in iProClass. If there is any calculated distance falling in the user-defined range, the protein entry and the qualified residues are written to the result file.

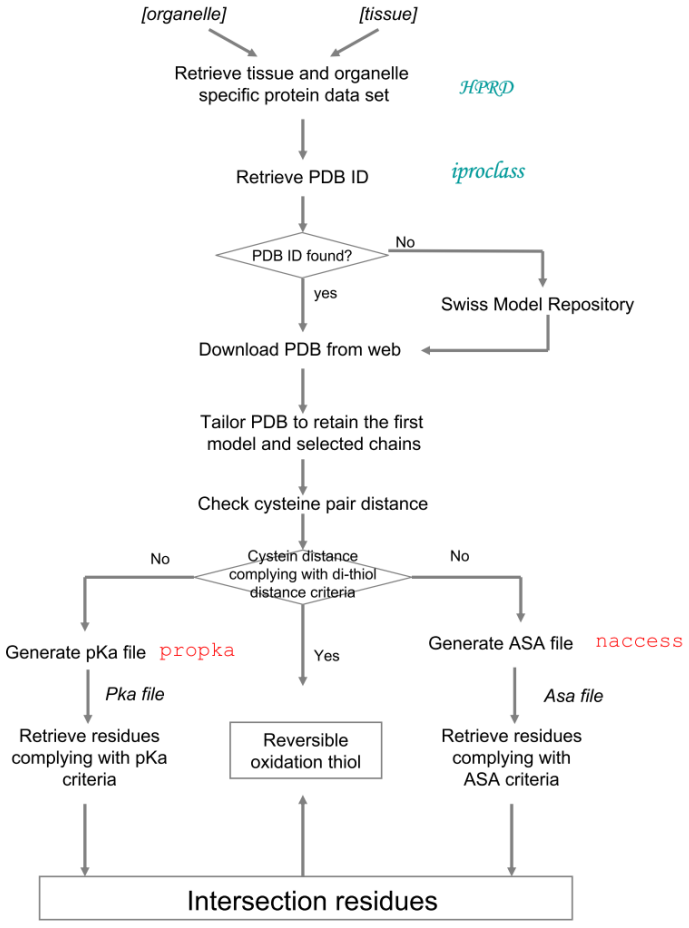

Fig. 2: The algorithm to predict protein with reversibly oxidized cysteinyl residues. The blue color denotes the database, and red denotes another standalone program.

If no cysteine-pair distance matches user's criteria, pK_a and ASA are calculated by PropKa and Naccess respectively with tailored PDB as the input. If the cysteine residue has the pK_a and ASA falling in the user-defined range and with the chain label as annotated in iProClass, the protein entry and the qualified residues are written to the result file.

During the execution of ROCD, the calculated cysteine-pair distance, pK_a, and ASA value for each cysteine residue are also written out to separate files for examination.

3.2 ROCD architecture

The architecture of ROCD is shown in Fig. 3. A command-line program and a web interface were created in this work. The program was implemented in Java, and the in-house HPRD and iProClass database [27] are stored in MySQL. The iProClass database is a collection of different molecular database accession numbers and makes the retrieval of corresponding molecular database accession numbers easier. The iProClass database flat file was downloaded from its website, and the data in each column of the flat file were parsed into different tables. In the web interface, users just provide: (i) Path to the file containing the SwissProt accession number, or to select tissue and organelle from the drop-down menu (ii) Criteria for the thiol-thiol distance, accessible surface area, and pK_a, (iii) file name to save the output files.

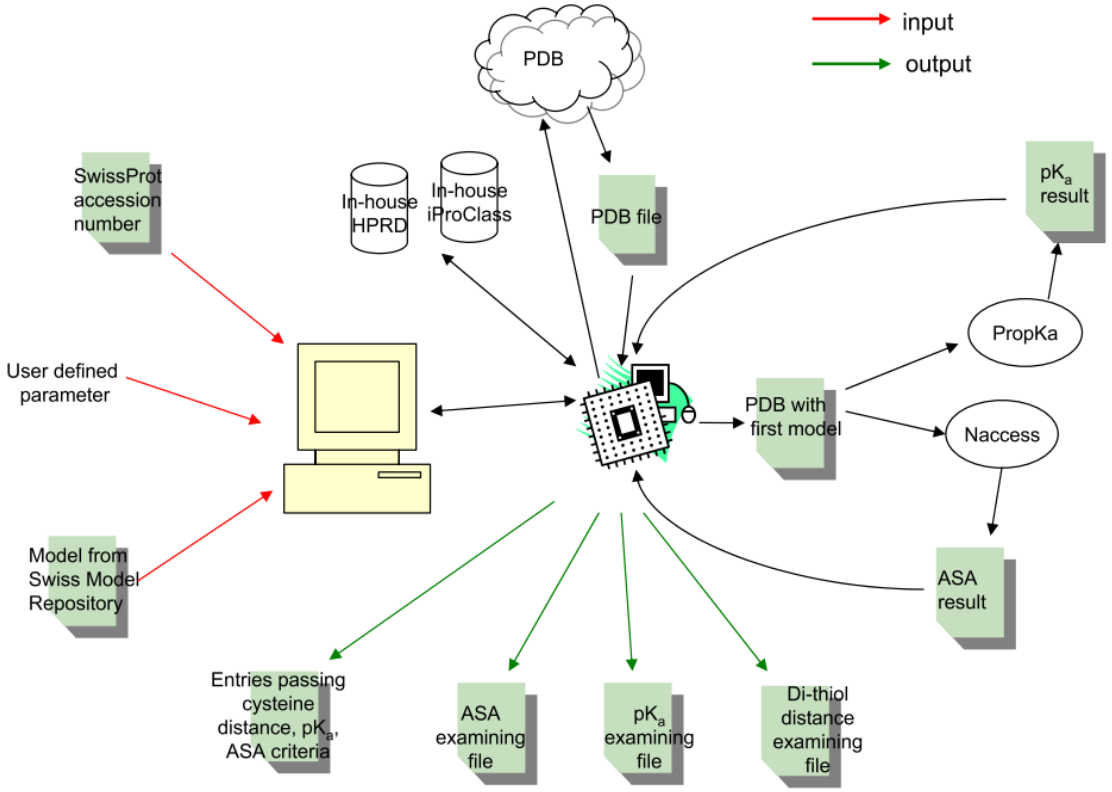

Fig. 3: The architecture of ROCD. Users are requested to provide three types of information to ROCD. ROCD utilizes three public databases (HPRD, iProClass, PDB) and two external tools (PropKa, Naccess) for the selection of reversibly oxidized cysteinyl residues.

In the back end, this program queries the in-house HPRD for tissue and organelle specific protein data set and iProClass database to find the PDB ID for each SwissProt accession number. Then the PDB file is downloaded from PDB database and is as the input for PropKa and Naccess. Finally, it generates four output files – one file for the distances of all cysteine pairs, one for all the calculated pK_a for cysteine residue, one for all the calculated ASA for the

SG atom, one for SwissProt accession numbers of potential proteins with reversibly oxidized cysteines.

3.3 Validation of ROCD prediction

The annotation of cysteine residues in Balanced Susceptible Cysteine Thiol Database (BALOSCTdb) from Sanchez et al.'s study was used as the gold standard to validate ROCD. BALOSCTdb contains 161 cysteine thiols that undergo reversible oxidation and 161 cysteine thiols that are not susceptible to oxidation. Each PDB ID in BALOSCTdb is tested by ROCD. The parameters were chosen as specified in Sanchez *et al*.- 6.2 for cysteine-cysteine distance, 1.3 for ASA, and 9.05 for pK_a.

3.4 Examination of thioredoxin target proteins in plant mitochondrion

Balmer *et al.* collected 46 thioredoxin target proteins in plant mitochondrion. The target protein of thioredoxin should contain reversibly oxidized cysteines which form transient disulfide bond with thioredoxin during interaction. We use ROCD to test the existence of reversibly oxidized cysteines for these 46 proteins.

4 Results

ROCD implements the algorithms suggested by Sanchez *et al*. [16] for reversibly oxidized cysteine detection. ROCD identifies proteins which fulfill the three criteria provided by the user: cysteine-cysteine distance, pK_a range, and solvent accessible area range, in a list of proteins provided by the user or with specific tissue and organelle localization.

Tab. 1: Result from testing ROCD on BALOSCTdb

		Actual condition	
		Non reversibly oxidized	**Reversibly oxidized**
Prediction result	**Non reversibly oxidized**	137	61
	Reversibly oxidized	24	100

After the implementation, the compliance of our prediction was checked with BALOSCTdb (Table 1). Our prediction achieved 62.1% accuracy for the cysteine residues which are marked "reversibly oxidized cysteine" in BALOSCTdb and 85.1 % for "non-reversibly oxidized cysteines". We also applied ROCD on 46 thioredoxin target proteins in Balmer *et al.* [15] with the reported 3 parameter values: (i) cysteine-cysteine distance $\leq$ 6.2 Å, (ii) accessible solvent area $\geq$ 1.3 $Å^2$, (iii) $pK_a \leq 9.05$(Fig. 3). Only 3 proteins have corresponding PDB IDs, and 36 unresolved ones have homology entries in SWISS MODEL Repository, and 3 could be modeled manually by Swiss Model. The modeled PDB structure from a template with less than 50% similarity to the query sequence was abandoned due to low reliability. The remaining proteins were inspected by ROCD, and the recall rate was 66.7% (Table 2).

Tab. 2: Result from applying ROCD on 46 thioredoxin target proteins in plants. The PDB ID in upper case letters are the exact structure entry for the protein, and the PDB ID in lower case is obtained by homology modeling. The two digits separated by underscore in Note1 column mean the cysteine residue pair, and Note2 column records the qualified residue number and residue name.

Uniprot Accession Number	Uniprot ID	Protein name	PDB ID	Modeled template	Sequence identity to template	Resideu pair with distance < 6.2	Residues with pK_a < 9.05 and ASA > 1.3
Q43644	NDUS1_SOLTU	NADH-ubiquinone oxidoreductase 75 kDa subunit, mitochondrial		2fug	26	114_128 100_111 100_128 215_218 215_262 111_114 100_114 212_262 167_173	
Q9LW15	Q9LW15_ARATH	Cytochrome c oxidase subunit Vb-like protein		1v54	37	122_146 146_149 122_149	
Q9ZPX5	DHSA2_ARATH	Succinate dehydrogenase [ubiquinone] flavoprotein subunit 2, mitochondrial		2fbw	67	175_177	
Q9FS87	IVD2_SOLTU	Isovaleryl-CoA dehydrogenase 2, mitochondrial		1ivh	65	177_230 326_336	
Q05046	CH62_CUCMA	Chaperonin CPN60-2, mitochondrial		1aon	56	244_305	
P37900	HSP7M_PEA	Heat shock 70 kDa protein, mitochondrial		2v7y	62	364_367	
P93541	DHE3_SOLLC	Glutamate dehydrogenase		1b26	48	382_387	
Q95P13	Q95P13_STRPU	Secreted frizzled-related protein precusor		1ijx	38	406_452 443_483 476_500 398_459 472_513	
P31023	DLDH_PEA	Dihydrolipoyl dehydrogenase, mitochondrial	1DXL			45_50	
O82514	KAD1_ARATH	Adenylate kinase 1		2c9y	53	60_110	
P93697	P93697_VIGUN	CPRD12 protein		2bgk	51	68_79	
P16048	GCSH_PEA	Glycine cleavage system H protein, mitochondrial	1HPC				124CYS
O48646	GPX6_ARATH	Probable phospholipid hydroperoxide glutathione peroxidase 6, mitochondrial		2p5q	72		105CYS 153CYS
P52902	ODPA_PEA	Pyruvate dehydrogenase E1 component subunit alpha, mitochondrial		1ni4	49		119CYS 53CYS
Q8LAV0	Q8LAV0_ARATH	Succinyl-CoA ligase beta subunit		1euc	49		133CYS 359CYS
Q9SIB9	ACO2M_ARATH	Aconitate hydratase 2, mitochondrial		2b3x	62		146CYS 914CYS
Q944P7	AMPL3_ARATH	Leucine aminopeptidase 3, chloroplastic		1gyt	34		209CYS 368CYS 451CYS
P55312	CATA2_SOLTU	Catalase isozyme 2		2j2m	47		230CYS 370CYS 400CYS 411CYS 420CYS 86CYS
Q9LEG5	Q9LEG5_SOLLC	Allene oxide cyclase		2brj	66		233CYS
O82450	O82450_ARATH	Branched-chain alpha-keto acid decarboxylase E1 beta subunit		1dtw	67		249CYS 86CYS
O48904	O48904_MEDSA	null		1smk	64		281CYS
P37225	MAON_SOLTU	NAD-dependent malic enzyme 59 kDa isoform, mitochondrial		1gq2	38		290CYS 452CYS
P29677	MPPA_SOLTU	Mitochondrial-processing peptidase subunit alpha		1hr6	32		298CYS
O64530	O64530_ARATH	YUP8H12R.17 protein		1boi	37		305CYS
Q9LKJ1	Q9LKJ1_ARATH	3-hydroxyisobutyryl-coenzyme A hydrolase		3bpt	39		308CYS
Q8RWN9	OPD22_ARATH	Dihydrolipoyllysine-residue acetyltransferase component 2 of pyruvate dehydrogenase complex, mitochondrial		3b8k	53		342CYS 385CYS 510CYS
P05493	ATPAM_PEA	ATP synthase subunit alpha, mitochondrial		2jdi	73		390CYS
P26969	GCSP_PEA	Glycine dehydrogenase		1wyt	36		587CYS 752CYS

		[decarboxylating], mitochondrial					
Q9S7E4	FDH_ARATH	Formate dehydrogenase, mitochondrial		2nac	52		58CYS
Q8LBK6	GRS15_ARATH	Monothiol glutaredoxin-S15		2wul	40		91CYS
P17614	ATPBM_NICPL	ATP synthase subunit beta, mitochondrial	1PYV				
P34899	GLYM_PEA	Serine hydroxymethyltransferase, mitochondrial		1eji	60		
P68209	SUCA1_ARATH	Succinyl-CoA ligase [GDP-forming] subunit alpha-1, mitochondrial		1euc	71		
P49364	GCST_PEA	Aminomethyltransferase, mitochondrial		1wsr	52		
Q43153	Q43153_SPIOL	null		1z7w	61		
Q9ZT91	EFTM_ARATH	Elongation factor Tu, mitochondrial		2hcj	93		
P52904	ODPB_PEA	Pyruvate dehydrogenase E1 component subunit beta, mitochondrial		2ozl	60		
Q9LEV3	UMP3_ARATH	Uncharacterized protein At5g10860, mitochondrial		2rc3	33		
P46643	AAT1_ARATH	Aspartate aminotransferase, mitochondrial		7aat	56		
O81233	O81233_GOSHI	Superoxide dismutase		1zte	57		
Q7XK22	Q7XK22_ORYSJ	null					
P93344	P93344_TOBAC	Aldehyde dehydrogenase (NAD+)					
P42056	VDAC2_SOLTU	Mitochondrial outer membrane protein porin of 36 kDa		2jk4	22		130CYS 235CYS
Q9SXV0	Q9SXV0_ORYSJ	Os03g0390400 protein					
Q9SUK9	P2C55_ARATH	Probable protein phosphatase 2C 55					
Q95I43	Q95I43_GADMO	MHC class Ia antigen		3bev	43	199_255 99_161	

5 Discussion

The goal of this work is to predict proteins that undergo reversible oxidization by examining the existence of reversibly oxidized cysteine. We implemented Sanchez *et al*'s algorithm to allow a high-throughput and automatic in silico prediction of thiol-disulfide transition proteins. There are several phases in the pipeline for achieving automation: database querying and data retrieving, external programs execution, and data processing. With the increasing speed of sequencing of new species, a high-throughput tool for protein property selection is needed.

The inventory of thiol-disulfide transition proteins is important for in silico construction of the redox regulatory network [4]. This redox regulatory network is a central and evolutionarily conserved feature of the cell. In addition, ROCD could serve as a subsidy to experimental approaches. The identification of reversibly oxidized thiols characterizes potential thioredoxin/glutaredoxin target proteins. In comparison to other thioredoxin/glutaredoxin target protein prediction methods, Sanchez *et al.* [16] provided a more complete framework to detect the necessary structure and physicochemical properties. In their leave-one-out cross-validation analysis, 80.1% accuracy was reported. By implementing the Sanchez algorithms and adapting their criteria, ROCD achieved an accuracy rate of 73.6%, the precision rate of 80.6%, when using BALOSCTdb as the gold standard. But users are also allowed to choose their own pK_a, ASA, and cysteine-cysteine distance value and let ROCD do the screening on their protein list.

We have applied ROCD on 46 thioredoxin target proteins in plant mitochondrion. Due to the shortage of resolved structure, the testing on these proteins relied mostly on modeled structures. After screening out the sequence identity lower than 50%, 24 proteins with structure were left, and 16 of them were detected to have reversibly oxidized residue by ROCD applying Sanchez *et al.*'s criteria.

The function of cysteinyl residues can be divided into several groups. These functional groups comprise redox-active catalytic, regulatory, structural, metal-coordinating, and catalytic non-redox thiol groups [28]. The Sanchez training data set includes the reversibly oxidized cysteines which belong to both the catalytic redox-active and regulatory groups. The Sanchez algorithm will predict reversibly oxidized cysteines with catalytic redox-active and regulatory function, but only the ones with regulatory function would be the primarily targeted cysteine residue for thioredoxin/glutaredoxin. A further refinement of ROCD should aim at implementing a subprogram based on additional criteria to filter out the cysteines with function in redox-dependent catalysis and thereby to restrict the output to redox regulatory cysteinyl residues.

Reliable verification of candidates for the redox network construction will either depend on experiments or involve text-mining to support the ROCD prediction. Once the confirmed target protein set is generated, the biological outcome can be inferred by the available metabolic network tool. In addition, it is intended to pack this tool into a plug-in in Cytoscape [21].

The identification of regulated target proteins is a critical step in the post-translational regulatory network construction. ROCD is applicable for the thioredoxin/glutaredoxin mediated redox regulatory network. Other regulatory networks, such as the phosphorylation network, also require a target prediction tool like ROCD, if the prediction algorithm is available.

Recent advances in molecular biology, biotechnology and related disciplines have generated a wealth of biological data deposited in diverse databases and literature. The identification and integration of information relevant for a defined purpose will help understand the interconnection between different biological components, for example, protein-protein interaction networks, gene regulatory networks or metabolic networks. A graphical network is usually constructed for explicit demonstration of the integrated biological knowledge. Since the network construction relies on the data stored in the biological database, the network construction work is hindered, once the required data is unavailable. This work provides a search strategy based on physicochemical and structural properties to fill the gap between specifically and limited knowledge deposited in literature and databases, and advancement of network construction.

ROCD is planned to develop a web interface which could be access at http://agbi.techfak.uni-bielefeld.de/ROCD.

References

[1] M. Martinez, A.I. Hernández, N. Martánez, M.L. Ferrándiz. Age-related increase in oxidized proteins in mouse synaptic mitochondria. Brain Res, 731(1-2): 246-248, 1996

[2] J.P. Blass, G.E. Gibson. The role of oxidative abnormalities in the pathophysiology of Alzheimer's disease. Rev Neurol(Paris), 147(6-7):513-25, 1991

[3] I. Ferrer. Early involvement of the cerebral cortex in Parkinson's disease: convergence of multiple metabolic defects. Prog Neurobiol, 88(2):89-103, 2009

[4] K.J. Dietz. Redox signal integration: from stimulus to networks and genes. Physiol Plant. 133(3): 459-468, 2008.

[5] K.J. Dietz. Redox control, redox signaling, and redox homeostasis in plant cells. Int Rev Cytol.: 141-193, 2003.

[6] Y. Meyer, B.B. Buchanan, F. Vignols, J.P. Reichheld. Thioredoxins and glutaredoxins: Unifying elements in redox biology. Annu Rev Genet, 43: 335-367, 2009

[7] A. Holmgren, C. Johansson, C. Berndt, M.E. Lönn, C. Hudemann, C.H. Lillig. Thiol redox control via thioredoxin and glutaredoxin systems. Biochem Soc Trans, 33(Pt 6):1375-7, 2005

[8] M. Valko, D. Leibfritz, J. Moncol, M.T. Cronin, M. Mazur, J. Telser. Free radicals and antioxidants in normal physiological functions and human disease. Int J Biochem Cell Biol, 39(1): 44-84, 2007

[9] D. Mustacich, G. Powis. Thioredoxin reductase. Biochem J, 346 Pt 1:1-8, 2000

[10] T. Hisabori, S.Hara, T. Fujii, D. Yamazaki, N. Hosoya-Matsuda, K. Motohashi. Thioredoxin affinity chromatography: a useful method for further understanding the thioredoxin network. J Exp Bot, 56(416):1463-8, 2005

[11] N. Rouhier, A. Villarejo, M. Srivastava, E. Gelhaye, O. Keech, M. Droux, I. Finkemeier, G. Samuelsson, KJ. Dietz, JP. Jacquot G., Wingsle. Identification of plant glutaredoxin targets. Antioxid. Redox Signal. 7(7-8): 919-929, 2005.

[12] H. Yano, J.H. Wong, Y.M. Lee, M.J. Cho, B.B. Buchanan. A strategy for the identification of proteins targeted by thioredoxin. Proc Natl Acad Sci USA, 98(8):4794-9, 2001

[13] E. Ströher, K.J. Dietz. The dynamic thiol-disulphide redox proteome of the Arabidopsis thaliana chloroplast as revealed by differential electrophoretic mobility. Physiol Plant. 133(3): 566-583, 2008.

[14] S.M. Marino, V.N. Gladyshev. Structure-based approach for detection of thiol Oxidoreductases and their catalytic redox-active cysteine residues. PLoS Comput Biol, 5(5):e1000383, 2009

[15] Y. Balmer, W.H. Vensel, C.K. Tanake, W.J. Hurkman, E. Gelhaye, N. Rouhier, J-P. Jacquot, W. Manieri, P. Schürmann, M. Droux, B.B. Buchanan. Thioredoxin links redox to the regulation of fundamental processes of plant mitochondria. Proc Natl Acad Sci USA 101:2642-2647, 2004

[16] R. Sanchez, M. Riddle, J. Woo, J. Momand. Prediction of reversibly oxidized protein cysteine thiols using protein structure properties. Protein Science, 17:473-481, 2008

[17] H. Ogata, S. Goto, K. Sato, W. Fujibuchi, H. Bono, M. Kanehisa. KEGG:Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res, 27(1):29-34, 1999

[18] G. Joshi-Tope, M. Gillespie, I. Vastrik, P. D'Eustachio, E. Schmidt, B. de Bono, B. Jasal, G.R. Gopinath, G.R. Wu, L. Matthews. Reactome: a knowledgebase of biological pathways. Nucleic Acids Res, 33(Database issue):D428-32, 2005

[19] R. Caspi, H. Foerster, C.A. Fulcher, P. Kaipa, M. Krummenacker, M. Latendresse, S. Paley, S.Y. Rhee, A.G. Shearer, C. Tissier, T.C. Walk, P. Zhang, P.D. Karp. The MetaCyc Database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. Nucleic Acids Res, 36(Database issue):D623-31, 2008

[20] I. Schomburg, A. Chang, O. Hofmann, C. Ebeling, F. Ehrentreich, D. Schomburg. BRENDA: a resource for enzyme data and metabolic information. Trends Bioche Sci, 27(1):54-6, 2002

[21] P. Shannon, A. Markiel, O. Ozier, N.S. Baliga, J.T. Wang, D. Ramage, N. Amin, B. Schwikowski, T. Ideker. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res, 13(11):2498-504, 2003

[22] Z. Hu, J. Mellor, J. Wu, C. DeLisi. VisANT: an online visualization and analysis tool for biological interaction data. BMC Bioinformatics, 5:17, 2004

[23] B.H. Junker, C. Klukas, F. Schreiber. Vanted: a system for advance data analysis and visualization in the context of biological networks. BMC Bioinformatics, 7:109, 2006

[24] T. Schwede, J. Kopp, N. Guex, M.C. Peitsch. SWISS-MODEL: An automated protein homology-modeling server. Nucleic Acids Res, 31(13):3381-5, 2003

[25] S. Hubbard, J. Thornton. Naccess. http://www.bioinf.manchester.ac.uk/naccess/

[26] H. Li, A.D. Robertson, J.H. Jensen. Very fast empirical prediction and rationalization of protein pKa values. Proteins, 61(4):704-21, 2005

[27] C.H. Wu, H. Huang, A. Nikolskaya, Z. Hu, W.C. Barker. The iProClass integrated database for protein functional analysis. Comput Biol Chem, 28: 87-96 (2004)

[28] D.E. Fomenko, S.M. Marino, V.N. Gladyshev. Functional diversity of cysteine residues in proteins and unique features of catalytic redox- active cysteines in thiol oxidoreductases. Mol Cells, 26(3):228-35, 2008

[29] T.S.K. Prasad *et al.* Human Protein Reference Database – 2009 update. *Nucleic Acids Res*, 37(Database issue):D767-72, 2009