Nucleic Acids Research, 2005, Vol. 33, Database issue ***D633–D636***
doi:10.1093/nar/gki052

Arabidopsis nucleolar protein database (AtNoPDB)

John W. S. Brown*, Peter J. Shaw[2], Paul Shaw[1] and David F. Marshall[1]

Gene Expression Programme and [1]Computational Biology Programme, Scottish Crop Research Institute, Invergowrie, Dundee DD2 5DA, Scotland, UK and [2]John Innes Centre, Norwich NR4 7UH, UK

Received August 13, 2004; Revised and Accepted October 1, 2004

ABSTRACT

The *Arabidopsis* Nucleolar Protein Database (http://bioinf.scri.sari.ac.uk/cgi-bin/atnopdb/home) provides information on 217 proteins identified in a proteomic analysis of nucleoli isolated from *Arabidopsis* cell culture. The database is organized on the basis of the *Arabidopsis* gene identifier number. The information provided includes protein description, protein class, whether or not the plant protein has a homologue in the most recent human nucleolar proteome and the results of reciprocal BLAST analysis of the human proteome. In addition, for one-third of the 217 *Arabidopsis* nucleolar proteins, localization images are available from analysis of full-length cDNA–green fluorescent protein (GFP) fusions and the strength of signal in different parts of the cell—nucleolus, nucleolus-associated structures, nucleoplasm, nuclear bodies and extra-nuclear—is provided. For each protein, the most likely human and yeast orthologues, where identifiable through BLASTX analysis, are given with links to relevant information sources.

INTRODUCTION

The nucleolus is the most prominent sub-structure of the nucleus. Its main function lies in transcription of ribosomal RNA (rRNA) gene units, processing and modification of precursor rRNA (pre-rRNA) and ribosomal subunit assembly (1). These processes require a large number of protein and small nucleolar RNA (snoRNA) components. Some snoRNAs are involved in cleavage of pre-rRNAs to generate the 18S, 25S and 5.8S rRNAs, while the majority are required for 2′-*O*-ribose methylation or pseudouridylation of specific nucleotides (2,3). In addition, the nucleolus has been implicated in a variety of other functions, including biogenesis or transport of a range of RNAs and RNPs, mRNA maturation, cell cycle control and, very recently, stress responses (4–7). Thus, the nucleolus is a complex and multifunctional component of the nucleus.

The structures of plant and mammalian nucleoli show some significant differences (8). When observed with the help of the transmission electron microscope, the mammalian nucleolus often shows three different regions in nucleoli: small, lightly staining structures called fibrillar centres (FC), surrounded by areas of densely stained material termed the dense fibrillar component (DFC), further surrounded by a particulate region called the granular component (GC). In contrast, in plant nucleoli, the DFC is less densely stained and occupies a much larger fraction (up to 70%) of the nucleolus. In addition, many plant nucleoli contain a central region called the nucleolar cavity, whose function is as yet unknown (9).

The purification of cellular structures, such as nuclear domains or bodies, and the determination of their protein components provide information on possible functions and dynamic interactions occurring in these domains. In addition, the localization of proteins to these domains may reflect interactions of components, assembly pathways of complexes or sequestration of components or complexes. Proteomic approaches have recently been applied to purified nucleoli in human [(10,11); A. I. Lamond and M. Mann, unpublished data] and *Arabidopsis* (P. J. Shaw and J. W. S. Brown, unpublished data). In the most recent study, around 700 proteins were identified in the human nucleolus. These studies have demonstrated the variety of the nucleolar protein complement possibly reflecting the range of functions in which the nucleolus may be involved. In the *Arabidopsis* nucleolar preparation, 217 proteins have been identified so far. Many proteins were known nucleolar proteins or proteins involved in ribosome biogenesis. As in the human analyses, the presence of some proteins, such as spliceosomal and snRNP proteins, and translation factors, was unexpected. In addition, proteins of unknown function which were either plant-specific or conserved between the human and plant nucleolar proteomes were identified. Finally, some plant proteins with human homologues were present in the plant nucleolar proteome but absent in that of human, suggesting differential localization or association with the nucleolus or differences in protein abundance in the nucleolus. The *Arabidopsis* Nucleolar

*To whom correspondence should be addressed. Tel: +44 1382 568533; Fax: +44 1382 562426; Email: jbrown@scri.sari.ac.uk

The online version of this article has been published under an open access model. Users are entitled to use, reproduce, disseminate, or display the open access version of this article for non-commercial purposes provided that: the original authorship is properly and fully attributed; the Journal and Oxford University Press are attributed as the original place of publication with the correct citation details given; if an article is subsequently reproduced or disseminated not in its entirety but only in part or as a derivative work this must be clearly indicated. For commercial re-use permissions, please contact journals.permissions@oupjournals.org.

© 2005, the authors

SCRI Bioinformatics

Arabidopsis Nucleolar Protein Database

Bioinformatics Home

Arabidopsis Nucleolar Protein Database - Table of Nucleolar Proteins

Table Key
Localisation = **NO/no** = nucleolar; **NU/nu** = nuclear; **NAS/nas** = nucleolus-associated structures; **NB/nb** = nuclear bodies; **EXN/exn** = extranuclear. Capitals represent strong labelling; lower case represents weak labelling. **nd** = not determined.
Human 692 proteins = The presence/absence of a homologue in the human nucleolar proteome dataset (692 proteins).

Click on locus name to retrieve further information.

Locus	Arabidopsis Gene Descriptor	Protein Class	Localisation	Image Present	Human 692 proteins	Reciprocal BLAST
At1g02140	Mago Nashi-related protein	Exon junction complex	no:NU			●
At1g02780	60S ribosomal protein L19 (RPL19A)	Ribosomal protein	nd		●	●
At1g04510	Transducin/WD-40 repeat protein	Splicing factor	NU:NAS	●	●	●
At1g07920	Elongation factor 1-alpha (EF-1-alpha)	Translation	nd		●	●
At1g09590	60S ribosomal protein L21 (RPL21A)	Ribosomal protein	NO:NB	●	●	●
At1g09760	U2 small nuclear ribonuclearprotein A	snRNP protein	nd		●	●
At1g13440	Glyceraldehyde-3-phosphate dehydrogenase-related		no:EXN:NU		●	●
At1g14320	60S ribosomal protein L10 (RPL10A)	Ribosomal protein	nd		●	●
At1g14850	Nucleoporin 155k	Nuclear pore complex	nd			●
At1g16610	Arginine/serine-rich protein	Exon junction complex	NU:no	●		●
At1g18450	Actin-related protein 4 (ARP4)	Cell structure	nd		●	●
At1g18850	Expressed protein	Unknown - Plant-specific	nd			
At1g20580	Expressed protein	snRNP protein	nd		●	●
At1g20960	U5snRNP helicase	DEAX Protein	NO:NU:exn	●	●	●
At1g21690	Replication factor-related	DNA interacting protein	nd		●	●
At1g23290	60S ribosomal protein L27A (RPL27aB)	Ribosomal protein	nb:NO		●	●
At1g24310	Expressed protein	Nuclear pore complex	nu:no			●
At1g26880	60S ribosomal protein L34 (RPL34A)	Ribosomal protein	NU	●	●	●
At1g33120	60S ribosomal protein L9 (RPL9+B92A)	Ribosomal protein	NO:nu	●	●	●
At1g43170	60S ribosomal protein L3 (RPL3A)	Ribosomal protein	NO	●	●	●
At1g45000	26S proteasome regulatory particle triple-A ATPase subunit4-related	Protein turnover	NB	●		●
At1g48620	Expressed protein	Unknown - plant-specific - DNA binding	nd			
At1g48850	Chorismate synthase	Organellar (chloroplast)	nd			
At1g48920	Nucleolin, putative	Nucleolar protein	nd		●	●
At1g51060	Histone H2A, putative	DNA interacting protein	no:NU		●	●
At1g51510	Y14	Exon junction complex	nu:NO:NB			●
At1g54060	Expressed protein	Unknown - Plant-specific	NAS:NO	●		
At1g54690	Histone H2A	DNA interacting protein	nd		●	●
At1g54850	Expressed protein	Unknown - Plant-specific	no:NU:exn			
At1g55310	Arginine-serine rich protein	Splicing factor/SR protein	nd			●
At1g56070	Elongation factor-related	Translation	nd		●	●
At1g56110	Nucleolar protein Nop56, putative	snoRNP protein (C/D)	nd		●	●
At1g60850	RNA polymerase subunit	Transcription	nd		●	●
At1g61730	Expressed protein	Unknown - Plant-specific	nd			
At1g65030	transducin/WD40 repeat protein (5 repeats)	Unknown - At/Hs nucleolar protein	nd		●	●
At1g67430	60S ribosomal protein L17 (RPL17B)	Ribosomal protein	nd		●	●
At1g68790	Nuclear matrix constituent protein 1 (NMCP1)	Nuclear envelope - plant-specific	nd			
At1g74060	60S ribosomal protein L6 (RPL6B)	Ribosomal protein	nd		●	●
At1g76300	Small nuclear ribonucleoprotein-related	snRNP protein	NAS:nu:no	●	●	●
At1g77180	Expressed protein	Transcriptional coactivator	NB:NAS:no:NU	●	●	●
At1g79280	Expressed protein	RNA interacting - nuclear pore complex	nd		●	●
At1g80750	60S ribosomal protein L7 (RPL7A)	Ribosomal protein	nd		●	●
At2g03510	Expressed protein	Unknown - only in plant No	nd			●
At2g04390	40S ribosomal protein S17 (RPS17A)	Ribosomal protein	NO	●	●	●
At2g07698	Expressed protein	Organellar (mitochondrial)	nd		●	●
At2g16360	40S ribosomal protein S25 (RPS25A)	Ribosomal protein	nd		●	●
At2g18020	60S ribosomal protein L8 (RPL8A)	Ribosomal protein	nd		●	●
At2g18510	Pre-mRNA splicing factor SF3b	Splicing factor	nd			●
At2g19520	WD-40 repeat protein (MSI4)	DNA interacting protein	NU:no:NAS	●	●	●

Figure 1. Screenshot of part of *Arabidopsis* nucleolar protein table showing At number, gene descriptor, protein class, localization information, image availability, presence/absence of protein orthologues in the human nucleolar dataset and reciprocal BLAST results.

Protein database (AtNoPDB) (http://bioinf.scri.sari.ac.uk/cgi-bin/atnopdb/home) is a MySQL/Perl/Apache informatics resource, which provides information on the plant proteins identified to date together with comparisons to orthologous human and yeast proteins, and images of cellular localizations for over one-third of the proteins. The database will continue to expand as new proteins are identified.

CONTENT OF THE DATABASE

The database currently contains information on 217 *Arabidopsis* proteins identified in a proteomic analysis of nucleoli isolated from *Arabidopsis* cell cultures. The entry point to the database is through a number of topics on the Home page. The main data topic is '*Arabidopsis* nucleolar proteins' that presents a table listing the 217 proteins arranged by chromosome on the basis of the *Arabidopsis* gene identifier numbers (see the table screenshot in Figure 1). This table also contains the gene descriptor and protein class. The localization of over one-third of the proteins has been determined by expressing full-length cDNA–green fluorescent protein (GFP) fusions in *Arabidopsis* culture cells. The localization patterns are described as nucleolar (NO/no), nucleolus-associated structures (NAS/nas), nucleoplasm (NU/nu), nuclear bodies (NB/nb) or extra-nuclear (EXN/exn) or combinations thereof, where upper and lower case letters indicate strong and weak labeling, respectively. The term 'nucleolar-associated structures' describes labelling of sub-regions of the nucleolus or cap-like regions closely associated with the nucleolus: the nature and function of these structures is currently unknown. The plant proteins have been compared with the most recent list of 692 human nucleolar proteins and the presence of a homologue in the human nucleolar proteome is indicated. Finally, the *Arabidopsis* protein sequences have been compared with human proteins using BLAST (12) and the top human hit has been again compared with *Arabidopsis*. In the majority of cases, the original *Arabidopsis* protein or a closely related protein was obtained in the reciprocal BLAST as indicated in the table.

From the main table, clicking on the *Arabidopsis* locus number gives access to an individual page for each protein/gene. Where an image of GFP fusion protein localization is available (as indicated by the green dot in the master table), the image is presented here along with a description of the labelling pattern. Information on the *Arabidopsis* gene/protein is obtained via links to additional information resources in the Arabidopsis Information Resource (TAIR), the Munich Information Centre for Protein Sequences (MIPS) and Entrez. In addition, the top BLAST result from the comparison to the human nucleolar dataset is provided with a link to the human nucleolar protein database via the IPI number. Finally, links to information sources on human and yeast homologues are provided via Entrez and the *Saccharomyces* Genome Database (SGD).

The distribution of the 217 proteins by protein class is available as is the complete library of GFP fluorescence images. Details of the comparison between the *Arabidopsis* and human 692 nucleolar protein datasets are provided. Finally, a list of relevant publications and links to relevant databases are given along with search capabilities on the basis of AGI number, gene description and protein class, and information on feedback and submission of data to AtNoPDB.

DATABASE ACCESS AND FUTURE OF DATABASE

The database provides an interface to comparative proteomic information for each of the *Arabidopsis* nucleolar proteins so far identified. As more proteins are identified, these will be added, ultimately providing a dataset which will allow a full comparison with the human nucleolar proteome. We are currently undertaking a comprehensive comparison of the plant and mammalian nucleolar proteomes based on a combined approach of alignment, structure and phylogeny. This comparative information together with protein motifs and structure will be integrated into the database at a later date. Information of the multigene family organization of many of the *Arabidopsis* proteins and comparative data from homologous proteins in other plant species will also be added.

LINKS

Links are provided to other information sources (TAIR, MIPS, etc.) as detailed above, to the current human nucleolar proteome database of 271 proteins (10), the Plant snoRNA database and links will be established to the new human nucleolar protein database which is currently being developed. We have established a collaboration with Dr Rebecca Ernest at MIPS through which we will provide a BioMOBY (http://biomoby.org/) based webservice integration with the PLANET consortium of *Arabidopsis* information resources (http://mips.gsf.de/proj/planet/).

ACKNOWLEDGEMENTS

This research was supported by funding from the Scottish Executive Environment and Rural Affairs Department (SEERAD) to SCRI and the Biotechnology and Biological Sciences Research Council of the UK (BBSRC) to the John Innes Centre.

REFERENCES

1. Venema,J. and Tollervey,D. (1999) Ribosome synthesis in S*accharomyces cerevisiae*. *Annu. Rev. Genet*., **33**, 261–311.
2. Filipowicz,W. and Pogačić,V. (2002) Biogenesis of small nucleolar ribonucleoproteins. *Curr. Opin. Cell Biol*., **14**, 319–327.
3. Kiss,T. (2002) Small nucleolar RNAs: an abundant group of non-coding RNAs with diverse cellular functions. *Cell*, **109**, 145–148.
4. Pederson,T. (1998) The plurifunctional nucleolus. *Nucleic Acids Res*., **26**, 3871–3876.
5. Carmo-Fonseca,M., Mendes-Soares,L. and Campos,I. (2000) To be or not to be in the nucleolus. *Nature Cell Biol*., **2**, E107–E112.
6. Olson,M.O.J., Dundr,M. and Szebeni,A. (2000) The nucleolus: an old factory with unexpected capabilities. *Trends Cell Biol*., **10**, 189–196.
7. Rubbi,C.P. and Milner,J. (2003) Disruption of the nucleolus mediates stabilisation of p53 in response to DNA damage and other stresses. *EMBO J*., **22**, 6068–6077.

8. Shaw,P.J. and Jordan,E.G. (1995) The nucleolus. *Annu. Rev. Cell Dev. Biol.*, **11**, 93–121.
9. Brown,J.W.S. and Shaw,P.J. (1998) Small nucleolar RNAs and pre-rRNA processing in plants. *Plant Cell*, **10**, 649–657.
10. Andersen,J.S., Lyon,C.E., Fox,A.H., Leung,A.K.L., Lam,Y.W., Steen,H., Mann,M. and Lamond,A.I. (2002) Directed proteomic analysis of the human nucleolus. *Curr. Biol.*, **12**, 1–11.
11. Scherl,A., Couté,Y., Déon,C., Callé,A., Kindbeiter,K., Sanchez,J.-C., Greco,A., Hochstrasse,D. and Diaz,J.-J. (2002) Functional proteomic analysis of human nucleolus. *Mol. Biol. Cell*, **13**, 4100–4109.
12. Altschul,S.F., Madden,T.L., Schäffer,A.A., Zhang,J., Zhang,Z., Miller,W. and Lipman,D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.