Arabidopsis nucleolar protein database (AtNoPDB)

John W. S. Brown*, Peter J. Shaw2, Paul Shaw1 and David F. Marshall1

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

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

ABSTRACT

The Arabidopsis Nucleolar Protein Database (http://biolinf.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 nucleolar RNA (snoRNA) components. Some snoRNAs are involved in ribosome biogenesis or association with the nucleolus or differences in protein abundance in the nucleolus. The structures of plant and mammalian nucleoli show some significant differences (2). 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 particular 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

Nucleic Acids Research, Vol. 33, Database issue © Oxford University Press 2005; all rights reserved
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/attnopdb/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 labeling 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 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