ELSEVIER

Journal of Cereal Science 41 (2005) 165–179

Journal of
CEREAL
SCIENCE

www.elsevier.com/locate/jnlabr/yjcrs

Review

Proteomics of wheat grain

D.J. Skylas[a,b,*], D. Van Dyk[c], C.W. Wrigley[b,d]

[a]*Australian Proteome Analysis Facility, Macquarie University, Level 4, Building F7B, Sydney, NSW 2109, Australia*
[b]*Value Added Wheat CRC Ltd, North Ryde, NSW 1670, Australia*
[c]*Minomic Pty, Ltd, Frenchs Forest, Sydney, NSW 2086, Australia*
[d]*Food Science Australia, North Ryde, NSW 1670, Australia*

Received 6 July 2004; revised 4 August 2004; accepted 16 August 2004

Abstract

The central dogma of molecular biology describes the flow of genetic information from DNA to RNA and to proteins in living biological systems. Newly emerging technologies are being applied, and continually developed, to elucidate interactions between these biomolecules at all stages during the flow of genetic information in biological systems, and in relation to specific conditions (the growth conditions of plants). These newly emerging technologies encompass genomics, transcriptomics and proteomics, as well as the rapidly expanding and exciting field of bioinformatic tools and interactive databases.

With the recent completion of the sequence of the genome of the 'model' plant *Arabidopsis thaliana*, a basis has been provided for the analysis of gene function in plants, which will no doubt have an impact on cereal plants as well. The importance of this 'model' genome project is enormous, as many important cereal crops, such as wheat, maize and rice, have large genomes and in some cases such as wheat are also polyploid, with related genes present on the different genomes. This may provide problems for the efficient and economical attempts to completely sequence these genomes, in the near future.

With the combination of these newly emerging technologies, the stage is now set for cereal chemistry to capitalise on advances being made widely in protein chemistry, to apply these new methods, and thereby bridge the traditional gap between DNA and proteins, between the genome and proteome. In doing so, we stand to learn more about the inheritance of grain-quality attributes, and also, possibly more importantly, to discover more about the effect of growth and storage conditions on grain quality, and their effects on processing. In this review, the main aspects of proteomics are discussed, as well as the current and future applications of proteomic technologies to cereal grain science.

Keywords: Grain quality; Wheat; Barley; Endosperm proteins; alpha-Amylase inhibitors; Puroindolines; Protein disulfide isomerase; Heat-shock proteins; Cultivar identification

1. Introduction

The central dogma of molecular biology for all living biological systems describes the flow of genetic information from DNA to mRNA, and on to the synthesis of polypeptides, which in turn are assembled into active protein molecules. Technologies are continually being developed and applied, to elucidate interactions between biological molecules, at all stages during the flow of genetic information in biological systems. These newly emerging technologies encompass 'genomics' (DNA), 'transcript-transcriptomics' (mRNA) and 'proteomics' (proteins) as well as the rapidly expanding and exciting field of

Abbreviations: 2-DE, two-dimensional gel electrophoresis; CID, collision-induced dissociation; DPA, days post-anthesis; ESI, electrospray ionisation; EST, expressed sequence tag; ICAT, isotope coded affinity tagging; IEF, isoelectric-focussing; IPG, immobilised pH gradient; LC, liquid chromatography; MALDI, matrix-assisted laser desorption/ionisation; MS, mass spectrometry; MS/MS, tandem mass spectrometry; MudPIT, multi dimensional protein identification technology; PDI, protein disulfide isomerase; p*I*, isoelectric point; PMF, peptide-mass fingerprinting; PSD, post-source decay; PVDF, polyvinylidene difluoride; SCX, strong cation exchange; SDS-PAGE, sodium dodecyl sulfate polyacrylamide gel electrophoresis; TOF, time-of-flight.

* Corresponding author. Tel.: +61 2 9850 6217; fax: +61 2 9850 6200.
E-mail address: dskylas@proteome.org.au (D.J. Skylas).

0733-5210/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jcs.2004.08.010

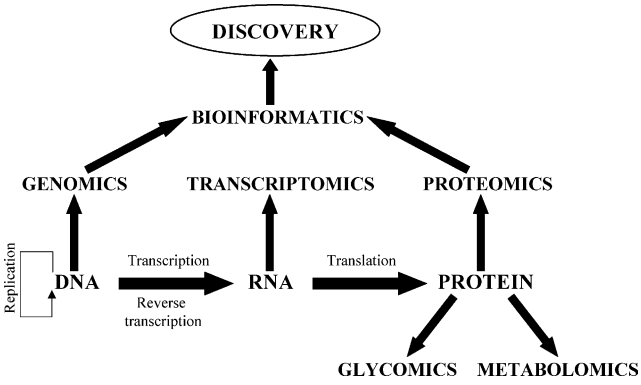

Fig. 1. Emerging technologies for the elucidation of genome and proteome interactions.

'bioinformatics', which provides complementary tools and interactive sequence databases (Fig. 1). Subsequently, the use of the '-omics' suffix has been extended further to cover several other aspects of the constitution of individual organisms that derive from its genome; including 'glyco-glycomics' (carbohydrates; Feizi et al., 2003; Hirabayashi and Kasai, 2002; Packer and Harrison, 1998) and 'metabo-metabolomics' (metabolites; Fiehn, 2002; Sumner et al., 2003; Weckwerth, 2003).

2. From genomics to proteomics

2.1. Model plant systems

Genome initiatives aim to sequence the whole genomes of organisms, to deduce their complete nucleotide sequences. The first genome completely sequenced was that of bacteriophage Lambda (Sanger et al., 1982). With increasing advances in DNA sequencing and related software-analysis technologies, larger genome initiatives have commenced and others have already been completed, including numerous prokaryotic microbes, as well as eukaryotes such as the model plant species *Arabidopsis thaliana* (The Arabidopsis Genome Initiative, 2000).

A. thaliana was selected as a plant 'model' due to the small size of its genome (approximately 120 Mb), comprising only five chromosomes, as well as its short lifecycle and extensive seed production. It is also related to major brassica crops and therefore provides a valuable model for these species. The major outcome of this sequencing initiative was the discovery of approximately 26,000 genes, of which, about 15,000 appear to be different. Research into gene function has only been pursued for a relatively small number of these genes. The genes were automatically classified according to their similarity to known protein sequences present in databases. Putative functions could be assigned for the majority of the genes analysed, with the remainder still unclassified. The majority of genes classified have putative functions in metabolism, gene regulation and plant-defence mechanisms (The Arabidopsis Genome Initiative, 2000).

Table 1
Genome size for some commercially important cereal crops (adapted from The Arabidopsis Genome Initiative, 2000)

Cereal	Genome size (Mb)
Rice (*Oryza sativa*)	420
Maize/corn (*Zea mays*)	2500
Barley (*Hordeum vulgare*)	4800
Wheat (*Triticum aestivum*)	16,000

Attention has now shifted towards functional genomic studies of *A. thaliana* with the aim of identifying and characterising all proteins expressed by the genome (that is, its proteome), with this information being made accessible on databases. This is a publicly funded project, which has recently commenced under the title '*Arabidopsis* 2010'.

2.2. Cereal genome initiatives

With the completion of the genome sequence of *A. thaliana*, many international plant scientists are focusing on sequencing the genomes of commercially important crops, particularly rice (Adam, 2000). Many cereal genomes are very large, with significant gene redundancy, which for some cereals is a result of polyploidy and local gene duplications. The genome sizes of a range of commercially important grain crops are listed in Table 1. The range is approximately 40-fold, from the genome of rice (420 Mb) to that of wheat (16,000 Mb). The relatively small genome size of rice and its importance as a food grain, particularly in developing countries, have made rice a primary target for genome sequencing. The International Rice Genome Sequencing Project (IRGSP) (Adam, 2000; http://rgp.dna.affrc.go.jp/IRGSP/index.html) involves 10 member countries: Japan, the United States of America, China, Taiwan, Korea, India, Thailand, France, Brazil, and the United Kingdom. Already, the IRGSP has published some hundreds of mega-bases of non-overlapping nucleotide sequences of the rice genome in public databases.

Genome initiatives such as these will contribute enormously to our understanding of plants in general, and to quality-related attributes of grains specifically. However, the existence of an open reading frame in a genome does not necessarily imply the existence of a functional gene, nor can its role in the plant be clearly understood from the genome sequence. Therefore, the first stage in annotating any genome is to verify the range of gene-products, namely, the polypeptides synthesised in a specific situation (that is, its proteome).

3. Proteomics and grain science

Characterising gene expression in organisms usually requires the analysis of either the mRNA or protein.