Cereal genes similar to Snf2 define a new subfamily that includes human and mouse genes

Abstract Genes from the SNF2 family play important roles in transcriptional regulation, maintenance of chromosome integrity and DNA repair. This study describes the molecular cloning and characterization of cereal genes from this family. The predicted proteins exhibit a novel C-terminal domain that defines a new subfamily designated SNF2P that includes human and mouse proteins. Comparison between genomic and cDNA sequences showed that cereal Snf2P genes consisted of 17 exons, including one only 8 bp long. Two barley alleles differed by the presence of a 7.7-kb non-LTR retrotransposon in intron 6. An alternative annotation of the orthologous Arabidopsis gene would improve its similarity with the other members of the subfamily. Intron 2 was not spliced out in approximately half of the rice Snf2P mRNAs present in leaves, resulting in a premature stop codon. Transcripts from the barley and wheat Snf2P genes were found in apexes, leaves, sheaths, roots and spikes. The Snf2P genes exist as single copies on wheat chromosome arm 5A\textsuperscript{mL} and in the colinear regions on barley chromosome arm 4HL and rice chromosome 3. High-density genetic mapping and RT-PCR suggest that Snf2P is not a candidate gene for the tightly linked vernalization gene Vrn2.

Keywords SNF2 family · Alternative splicing · Comparative genomics · Wheat · Vernalization gene Vrn2

Introduction

Members of the SNF2 family of proteins share the SNF2 domain, defined by the presence of seven motifs with sequence similarity to DNA helicases (Coleman et al. 2000). Despite the presence of these helicase-related motifs, none of the proteins from this family has yet been demonstrated to have a helicase function. In general, members of the SNF2 family show the capacity to use the energy released by their DNA-dependent ATPase activity to stabilize or perturb protein-DNA interactions (Pazin and Kadonaga 1997; Muthuswami et al. 2000).

Among the best-characterized members of this family are the genes included in the SNF2 subfamily. Recent whole-genome expression studies have shown that genes from the SNF2 subfamily control the transcription of about 6% of all genes in Saccharomyces cerevisiae (Kingston et al. 1996; Sudarsanam et al. 2000), and their products are required in vivo for the establishment of a transcriptionally active chromatin structure (Gavin and Simpson 1997; Wu and Winston 1997). These genome-wide expression analyses also suggested that control was exerted at the level of individual promoters rather than over chromosomal domains (Sudarsanam et al. 2000). Genetic and biochemical studies of SNF2 and related proteins...
indicated that they can destabilize nucleosome structure and thereby facilitate the binding of transcription factors to chromatin in promoter regions (Kingston et al. 1996).

Whole-genome mRNA expression studies suggested that genes from the SNF2 family could also act as transcriptional repressors. Almost half (25–40%) of the genes affected in defective snf2 mutants in yeast express increased levels of mRNA. Another example is the *Drosophila* *domino* gene, which contributes to the silencing of homeotic genes (Ruhf et al. 2001).

Proteins from the SNF2 family have been classified into 13 subfamilies with a range of biological functions, including transcriptional regulation, maintenance of chromosome integrity, DNA repair, and maintenance of genomic methylation (Eisen et al. 1995; Jeddeloh et al. 1999; Coleman et al. 2000). Most of the proteins included in the studies by Coleman et al. (2000) were from bacteria, yeast, *Drosophila*, mouse, and human; they looked at no examples from plant species. Recently, numerous proteins related to members of the different SNF2 subfamilies have been predicted from the genome sequence of *Arabidopsis thaliana* (Lin et al. 1999; Kaneko et al. 2000; Salanoubat et al. 2000), but no formal classification was attempted.

We found sequences homologous to one of the Snf2-related Arabidopsis genes (GenBank AC007659_19) in diploid wheat (*Triticum monococcum*, L., hereafter referred to as wheat), barley and rice during a chromosome walk toward the vernalization gene *Vrn2* of wheat (Dubcovsky et al. 1998; Dubcovsky 2001; L. Yan and J. Dubcovsky, unpublished results). *Vrn2* is involved in the repression of flowering in barley and diploid wheat (Tranquilli and Dubcovsky 1999), and is located on the distal part of the long arm of homoeologous group 5 (in a segment translocated to chromosome 5A in both diploid and polyploid wheats; Dubcovsky et al. 1998). This gene is different from *Vrn-B1*, which was known as *Vrn2* in the old *Vrn* classification (McIntosh et al. 1998).

The involvement of different members of the SNF2 family in transcriptional repression (Ruhf et al. 2001) and in the maintenance of genomic methylation (Jeddeloh et al. 1999), and the absence of recombination in a mapping population of 150 gametes, suggested that the Snf2-related gene might be a good candidate for the *Vrn2* gene. Sequences from complete cDNAs from barley and rice were used to determine the gene structure in cereals and to compare genes from winter and spring barley varieties. Expression profiles and high-density genetic maps are presented to show that this *Snf2*-related gene in Triticeae is not a strong candidate gene for the vernalization gene *Vrn2*. In spite of this negative result, valuable information was obtained on the structure and expression of the *Snf2*-related genes in these experiments. We also show that the proteins coded by these plant genes, together with proteins from human and mouse, define a new subfamily within the SNF2 family.

### Materials and methods

**Sequence alignment and nomenclature**

The 350 amino acids extending from the C-terminus of the Arabidopsis protein (GenBank AC007659_19) and the related proteins identified in this study were used to search the non-redundant database using BLASTP (Altschul et al. 1997). Human protein AAH01171 and mouse protein NP_080815 were the only ones to show a significantly similar C-terminal domain, and were aligned with the proteins encoded by the plant genes using the multiple sequence alignment program ClustalW1.8 (Thompson et al. 1994). The amino acid sequence alignment figure was produced with Boxshade 3.21 (http://www.ch.embnet.org/software/BOX_form.html).

BLASTP searches using the conserved SNF2 domain showed that the most closely related proteins belong to the *SNF2, SNF2L, CHD1* and *YFK8* subfamilies. A phylogenetic study was performed using the six SNF2-related proteins listed above plus three members of each of the four most related SNF2 subfamilies and one member of the other subfamilies as defined by Coleman et al (2000). The accession Nos. for the members of the SNF2 subfamily included in this study were S66910 (human), M61703 (yeast), and NP_187252 (Arabidopsis); for the SNF2L subfamily M89907 (human), S46122 (yeast), and NP_187291 (Arabidopsis); for the CHD1 subfamily XP_004000 (human), NP_011091 (yeast), and AAD28668 (Arabidopsis); and for the YFK8 subfamily NP_032260 (mouse), P43610 (yeast), and NP_201476 (Arabidopsis). The Arabidopsis member of the YFK8 subfamily is the *DDM1* gene, which is involved in the maintenance of genomic DNA methylation (Jeddeloh et al. 1999). Only one member each was included from the less related subfamilies *CSB* (NP012569, yeast), *RAD54* (M63232, yeast), *YSCD* (Z48618, yeast), and *MOT1* (M83224, yeast). Finally, only one member each was included from the two pairs of distantly related groups including subfamilies *HEPA1-HARP* (XP046726, human) and *HEPA2-RAD16* (M86929, yeast) (Coleman et al. 2000).

Phylogenetic trees were generated from the ClustalW sequence alignments using multiple distance- and parsimony-based methods available in the MEGA2.1 computer software package (http://www.megasoftware.net/; see Kumar et al. 1994). All sites containing alignment gaps were removed before the calculations (complete-deletion option). Distances between each pair of proteins were calculated and trees were constructed using four different methods to determine the robustness of the groups to variations in clustering techniques. The four techniques compared in this study included Neighbor-Joining (NJ), the Unweighted Pair Group Method using Averages (UPGMA), Maximum Parsimony (MP) and Minimum Evolution (ME) with the default parameters implemented in MEGA2.1. Consensus trees and confidence values for the nodes were calculated for each of the four analyses using 500 bootstrap (MEGA2.1).

BLASTP searches of the C-terminal domain and consistent clusters from the phylogenetic studies were used to define the subfamilies. Based on these criteria it will be shown that the group of proteins including the Arabidopsis AC007659_19, the cereal proteins described in this study, the human AAH01171 and mouse NP_080815 define a new subfamily that will be designated *SNF2P* throughout the text. Genes coding for the proteins included in the *SNF2P* subfamily will be indicated in lower case italics (*Snf2P*).

Selection and sequencing of BAC clones

BAC clones from wheat, barley and rice were selected with the genomic RFLP probes UCW1 (Nuecellen; Chen and Foolad 1997) and UCW2, which were previously found to be tightly linked to the vernalization gene *Vrn2* (Dubcovsky et al. 1998; L. Yan and...