Widespread occurrence of a highly conserved RING-H2 zinc finger motif in the model plant *Arabidopsis thaliana*

Rikke B. Jensen, Karin L. Jensen, Hans M. Jespersen, Karen Skriver*

Institute of Molecular Biology, University of Copenhagen, Øster Farimagsgade 2A, DK-1353 Copenhagen K, Denmark

Received 27 July 1998; revised version received 27 August 1998

Abstract Several novel *Arabidopsis thaliana* proteins containing a RING-H2 zinc finger motif were predicted after database searches. Alignment of 29 RING-H2 finger sequences shows that the motif is strikingly conserved in otherwise unrelated proteins. Only short, non-conserved polar/charged sequences distinguish these domains. The RING-H2 domain is most often present in multi-domain structures, a number of which are likely to contain a membrane-spanning region or an additional zinc finger. However, there are several small (126–200 residues) proteins consisting of an N-terminal domain, rich in aliphatic residues, and a C-terminal RING-H2 domain. Reverse-transcription PCR suggests that the RING-H2 genes are widely expressed at low levels.

Key words: Expressed sequence tag; Reverse transcription polymerase chain reaction; RING-H2; Bacterial artificial chromosome; *Arabidopsis thaliana*

1. Introduction

The RING finger, named after the Really Interesting New Gene 1 [1], was originally identified as a cysteine-rich, putative zinc chelating motif in otherwise unrelated proteins [2]. The $C_3H_2C_3$ RING-H2 variant contains a histidine ligand in place of the fourth cysteine of the classical RING finger motif [3]. RING finger proteins have diverse cellular functions but are often involved in formation of multiprotein complexes [4]. Only a few plant RING finger-containing gene products have been described. *Arabidopsis thaliana* COP1 contains a RING finger and can repress photomorphogenesis [5], while *ATL2* is an auxin-inducible, early response gene encoding a protein with a RING-H2 finger [6]. No function has been assigned to the RING-H2 Pzf proteins from *Lotus japonicus* or *Glycine max* [7] or the *Arabidopsis* RING containing gene product A-RZF [8].

Based on the above, proteins with RING fingers are likely to have important regulatory functions in plants. We have examined their diversity in *Arabidopsis*, since this will allow selection of genes and proteins for functional studies in a model plant.

*Corresponding author. Fax: (45) (35) 32 20 75.
E-mail: cinacat@biobase.dk

Abbreviations: ABRC, *Arabidopsis* Biological Resource Center; BAC, bacterial artificial chromosome; EST, expressed sequence tag; IEEHV, immediate early equine herpes virus; PML, acute promyelocytic leukaemia proto-oncoprotein; RHN, RING-H2 group N; RING, really interesting new gene zinc finger; RT, reverse transcription

2. Materials and methods

2.1. Identification and analysis of RING-H2 cDNA clones, genes and predicted proteins

Keyword searches for RING finger sequences were performed in the expressed sequence tag (EST) and GenBank databases, and identified sequences were used as search strings in tblastn [9] searches. Genomic and cDNA sequences were cross-examined with blastn [9] to identify corresponding genes and cDNAs. Genes which were not defined in the annotation of bacterial artificial chromosomes (BAC) were analyzed using NetPlantGene [10]. Proteins predicted only from genomic data were named by the accession code of the parent BAC followed by an arbitrary number (Table 1). Sequence context analysis [11] and prediction of TATA boxes were used to help define translation initiation sites. Predicted protein sequences were examined for similarity to known protein sequences by blastp [9], amino acid composition and arrangement by SAPS [12], motifs by ScanProsite [13] and COILS [14], sorting signals and membrane-spanning regions by PSORT [15], and secondary structures by PHDsec [16]. CLUSTAL X [17] was used for sequence alignments.

2.2. DNA sequencing

cDNA clones were acquired from the *Arabidopsis* Biological Resource Center (ABRC), Ohio State University. Each base in the nucleotide sequences was unambiguously determined on both strands. Sequences were analyzed with Sequencher version 3.0 (Gene Codes Corporation) and will appear in the DDBJ, EMBL, and GenBank databases (Table 1).

2.3. Reverse transcription and long-range PCR

Seeds of *Arabidopsis* (Columbia) were grown in soil in the green house. Tissues were harvested, separated and immediately frozen in liquid nitrogen. Total RNA was isolated by phenol-chloroform extraction and LiCl precipitation [18], followed by DNase treatment and thermal inactivation of the DNase. The mRNA expression pattern was analyzed using a reverse transcription (RT)-PCR kit (Perkin Elmer). RT of 0.1 μg RNA was carried out according to the manufacturer. The RT-reaction was diluted 10-fold in a 20-μl PCR reaction (Perkin Elmer), and PCR performed using 30 cycles at 95°C for 1 min and 60°C for 1 min. Primers (Fig. 4) used for the amplifications were designed to anneal to non-conserved coding or 3′ non-coding sequence regions and to span introns when possible (*RHA4a*, *RHF1a*, and *RHG1a*).

A complete *RHA4a* cDNA sequence was obtained by long-range PCR [19]. One μg DNA isolated from an *Arabidopsis* (Columbia) λ-PRL2 library from ABRC was used as template for PCR amplification with the *RHA4a* specific antisense primer 5′-GAATCAAACCGGTGACGT-3′ and a library specific primer. The reaction components were according to the Perkin Elmer XL PCR Kit employing the Hot Start technique with a lower and an upper mix initially separated by a wax bead to reduce mispriming.

3. Results and discussion

3.1. Identification of RING-H2 encoding cDNA clones and genes

Database searches identified approximately 50 different, putative RING encoding *Arabidopsis* sequences, about 40 of which were of the RING-H2 type. Analysis of the sequences revealed an overrepresentation of RING-H2 fingers with the

PII: S0014-5793(98)01143-0

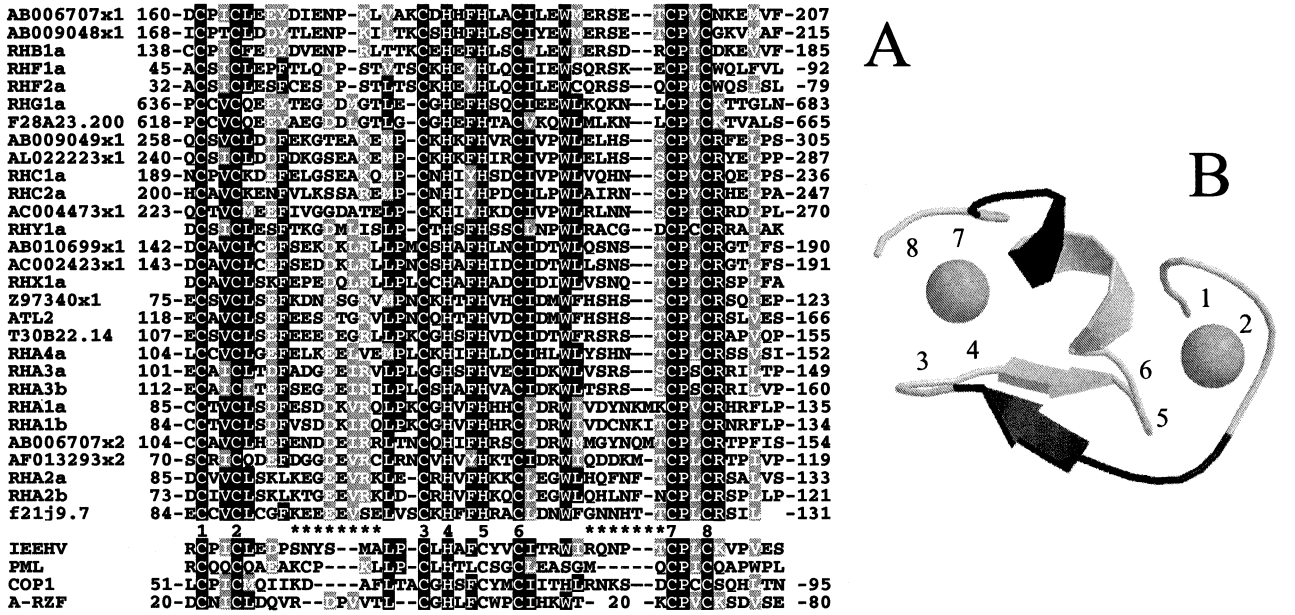

Fig. 1. Conserved and variable regions of the *Arabidopsis* RING-H2 fingers. A: Alignment of 29 *Arabidopsis* RING-H2 finger sequences. Residues conserved or chemically similar in 50% or more of the sequences are surrounded by black and grey, respectively. The zinc ligands (1–8) and variable sequence regions (asterisks) are marked. The sequences of four classical RING fingers, IEEHV [20], PML [21], COP1 [5], and A-RZF [8], are also shown. Twenty residues were omitted from the A-RZF sequence in the second variable region. B: MOLSCRIPT [27] ribbon diagram representation of the IEEHV RING finger. The zinc ions are shown as spheres. The approximate positions of the zinc ligands are labelled 1–8. The regions in IEEHV corresponding to the variable sequence regions in the *Arabidopsis* proteins, based on the alignment in A, are shown in black.

Table 1
cDNAs and genes encoding putative RING-H2 *Arabidopsis* proteins

Predicted protein	No. of ESTs	Acc. no. cDNA	Acc. no. BAC	Coding region in BAC[a]	Length (AA)	Group
RHA1a	1	AF078683[b]	–		159	A
RHA1b	3	AF078821[b]	–		157	A
RHA2a	7	AF078822[b]	–		155	A
RHA2b	3	AF078823[b]	–		147	A
RHA3a	3	AF078824[b]	AC002329	C(76 764–77 321) (0)	185	A
RHA3b	1	AF078825[b]	–		200	A
RHA4a	2	AF079178[b]	AC002343	82 029–83 339 (2)	174	A
AB006707x2	0	–	AB006707	C(50 398–50 928) (0)	176	A
AF013293x2	0	–	AF013293	C(80 458–80 838) (0)	126	A
F21J9.7	0	–	AC000103	C(27 987–28 382) (0)	131	A
RHB1a	1	AF079179[b]	AF013293	C(64 426–65 522) (3)	190	B
AB006707x1	1	–	AB006707	C(22 840–23 906) (3)	212	B
AB009048x1	0	–	AB009048	34 573–35 804 (3)	221	B
RHC1a	4	AF079180[b]	AC002409	13 438–14 424 (0)	328	C
RHC2a	1	AF079186[b]	AC003000	C(10 039–11 244) (0)	401	C
AB009049x1	1	–	AB009049	31 373–32 563 (0)	396	C
AC004473x1	0	–	AC004473	C(82 715–83 698) (0)	327	C
AL022223x1	1	–	AL022223	63 208–64 278 (0)	356	C
AB010699x1	0	–	AB010699	38 659–39 789 (0)	376	D
AC002423x1	0	–	AC002423	C(59 333–60 442)[c] (0)	369	D
ATL2	4	L76926 [5]	–		284	E
T30B22.14	0	–	AC002535	36 023–36 706 (0)	227	E
Z974340x1	0	–	Z97340	18 969–19 676 (0)	235	E
RHF1a	1	AF079181[b]	Z97335	153 936–156 164 (7)	371	F
RHF2a	6	AF079182[b]	–		375	F
RHG1a	3	AF079183[b]	AB008264	C(45 571–48 093) (4)	691	G
F28A23.200	0	–	AL021961	83 079–85 944 (4)	666	G
RHX1a	1	AF079184[b]	–		–	
RHY1a	3	AF079185[b]	–		–	

[a]First base in predicted start codon to last base in predicted stop codon. Number of introns in indicated in parentheses. C()-complement.
[b]Sequenced in this work.
[c]High throughput genome sequence, position may change.

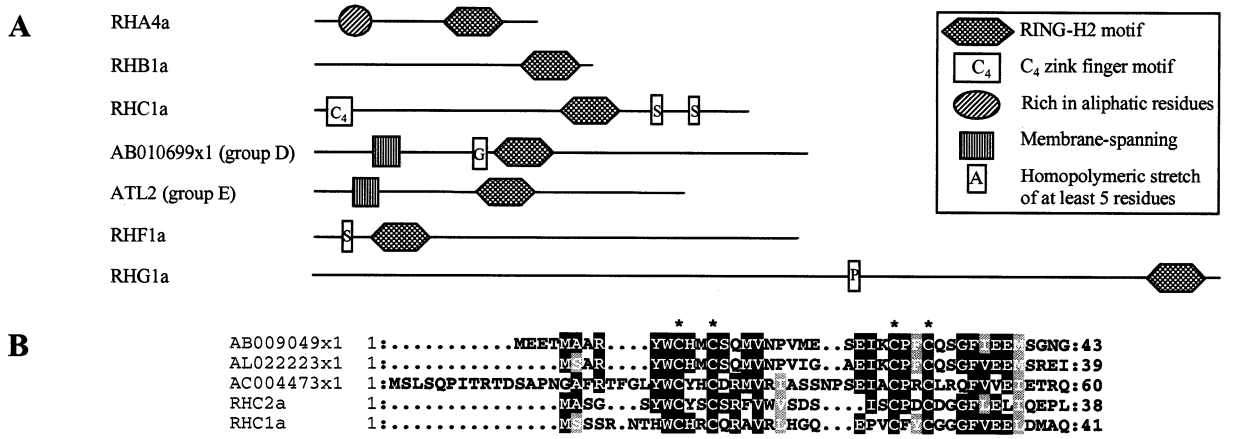

Fig. 2. Predicted *Arabidopsis* RING-H2 proteins. A: Structural outline of proteins representing the tentative groups A–G showing the relative size of the proteins and the position of the RING-H2 finger. B: Alignment of the N-terminal region of the RHC proteins with conserved cysteines (asterisks) of their putative C_4 zinc finger motif marked.

consensus $CX_2CX_{14-15}CXHX_2HX_2CX_3WX_{6-8}CPXC$, a much more stringent consensus than the general $CX_2CX_{9-39}CX_{1-3}HX_{2-3}[C,H]X_2CX_{4-48}CX_2C$ RING finger motif [4]. Twenty-nine predicted proteins containing the restricted RING-H2 consensus sequence are presented in Table 1 together with information about the cDNAs and BACs from which they are derived.

3.2. Short polar/charged regions discriminate highly conserved RING-H2 fingers

Alignment of the 29 RING-H2 finger sequences from the proteins in Table 1 demonstrates the high degree of amino acid conservation and the smaller variation in the spacing between ligand number two and three and six and seven than in the general RING finger motif (Fig. 1). Because of the similarity between our RING-H2 consensus sequence and the sequence of the immediate early equine herpes virus (IEEHV) RING finger its structure may serve as a model for the *Arabidopsis* RING-H2 fingers (Fig. 1B) [20]. Similar structural features of the IEEHV and acute promyelocytic leukaemia proto-oncoprotein (PML) RING fingers [4,21], despite the greater sequence divergence between these than between the *Arabidopsis* RING-H2 and the IEEHV fingers, further support the use of the IEEHV structural model. According to this, the *Arabidopsis* RING-H2 fingers would be built of a hydrophobic core with conserved cysteines and histidines forming a cross-brace zinc ligation system and the regions with the most diverse sequences exposed. These short, exposed, variant sequence regions are dominated by polar and charged residues (Fig. 1) which may be well-suited for macromolecular interactions.

3.3. Structural diversity of Arabidopsis RING-H2 proteins

The RING-H2 proteins reported in Table 1 were analyzed as described in Section 2. The proteins were tentatively assigned to groups named RHN for RING-H2 group N based on structural characteristics (RHA) or sequence similarity (RHB-RHG) outside the RING-H2 domain. The RING-H2 finger is localized at different positions in the proteins representing the different groups depicted in Fig. 2A. For most of the groups structural features in addition to the RING-H2 finger were recognized. Several of the RHA, and all of the

Fig. 3. Alignments of predicted RHA sequences (Table 1). Cysteines and histidines of the RING-H2 finger motif (1–8) and the leucine zipper motif in RHA4a (a) are marked. Predicted membrane-spanning regions are underlined, and α-helix (H) and β-strand (E) predictions are shown for the N-terminal region of RHA2a and RHA4a.

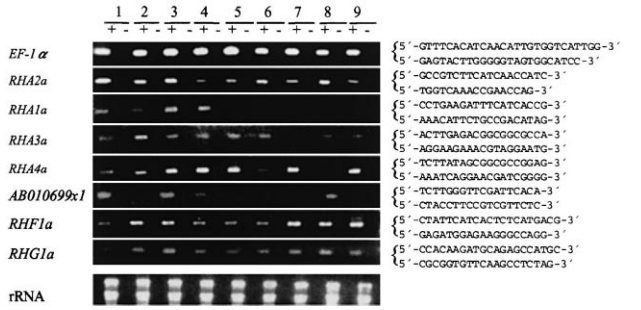

Fig. 4. Accumulation of RING-H2 and *EF-1α* mRNA. Total *Arabidopsis* RNA was tested using RT-PCR, in the presence (+) and absence (−) of RT. The following RNA templates were used: lane 1: 7-day-old etiolated seedlings; lane 2: 40-day-old light grown plants; lane 3: 7-day-old light grown plants; lane 4: stems; lane 5: flowers; lane 6: green siliques; lane 7: cauline leaves; lane 8: seeds; lane 9: roots. 'rRNA' is 10 μg of the RNA template sample stained with ethidium bromide. Amplicons fractionated through a 2% agarose gel are shown.

RHD and RHE proteins, were predicted to contain a membrane-spanning region [6,15], as discussed below. The RHC proteins contain an N-terminal, putative C_4 zinc finger motif (Fig. 2B) [22] with the consensus sequence $CX_2CX_{11-15}CX_2C$. A combination of two zinc fingers with different functions is also found in, for instance, the RING-finger B-box proteins [3]. No clear structural features were identified for the RHB, RHF, and RHG proteins, except for homopolymeric amino acid stretches present in many of the RING-H2 proteins.

3.4. RHA, a group of small RING-H2 finger proteins

The RHA proteins were grouped together because of their small size, 126–200 residues, and their structurally similar N-terminal regions (Fig. 3). The relatively simple structure of the RHA proteins provides an opportunity to elucidate the function of the separate domains and, therefore, the proteins will be described in detail.

The RHA proteins possess at least two distinct domains (Fig. 3). Only the C-terminal RING-H2 regions show pronounced sequence similarity with proteins in databases. The N-terminal regions are characterized by a high content of aliphatic residues and are predicted to form membrane-spanning, α-helical regions in several of the proteins [15,16]. Since they, except in the case of AB006707x2, are unlikely to represent cleavable signal peptides [23], a function in membrane anchoring is likely. The RHD and RHE proteins probably also function when associated with or bound to membranes as previously reported for several RING-H2 proteins [3]. Considering other possible functions for the RHA proteins, it is interesting to note that RHA4a contains a leucine zipper $LX_6LX_6LX_6L$ motif [24]. However, based on the high density of aliphatic residues in this region and a low probability of forming a coiled-coil structure [14], the motif is unlikely to form a leucine zipper.

3.5. Accumulation of the RING-H2 gene products

RT-PCR was used to examine the accumulation of selected RING-H2 gene products, mostly of group A, the proteins of which have been described in most detail (Fig. 4). RNA, representing the material used for generation of the ESTs [25], was used as template. RT-PCR with *RHA2a* specific primers resulted in a strong band of the expected size in all samples tested. In contrast, the results shown in Fig. 4 suggest a differential accumulation of mRNA encoded by the *RHA1a*, *RHA3a*, and *RHA4a* genes. The analyses also showed that differential expression of the AB10699x1 gene (group D), for which no EST clone is available, does occur. The *RHF1a* (group F) and *RHG1a* (group G) genes are also expressed, in both cases in all tissues tested.

Although RT-PCR cannot be used for quantitative measurements of specific gene products, it can give some information about the extent of mRNA accumulation. A 10-fold reduction in the amount of RNA template severely reduced the band intensity for reactions with RING-H2 primers, whereas no reduction in band intensity was observed for reactions specific for the constitutively expressed *EF-1α* control [26] (data not shown). Low levels of expression of the RING-H2 genes are also suggested by the low number of the corresponding ESTs (Table 1). The analyses suggest that some of the RING-H2 genes are widely expressed, and that their protein products may participate in general cellular functions. However, induction by specific conditions or accumulation in specialized organs would not be revealed in the analyses performed.

3.6. Predominance of the RING-H2 motif

New *Arabidopsis* genes encoding highly conserved RING-H2 fingers appear constantly in the databases. Although a certain bias resulting from our search strategy cannot be excluded, the results obtained in this study suggest that in *Arabidopsis* the RING-H2 variant is more widespread than the classical RING finger. The motif has been conserved in different contexts in otherwise unrelated proteins with presumably diverse cellular functions. This suggests that the domain is a stable and functional framework and that the short divergent polar regions are sufficient for specific functions.

4. Conclusions

The function of the RING-H2 proteins cannot be predicted solely on the basis of homology to proteins of known function, since it is likely to be determined by the cooperativity of several independent domains. A large catalogue of the proteins will allow selection of models for both structural and physiological studies. A screen for gene-knock-out mutants or transgenic plants over- or underexpressing the proteins may help delineate their physiological roles. Thus, reverse-genetics can complement other studies of *Arabidopsis* which typically involve phenotypic mutant screens followed by gene isolation.

Acknowledgements: We thank Drs. Stanley Brown and John Mundy for comments on the manuscript. This work was supported by the Danish Natural Science Research Council (9502825).

References

[1] Hanson, I.M., Poustka, A. and Trowsdale, J. (1991) Genomics 10, 417–424.
[2] Freemont, P.S., Hanson, I.M. and Trowsdale, J. (1991) Cell 64, 483–484.
[3] Freemont, P.S. (1993) Ann. NY Acad. Sci. 684, 174–192.
[4] Borden, K.L.B. and Freemont, P.S. (1996) Curr. Opin. Struct. Biol. 6, 395–401.
[5] Deng, X.W., Matsui, M., Wei, N., Wagner, D., Chu, A.M., Feldmann, K.A. and Quail, P.H. (1992) Cell 71, 791–801.

[6] Martinez-Garcia, M., Garciduenas-Pina, C. and Guzman, P. (1996) Mol. Gen. Genet. 252, 587–596.
[7] Schauser, L., Christensen, L., Borg, S. and Poulsen, C. (1995) Plant Physiol. 107, 1457–1458.
[8] Zou, J. and Taylor, D.C. (1997) Gene 196, 291–295.
[9] Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) J. Mol. Biol. 215, 403–410.
[10] Hebsgaard, S.M., Korning, P.G., Tolstrup, N., Engelbrecht, J., Rouze, P. and Brunak, S. (1996) Nucleic Acids Res. 24, 3439–3452.
[11] Joshi, C.P., Zhou, H., Huang, X. and Chiang, V.L. (1997) Plant Mol. Biol. 35, 993–1001.
[12] Brendel, V., Bucher, P., Nourbakhsh, I., Blaisdell, B.E. and Karlin, S. (1992) Proc. Natl. Acad. Sci. USA 89, 2002–2006.
[13] Bairoch, A., Bucher, P. and Hofmann, K. (1997) Nucleic Acids Res. 25, 217–221.
[14] Lupas, A., Van Dyke, M. and Stock, J. (1991) Science 252, 1162–1164.
[15] Nakai, K. and Kanehisa, M. (1992) Genomics 14, 897–911.
[16] Rost, B., Sander, C. and Schneider, R. (1994) Comput. Appl. Biosci. 10, 53–60.
[17] Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F. and Higgins, D.G. (1997) Nucleic Acids Res. 25, 4876–4882.
[18] Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
[19] Mundy, J., Mayer, R. and Chua, N.-H. (1995) Plant Mol. Biol. Rep. 13, 156–163.
[20] Barlow, P.N., Luisi, B., Milner, A., Elliott, M. and Everett, R. (1994) J. Mol. Biol. 237, 201–211.
[21] Borden, K.L.B., Boddy, M.N., Lally, J., O'Reilly, N.J., Martin, S., Howe, K., Solomon, E. and Freemont, P.S. (1995) EMBO J. 14, 1532–1541.
[22] Berg, J.M. and Shi, Y. (1996) Science 271, 1081–1085.
[23] Nielsen, H., Engelbrecht, J., Brunak, S. and von Heijne, G. (1997) Protein Eng. 10, 1–6.
[24] Landschulz, W.H., Johnson, P.F. and McKnight, S.L. (1988) Science 240, 1759–1764.
[25] Newman, T., de Bruijn, F.J., Green, P., Keegstra, K., Kende, H., McIntosh, L., Ohlrogge, J., Raikhel, N., Somerville, S., Thomashow, M., Retzel, E. and Somerville, C. (1994) Plant Physiol. 106, 1241–1255.
[26] Axelos, M., Bardet, C., Liboz, T., Le Van Thai, A., Curie, C. and Lescure, B. (1989) Mol. Gen. Genet. 219, 106–112.
[27] Kraulis, P.J. (1991) J. Appl. Crystallogr. 24, 946–950.