

Computational analyses and annotations of the *Arabidopsis* peroxidase gene family

Lars Østergaard^a, Anders G. Pedersen^b, Hans M. Jespersen^a, Søren Brunak^b,
Karen G. Welinder^{a,*}

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

^bCenter for Biological Sequence Analysis, The Technical University of Denmark, Building 208, DK-2800 Lyngby, Denmark

Received 6 June 1998; revised version received 3 July 1998

Abstract Classical heme-containing plant peroxidases have been ascribed a wide variety of functional roles related to development, defense, lignification, and hormonal signaling. More than 40 peroxidase genes are now known in *Arabidopsis thaliana* for which functional association is complicated by a general lack of peroxidase substrate specificity. Computational analysis was performed on 30 near full-length *Arabidopsis* peroxidase cDNAs for annotation of start codons and signal peptide cleavage sites. A compositional analysis revealed that 23 of the 30 peroxidase cDNAs have 5' untranslated regions containing 40–71% adenine, a rare feature observed also in cDNAs which predominantly encode stress-induced proteins, and which may indicate translational regulation.

© 1998 Federation of European Biochemical Societies.

Key words: Adenine-rich 5' untranslated region; Neural network prediction; Peroxidase; Plant cDNA; Translational control

1. Introduction

Plant peroxidases containing heme at their catalytic center have been ascribed a variety of biological functions, including intracellular scavenging of peroxide, participation in extracellular defense processes, lignification, and hormonal signaling [1–4]. The enzymatic action of peroxidase is well-understood and results in efficient conversion of hydrogen peroxide to water and formation of radicals (A[•]) from a variety of organic and inorganic substrates (HA): $H_2O_2 + 2HA \rightarrow 2H_2O + 2A^{\bullet}$ [5]. These radicals may continue in non-enzymatic reactions including polymerization and depolymerization depending on their chemical nature and the environment. Peroxidase isoenzymes show little substrate specificity, and it has therefore been difficult to assign precise biological functions to them.

At the same time, it has become increasingly clear that plants express large numbers of peroxidase isoenzymes. Nine different peroxidase genes or proteins have been known in horseradish for several years [6]. In order to understand the complexity of biological functions of plant peroxidases, we have attempted to provide an overview of the repertory of peroxidase genes in a single plant, *Arabidopsis thaliana*, taking advantage of the international effort to produce EST (ex-

pressed sequence tag) clones and ultimately, the entire genomic sequence. This has resulted so far in the identification and complete sequencing of cDNAs encoding three novel intracellular ascorbate-type peroxidases, in addition to two previously known [7], and 24 near full-length clones encoding novel ER-targeted classic-type peroxidases (Welinder et al., in preparation). The protein products of these 24 clones have never been observed before and it is not trivial to assign authentic initiation codon and sites of post-translational modifications. Here, we use neural network-based computer programs to predict start codons as well as signal peptide cleavage sites for the 24 peroxidases.

A striking result, which we obtained by analyzing the peroxidase cDNAs, was an unusually high content (46% on average) of adenine residues in the 5' untranslated regions (UTRs). Searching for 5' UTR sequences similar to peroxidase mRNA 5' UTRs, cDNA clones encoding stress-related proteins were dominant. Since cap-independent translation has been observed in situations of stress [8], we suggest that expression of these peroxidases is regulated at the translational level.

Here, we also stress the importance of using computer-based calculations to obtain more precise annotations of mRNA sequences before submission to the database. This is beneficial for scientists working in related fields and crucial in the construction of useful prediction programs.

2. Materials and methods

2.1. *Arabidopsis* peroxidases

The EST database (dbEST) [9,10] was searched for peroxidase-like sequences from *Arabidopsis*, resulting in the identification of 200 cDNAs encoding 24 different peroxidases belonging to the class of classic ER-targeted peroxidases [11]. Beside these, six peroxidase-encoding EST clones were identified which only contained part of the coding region and were thus not included in the calculations. Two peroxidase ESTs were considered identical if they shared a segment of 100 or more nucleotides with more than 95% identity. We obtained the EST clones from the *Arabidopsis* Biological Resource Center at Ohio State University, and completed the sequencing ([12], Welinder et al., in preparation). Four previously published peroxidase cDNAs [13–15], and two EST variants [16] were also included in the present analysis.

2.2. *Arabidopsis* reference cDNAs

A redundancy reduced database containing 524 *Arabidopsis* cDNAs was used for analysis of possible unique traits in *Arabidopsis* peroxidase UTRs (Pedersen and Nielsen, in preparation). A redundancy reduced *Arabidopsis* genomic database which contain 129 genes [17] was used for intron analysis.

2.3. Sequence analysis tools

In this paper we used neural networks for two purposes: prediction

*Corresponding author. Present address: Biotechnology Laboratory, University of Aalborg, Sohngaardsholmsvej 57, DK-9000 Aalborg, Denmark. Fax: (45) 98 14 25 55. E-mail: welinder@civil.auc.dk

of start codons ([18], <http://www.cbs.dtu.dk/services/NetStart/>) and prediction of signal peptides ([19], <http://www.cbs.dtu.dk/services/SignalP/>). Briefly, neural networks consist of a number of simple, non-linear computational units that operate in parallel [20]. By training with known data, it is possible to obtain networks that are able to distinguish different classes in input data, e.g. start codon vs. non-start codon, and signal peptide vs. non-signal peptide. Information from both positive and negative cases is taken into account, in such a way that residues and positions that are important will be weighted accordingly.

The neural networks used in this study were of the feed-forward type, and had three layers of neurons [20,21]: an input layer with a window scanning the nucleotide or amino acid sequence, a hidden layer, and an output layer. Details of the implementation have been published elsewhere [22,23]. Input sequences were presented to the networks by encoding the DNA or amino acid sequences into a binary string, using a coding scheme where each monomer is represented by a string of binary digits. In the case of DNA, A = 0001, C = 0010, G = 0100, and T = 1000 were used [20,24].

We investigated the RNA structure of the peroxidase UTRs using the program RNAfold [25–27]. Consensus signals in the 3' UTR were analyzed using the program Wordup [28,29].

For clarity, AUG will be used further on in referring to start codons even though the data set comprises cDNAs. Moreover, we will use the term near full-length about peroxidase cDNAs containing the entire putative open reading frame, since cDNAs in the *Arabidopsis* EST database rarely contain the authentic 5' end (own experience).

3. Results and discussion

3.1. *Arabidopsis peroxidases*

The majority of *Arabidopsis* ESTs come from a cDNA library representing five types of tissue [9] and some originate from other cDNA libraries from a variety of specific tissues [10]. A few peroxidase genes are represented by up to 40 entries in dbEST (ATP 1 and ATP 2) [12], some are absent (ATP Ca, ATP Cb, and ATP A2) [13–15], whereas eight occur only once. The translated amino acid sequences of the 30 mature peroxidases discussed in this study (Table 1) are 33–99% identical and give a dendrogram (phylogenetic tree) very similar to one composed of all known plant peroxidases irrespective of species (data not shown). Hence, the 30 *Arabidopsis* peroxidase mRNAs can be considered to be representative of those for all plant peroxidases. However, the total number of peroxidase-encoding genes in *Arabidopsis* is still unknown and may be higher than 100, considering that recently released *Arabidopsis* genomic sequences reveal 23 peroxidase genes of which only 11 are among the 40 presently known (Welinder et al., in preparation).

3.2. AUG context and prediction of translation initiation site

Translation initiation on most eukaryotic mRNAs is de-

Table 1
Annotation of start codon position, adenosine content, and signal peptide cleavage site

| Accession no. ^a | Peroxidase | Initiation | | % A in 5' UTR | Signal peptide | |
|-----------------------------|------------|-------------------------|---------------|---------------|-----------------------|----------------------|
| | | Nucleotide ^b | Score | | Cleavage ^c | Y score ^d |
| M58380 | ATP Ca | 54 | 0.570 | 28.3 | Q32/L22 | 0.722/0.151 |
| X71794 | ATP Cb | 35 | 0.508 | 26.5 | Q31 | 0.567 |
| M58381 | ATP Ea | 64 | 0.885 | 34.9 | Q30/C19 | 0.633/0.102 |
| X99952 | ATP A2 | 49 | 0.816 | 41.7 | Q31 | 0.746 |
| X98453 | ATP N | 23/44 | 0.537/0.674 | 40.9/48.8 | Q29 | 0.896 |
| X98189 | ATP 1a | 25 | 0.942 | 70.8 | V24 | 0.781 |
| X98313 | ATP 1b | 50 | 0.944 | 49.0 | V24 | 0.658 |
| X98190 | ATP 2a | 42 (49) | 0.357 (0.620) | 56.1 | E29 | 0.778 |
| X98317 | ATP 2b | 37 (44) | 0.448 (0.655) | 55.6 | E29/I22 | 0.778/0.322 |
| X98808 | ATP 3a | 56 | 0.920 | 49.1 | Q24 | 0.889 |
| X98773; X98318 ^e | ATP 4a | 33 | 0.832 | 65.6 | L23/K32 | 0.484/0.843 |
| X98809 | ATP 5a | 28/16 | 0.654/0.487 | 29.6/20.0 | F28 | 0.810 |
| X98774; X98320 ^e | ATP 6a | 30 | 0.833 | 48.3 | S24 | 0.778 |
| X98321 | ATP 7a | 57/48 | 0.585/0.259 | 50.0/51.1 | Q28 | 0.853 |
| X98855 | ATP 8a | 14 | 0.665 | 38.5 | Q26 | 0.871 |
| X98314; X98856 ^e | ATP 9a | 25 | 0.735 | 33.3 | Q26 | 0.889 |
| X98928; X98323 ^e | ATP 10a | 39 | 0.878 | 42.1 | Q26 | 0.805 |
| X98802 | ATP 11a | 46 | 0.769 | 40.0 | I22 | 0.886 |
| X98775; X98319 ^e | ATP 12a | 45 | 0.843 | 59.1 | Q24 | 0.896 |
| X98776; X98322 ^e | ATP 13a | 50 | 0.797 | 55.1 | Q23 | 0.826 |
| X98803 | ATP 14a | 23 | 0.779 | 54.5 | Q27 | 0.781 |
| X99097 | ATP 15a | 44 | 0.883 | 41.9 | Q24 | 0.840 |
| X98777; X98315 ^e | ATP 16a | 32 (365) | 0.345 (0.658) | 41.9 | Q30 | 0.819 |
| X99096; X98316 ^e | ATP 17a | 65 | 0.887 | 42.2 | L23 | 0.645 |
| X98804 | ATP 18a | 13 | 0.955 | 66.7 | H24 | 0.868 |
| X98805 | ATP 19a | 31 | 0.775 | 46.7 | Q24 | 0.892 |
| X98806 | ATP 20a | 29 | 0.600 | 53.6 | Q31 | 0.719 |
| X98807 | ATP 21a | 49 | 0.876 | 25.0 | Q26 | 0.841 |
| Y08781 | ATP 22a | 23 (150) ^f | 0.328 (0.560) | 40.9 | Q24 | 0.890 |
| Y08782 | ATP 23a | 55 | 0.589 | 42.6 | F21 | 0.587 |

^aReferences of published sequences are in the text and listed with GenBank links at <http://biobase.dk/~welinder/prx.html>.

^bPresent versus data base annotation of start codon. Only one position is given in case of agreement. Numbers in parentheses indicate predictions which are not valid and which probably reflect the error rate of the program.

^cPresent versus data base annotation of the N-terminal residue of the putative mature peroxidase. Only one position is given in case of agreement.

^dSee description of Y score in legend to Fig. 1.

^eTwo entries of the same cDNA have been deposited in GenBank.

^fFor ATP 22a, no full-length clone was obtained containing the entire open reading frame. However, the remaining sequence was obtained from the chromosome sequencing project and the prediction program at http://l25.itba.mi.cnr.it/~webgene/wwwHC_tata.html [30] was used to annotate the TATA box.

pendent on a cap structure at the 5' end and occurs at the first AUG codon in accordance with the scanning behavior of the ribosomal 40S subunit [31]. In cases where the first AUG is not recognized, it is usually because the sequence surrounding the AUG is not optimal for initiation. In eukaryotes, the start codon context 5'-GCCACCaugG-3' has been suggested and, furthermore, several of these nucleotides were found to be optimal for initiation [31]. These data largely relate to vertebrates, whereas studies of plant sequences indicate that plant mRNAs from dicots have a different start codon preference, 5'-AAAaugGC-3' ([32,33], Pedersen and Nielsen, in preparation).

The correct reading frame is easily identified for a peroxidase-encoding cDNA, since the most conserved amino acid sequence around the catalytic distal histidine is found only

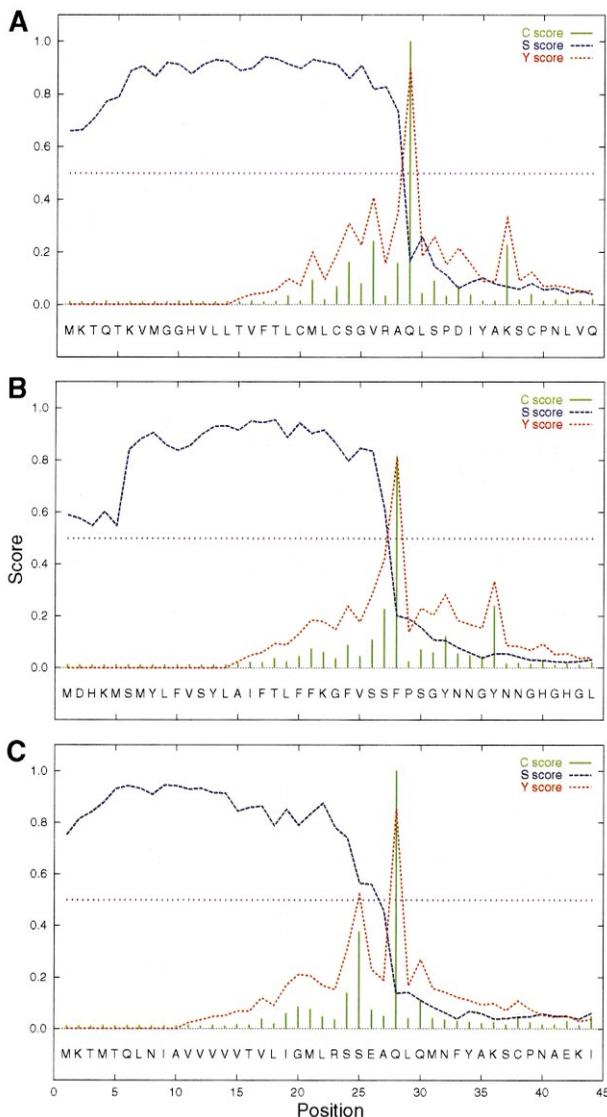


Fig. 1. Prediction of signal peptide cleavage sites to generate mature proteins of (A) ATP N, (B) ATP 5a, and (C) ATP 7a. Sequences were submitted to the SignalP server [19]. The C score is the probability of cleavage at a certain position, the S score gives the probability of an amino acid in a certain position to be situated in a signal peptide, while Y is obtained by combining the height of the C score with the slope of the S score. The prediction of cleavage site location is optimized by observing where the C score is high and the S score changes from a high to a low value.

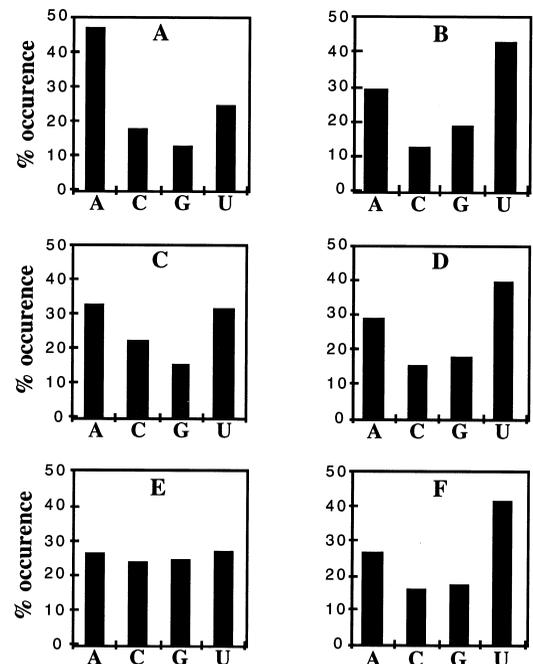


Fig. 2. Nucleotide distribution in (A) 5' UTRs from 30 *Arabidopsis* peroxidase cDNAs, (B) 3' UTRs from 30 *Arabidopsis* peroxidase cDNAs, (C) 5' UTRs from 524 *Arabidopsis* cDNAs, (D) 3' UTRs from 524 *Arabidopsis* cDNAs, (E) coding regions from 524 *Arabidopsis* cDNAs, and (F) introns from 129 *Arabidopsis* genes.

60–70 residues from the initiating methionine. This is approximately 250 nucleotides from the 5' end of a full-length mRNA. In the absence of experimental data, an automatic assignment [18] of the most likely start codons was performed (Table 1). For the majority of the peroxidase-encoding cDNAs, the 5' proximal AUG was predicted as start codon; however, in seven cases, the network predicted alternative start codons. Four of these either contained the predicted start codon in another reading frame (ATP 2a and ATP 2b) or the predicted AUG was located too far downstream for the open reading frame to encode a peroxidase (ATP 16a and ATP 22a). These cases probably reflect a low expression level of these peroxidase genes with inefficient initiation of translation from the 5' proximal start codon, but may also represent the low error-rate of the program (10%) [18]. For the remaining three peroxidase-encoding cDNAs, ATP N, ATP 5a, and ATP 7a, a strong alternative in-frame start codon is located 3–7 triplets downstream from the first one. This could mean that if the first AUG triplet is missed, translation can start efficiently on the second – a phenomenon known as leaky scanning [31]. It is worth noting that in the case of ATP N, the region between the two methionines contains 57% adenine while the first region has 40%. This may suggest a special function for the intervening sequence (further discussion of the adenine-rich peroxidase 5' UTRs below).

The translation initiation context determined for *Arabidopsis* reference mRNAs [18] and peroxidase mRNAs was compared and no significant difference was found as judged by a χ^2 -test (data not shown).

3.3. Prediction of signal peptide

The peroxidase sequences analyzed here are all ER-targeted by typical eukaryotic signal peptides [34]. No N-terminal ami-

Table 2
Arabidopsis genes with 5' UTRs containing >40% adenosine

| Acc.no. | Product | Reg. ^a | % A ^b | Length ^c |
|---------|-----------|-------------------|------------------|---------------------|
| X55053 | Cor 6.6 | 1 | 61.2 | 49 |
| X90959 | Cor 47 | 1 | 40.7 | 108 |
| X67670 | Lti 65 | 1 | 50.6 | 85 |
| X67671 | Lti 78 | 1 | 46.9 | 81 |
| X51474 | Atkin 1 | 1, 2 | 53.3 | 60 |
| X62281 | Atkin 2 | 1, 2 | 58.3 | 60 |
| X68042 | Atrab 18 | 1, 2 | 40.4 | 89 |
| L04637 | Lipoxy | 2, 3, 4 | 48.8 | 80 |
| X53435 | Aux 2-11 | 5 | 45.6 | 68 |
| X53436 | Aux 2-27 | 5 | 52.4 | 42 |
| S70188 | Saur-ac 1 | 5 | 47.0 | 83 |
| U27609 | TCH 4 | 5, 6 | 52.0 | 50 |
| Z19050 | Nia 1 | 7 | 41.3 | 116 |
| X68146 | Athb-2 | 8 | 44.7 | 76 |
| L19354 | UBC 4 | 9 | 55.8 | 86 |
| L41244 | Thi 2.1 | 10 | 43.1 | 51 |
| L41245 | Thi 2.2 | 10 | 58.8 | 17 |

^aRegulatory functions are 1: cold-induced, 2: abscisic acid-induced, 3: pathogen-induced, 4: jasmonate-induced, 5: hormone-regulated, 6: environmentally regulated, 7: encoding chlorate resistance, 8: far-red-light-induced, 9: ubiquitinating, and 10: other than pathogen-induced.

^bAdenine content in 5' UTR.

^cNumber of nucleotides in 5' UTR.

no acid sequence has been determined experimentally so far for a mature *Arabidopsis* peroxidase. Therefore, we applied a computational method analyzing both the residues within the putative signal, and around the putative cleavage site based on a network trained on experimental eukaryotic data [19]. Three selected examples are shown in Fig. 1, and all predictions are compiled in Table 1. The N-termini of horseradish peroxidases HRP C, HRP E5, and HRP A2 (highly similar to *Arabidopsis* peroxidases ATP Cb and ATP Ca, ATP Ea, and ATP A2, respectively) have been determined by experimental protein sequence analysis ([35,36], Rasmussen et al., in preparation). In all cases, the mature peroxidases start with a glutamine modified to a pyroglutamyl residue. The method used here generally provided clear-cut predictions with high scores of significance and for the majority of the peroxidases, cleavage was predicted to give rise to an N-terminal glutamine residue.

In the three cases where in-frame start codons were predicted 3–7 codons downstream, it is notable that the signal peptide (S) score is lower for the residues between the first and the second methionines, encoded by the two AUGs, than for the rest of the putative signal. This suggests that, if the first start codon is not recognized, the second AUG can be used in all three cases without loss of signaling function. Deleting the amino acids between the first and second methionine indeed resulted in equal predictions with scores of similar significance (data not shown).

3.4. Characterization of UTRs

A striking result of the present analysis is the observation that the 5' UTRs of *Arabidopsis* peroxidases contain significantly more adenine than 5' UTRs from *Arabidopsis* reference cDNAs (Fig. 2A,C). Twenty-three out of 30 near full-length peroxidase cDNAs have 40–71% A residues compared to an average of 32% for the reference cDNAs. The remaining seven peroxidases (ATP Ca, Cb, Ea, 5a, 8a, 9a, and 21a) have normal A contents of 25–38%. An analysis of 71 5' UTRs from

peroxidase cDNAs from other plants showed that 24 had A >40%. The smaller fraction of A-rich peroxidase 5' UTRs from other plants probably reflects the non-random cloning strategies, since the probes often have been designed based on the sequence of HRP C [37,38] and in the case of horseradish [37] and *Arabidopsis*, 5' UTRs from HRP C-type peroxidases are not A-rich.

The efficiency of translation initiation from a eukaryotic mRNA is to a large degree determined by the length and secondary structure of the 5' UTR [31]. Eukaryotic mRNAs have 5' UTRs, that range from three to several hundred nucleotides with an average length of about 90 [39]. It seems that a moderately long (>20 nt) and unstructured 5' UTR is necessary and sufficient for efficient initiation of translation, while a heavily structured leader is a major obstacle to the scanning 40S ribosomal subunit [31]. The A-rich peroxidase 5' UTRs presumably have very little potential to form stable secondary structures. Analyzing all regions of the *Arabidopsis* peroxidase cDNAs using the RNAfold program indeed suggested that both the 5' UTRs and 3' UTRs are generally unstructured.

It is also remarkable that the nucleotide compositions in introns and 3' UTRs (Fig. 2B,D,F) are very similar. In accordance with Joshi [40], we found that 34% of 3' UTRs of the *Arabidopsis* redundancy reduced data set contained the highly conserved polyadenylation signal AAUAAA known from mammals [41] located 20–25 nt upstream from the polyadenylation site. This is significantly more than expected on a random basis. However, since 66% do not contain the AAUAAA signal this clearly indicates differences in the recognition between plants and mammals. The lack of a single conserved polyadenylation signal and the very similar nucleotide compositions in introns and 3' UTRs seems to support the previously published hypothesis of a mechanistic interplay between splicing and polyadenylation that has been suggested to occur in plants [42].

In order to investigate if a high A content is a general feature of 5' UTRs from mRNAs encoding a special group of proteins, *Arabidopsis* 5' UTRs that contained >40% A were extracted from the 524 cDNA sequences. Interestingly, 16 out of 24 fulfilling these criteria appeared to encode proteins induced by biotic or abiotic stress or by hormones (Table 2) while cDNAs containing <25% A in their 5' UTR only included 10% of stress-related genes (data not shown). It has not been possible to identify a common motif for the A-rich leaders. On the contrary, the adenosines seem to be dispersed evenly throughout the leader.

Based on the unusual A-rich character of the majority of the peroxidase 5' UTRs from *Arabidopsis*, and the 5' UTRs from mRNAs encoding stress-related *Arabidopsis* proteins, we suggest that a common regulation mechanism between these two groups exists. The A-rich quality of these 5' UTRs may affect transcription, translation, and/or mRNA stability; however, since translation regulation usually is mediated through the 5' UTR and since no common motif was found which could act as a promoter element, we suggest that expression of these genes are under translational control. For several of the genes in Table 2, promoters have been cloned and elements responsible for regulating transcriptional activity identified [43–45]. It is, however, not unusual that regulation occurs at both levels [46].

Cap-independent translation is linked to development

[47,48] and also occurs under conditions of stress, e.g. pathogen attack, wounding, heat, cold, or draught. During stress, proteins which are involved in protection of the organism must be expressed. It is therefore important that the mRNAs encoding these proteins can be translated in a cap-independent fashion. There are several examples of translational discrimination of stress-induced proteins [8,49,50], and it has been found that >80% of the 5' UTRs from heat shock proteins from invertebrates and plants are AU-rich [8]. Whether the putative cap-independent translation on these 5' UTRs occurs due to lack of secondary structure or through binding of specific proteins remains to be established. Here, we suggest that expression of the majority of peroxidases is stress- or developmentally induced at the translational level and that the A-rich regions of their 5' UTRs are able to direct cap-independent initiation.

Acknowledgements: This research was supported by a fellowship from the University of Copenhagen to L.Ø., the Danish Agricultural and Veterinary Research Council (5.23.26.10-1 to K.G.W.), the Danish National Science Research Council (9502825 to K.G.W.), and the Danish National Research Foundation (A.G.P. and S.B.). DNA sequencing of peroxidase EST clones was performed largely by Dr. Annemarie F. Justesen and Ms. Inger V. H. Kjærsgård, the Department of Protein Chemistry, and assisted by Dr. Søren K. Rasmussen, Risø National Laboratory, Denmark.

References

- [1] Chen, G.X. and Asada, K. (1992) *Plant Cell Physiol.* 33, 117–123.
- [2] Roberts, E. and Kolattukudy, P.E. (1989) *Mol. Gen. Genet.* 217, 223–232.
- [3] Wallace, G. and Fry, S. (1994) *Int. Rev. Cytol.* 151, 229–267.
- [4] Lagrimini, L.M., Gingas, V., Finger, F., Rothstein, S. and Liu, T.Y. (1997) *Plant Physiol.* 114, 1187–1196.
- [5] Dunford, H. B. (1991) in: *Peroxidases in Chemistry and Biology* (Everse, J., Everse, K.E., and Grisham, M.B., Eds.), pp. 1–23, CRC Press, Boca Raton, FL.
- [6] Welinder, K.G. (1992) in: *Plant Peroxidases 1980–1990, Topics and Detailed Literature on Molecular, Biochemical, and Physiological Aspects* (Penel, C., Gaspar, Th., and Greppin, H., Eds.), pp. 1–24, University of Geneva Press, Geneva.
- [7] Jespersen, H.M., Kjærsgård, I.V.H., Østergaard, L. and Welinder, K.G. (1997) *Biochem. J.* 326, 305–310.
- [8] Joshi, C.P. and Nguyen, H.T. (1995) *Nucleic Acids Res.* 23, 541–549.
- [9] 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.
- [10] Cooke, R., Raynal, M., Laudicé, M., Grellet, F., Delseny, M., Morris, P.C., Guerrier, D., Giraudat, J., Quigley, F., Clabault, G., Li, Y.F., Mache, R., Krivitzky, M., Gy, J.J., Kreis, M., Lecharny, A., Parmentier, Y., Marbach, J., Fleck, J., Clement, B., Philipps, G., Hervé, C., Bardet, C., Tremousaygue, D., Lescuré, B., Lacomme, C., Roby, D., Jourjon, M.F., Chabrier, P., Charpentreau, J.L., Desprez, T., Amselem, J., Chiapello, H. and Höfte, H. (1996) *Plant J.* 9, 101–124.
- [11] Welinder, K.G. (1992) *Curr. Opin. Struct. Biol.* 2, 388–393.
- [12] Kjærsgård, I.V.H., Jespersen, H.M., Rasmussen, S.K. and Welinder, K.G. (1997) *Plant Mol. Biol.* 33, 699–708.
- [13] Intaprak, C., Yamamoto, K., Fujiyama, K., Takano, M. and Shinmyo, A. (1993) *J. Ferment. Bioeng.* 75, 166–172.
- [14] Intaprak, C., Takano, M. and Shinmyo, A. (1994) *Plant Physiol.* 104, 285–286.
- [15] Østergaard, L., Abelskov, K.A., Mattsson, O. and Welinder, K.G. (1996) *FEBS Lett.* 398, 243–247.
- [16] Capelli, N., Tognolli, M., Flach, J., Overney, S., Penel, C., Grep-pin, H. and Simon, P. (1996) *Plant Physiol.* 112, 446.
- [17] Korning, P.G., Hebsgaard, S.M., Rouze, P. and Brunak, S. (1996) *Nucleic Acids Res.* 24, 316–320.
- [18] Pedersen, A.G. and Nielsen, H. (1997) in: *Proceedings of the Fifth International Conference on Intelligent Systems for Molecular Biology*, pp. 226–233, AAAI Press.
- [19] Nielsen, H., Engelbrecht, J., Brunak, S. and von Heine, G. (1997) *Protein Eng.* 10, 1–6.
- [20] Hertz, J., Krogh, A. and Palmer, R.G. (1991) *Introduction to the Theory of Neural Computation*, Addison-Wesley, Redwood City, CA.
- [21] Minsky, M. and Papert, S. (1969) *Perceptions: An Introduction to Computational Geometry*, MIT Press, Cambridge, MA.
- [22] Brunak, S., Engelbrecht, J. and Knudsen, S. (1990) *Nature* 343, 123.
- [23] Brunak, S., Engelbrecht, J. and Knudsen, S. (1991) *J. Mol. Biol.* 220, 49–65.
- [24] Qian, N. and Sejnowski, T.J. (1988) *J. Mol. Biol.* 202, 865–884.
- [25] Zuker, M. and Stiegler, P. (1981) *Nucleic Acids Res.* 9, 133–148.
- [26] Turner, D.H., Sugimoto, N. and Freier, S.M. (1988) *Annu. Rev. Biophys. Chem.* 17, 167–192.
- [27] McCaskill, J.S. (1990) *Biopolymers* 29, 1105–1119.
- [28] Pesole, G., Prunella, N., Liuni, S., Attimonelli, M. and Saccone, C. (1992) *Nucleic Acids Res.* 20, 2871–2875.
- [29] Pesole, G., Attimonelli, M. and Saccone, C. (1996) *Methods Enzymol.* 266, 281–294.
- [30] Milanese, L., Muselli, M. and Arrigo, P. (1996) *Comput. Appl. Biosci.* 12, 399–404.
- [31] Kozak, M. (1991) *J. Biol. Chem.* 266, 19867–19870.
- [32] Lütcke, H.A., Chow, K.C., Mickel, F.S., Moss, K.A., Kern, H.F. and Scheele, G.A. (1987) *EMBO J.* 6, 43–48.
- [33] Joshi, C.P., Zhou, H., Huang, X. and Chiang, V.L. (1997) *Plant Mol. Biol.* 35, 993–1001.
- [34] von Heine, G. (1986) *Nucleic Acids Res.* 14, 4683–4690.
- [35] Welinder, K.G. (1976) *FEBS Lett.* 72, 19–23.
- [36] Morita, Y., Mikami, B., Yamashita, H., Lee, J.Y., Aibara, S., Sato, M., Katsube, Y. and Tanaka, N. (1991) in: *Biochemical, Molecular, and Physiological Aspects of Plant Peroxidases* (Lobarewski, J., Greppin, H., Penel, C. and Gaspar, Th., Eds.), pp. 81–88, University of Geneva Press, Geneva.
- [37] Fujiyama, K., Takemura, H., Shinmyo, A., Okada, H. and Takano, M. (1990) *Gene* 89, 163–169.
- [38] Buffard, D., Breda, C., van Huystee, R.B., Asemoto, O., Pierre, M., Ha, D.B. and Esnault, R. (1990) *Proc. Natl. Acad. Sci. USA* 87, 8874–8878.
- [39] Kozak, M. (1991) *Gene Express.* 1, 111–115.
- [40] Joshi, C.P. (1987) *Nucleic Acids Res.* 15, 9627–9640.
- [41] Wickens, M. (1990) *Trends Biochem. Sci.* 15, 277–281.
- [42] Luehrsen, K.R. and Walbot, V. (1994) *Genes Dev.* 8, 1117–1130.
- [43] Jaglo-Ottosen, K.R., Gilmour, S.J., Zarka, D.G., Schabenberger, O. and Thomashow, M.F. (1998) *Science* 280, 104–106.
- [44] Stockinger, E.J., Gilmour, S.J. and Thomashow, M.F. (1997) *Proc. Natl. Acad. Sci. USA* 94, 1035–1040.
- [45] Yamaguchi-Shinozaki, K. and Shinozaki, K. (1994) *Plant Cell* 6, 251–264.
- [46] Nielsen, F.C., Gammeltoft, S. and Christiansen, J. (1990) *J. Biol. Chem.* 265, 13431–13434.
- [47] Fütterer, J. and Hohn, T. (1996) *Plant Mol. Biol.* 32, 159–189.
- [48] Teerink, H., Kasperaitis, M.A., De-Moor, C.H., Voorma, H.O. and Thomas, A.A. (1994) *Biochem. J.* 303, 547–553.
- [49] Hess, M.A. and Duncan, R.F. (1996) *Nucleic Acids Res.* 24, 2441–2449.
- [50] Pitto, L., Gallie, D.R. and Walbot, V. (1992) *Plant Physiol.* 100, 1827–1833.